Sources of Long-term Variability in Measurements of Lung Function*

Implications for Interpretation and Clinical Trial Design

Robert L. Jensen, PhD; John G. Teeter, MD; Richard D. England, MD, PhD; Heather M. Howell; Heather J. White, DVM; Eve H. Pickering, PhD; and Robert O. Crapo, MD, FCCP

Background: **The objective of the study was to characterize the biological and technical components of variability associated with longitudinal measurements of FEV_1 and carbon monoxide diffusing capacity (DLCO). Variability was apportioned to subject and instrument for five commercially available pulmonary function testing (PFT) systems: Collins CPL (Ferraris Respiratory; Louisville, CO); Morgan Transflow Test PFT System (Morgan Scientific; Haverhill, MA); SensorMedics Vmax 22D (VIASYS Healthcare; Yorba Linda, CA); Jaeger USA Masterscreen Diffusion TP (VIASYS Healthcare; Yorba Linda, CA); and Medical Graphics Profiler DX System (Medical Graphics Corporation; St. Paul, MN).**

Methods: **This was a randomized, replicated cross-over, single-center methodology study in 11 healthy subjects aged 20 to 65 years. Spirometry and DLCO measurements were performed at baseline, 3 months, and 6 months. Repetitive simulations of FEV_1 and DLCO were performed on the same instruments on four occasions over a 90-day period using a spirometry waveform generator and a DLCO simulator.**

Results: **The coefficient of variation associated with repetitive measurements of FEV_1 or DLCO in subjects was consistently larger than that associated with repetitive simulated waveforms across the five instruments. Instrumentation accounted for 13 to 58% of the total FEV_1 and 36 to 70% of the total DLCO variability observed in subjects. Sample size estimates of hypothetical studies designed to detect treatment group differences of 0.050 L in FEV_1 and 0.5 mL/min/mm Hg in DLCO varied as much as four times depending on the instrument utilized.**

Conclusions: **These results provide a semiquantitative assessment of the biological and technical components of PFT variability in a highly standardized setting. They illustrate how instrument choice and test variability can impact sample size determinations in clinical studies that use FEV_1 and DLCO as end points.** ***(CHEST 2007; 132:396–402)***

Key words: clinical trial design; diffusing capacity; diffusing capacity simulator; pulmonary function testing; pulmonary waveform generator; spirometry; variability

Abbreviations: ATS = American Thoracic Society; CV = coefficient of variation; DLCO = carbon monoxide diffusing capacity; PFT = pulmonary function testing; RMSCV = root mean square coefficient of variation

The sources of variation in lung function measurement have been described by Becklake and White,[1] but the relative contributions of technical (instrument) and biological elements have not been well studied, especially for carbon monoxide diffusing capacity (DLCO) and lung volumes. Much of the early focus was on the technical aspects, specifically improving spirometers. The development of standard spirometric waveforms and mechanical simulators that could deliver them accurately allowed reliable testing of spirometer performance and has been included in current American Thoracic Society (ATS) and European Respiratory Society spirometry standards.[2,3] In 1990, Nelson et al[4] studied 62 different spirometer models using a mechanical spirometry waveform simulator and the ATS standard

waveforms. Of the spirometers tested, 29% failed the tests and 14.5% were judged marginal. In all cases, specific problems were identified by the testing; when those problems were corrected, the spirometers passed.

Two of the authors (R.C. and R.J.) have since developed a DLCO simulator to eliminate biological

For related article see page 388
For editorial comment see page 367

qsources of variability,[5] thereby making it possible to better identify instrument errors. The DLCO simulator utilizes two precision syringes in conjunction with a high precision mix of gases (tracer gas and carbon monoxide) to simulate a range of physiologically relevant DLCO values. Although there is no mandate for its use, some manufacturers have adopted it as part of their quality control.

This study was designed to quantify biological variability in spirometry and DLCO measurements over a 6-month period using a variety of modern instruments. The impetus for the study was the need to determine sample size and measurement frequency for clinical studies.

Materials and Methods

This was a randomized, replicated, cross-over, single-center methodology study to assess intrainstrument and intrasubject variability of pulmonary function testing (PFT) measurements over a 6-month period. The local institutional review board approved the study protocol, and the study was conducted in accordance with the ethical principles of the Declaration of Helsinki. All subjects gave written informed consent.

Instruments Tested

One new instrument model from each of five PFT equipment manufacturers was purchased. The instrument manufacturers and models were as follows: Collins CPL (Ferraris Respiratory; Louisville, CO); Morgan Transflow Test PFT System (Morgan Scientific; Haverhill, MA); SensorMedics Vmax 22D (VIASYS Healthcare; Yorba Linda, CA); Jaeger USA Masterscreen Diffusion TP (VIASYS Healthcare; Yorba Linda, CA); and Medical Graphics Profiler DX System (Medical Graphics Corporation; St. Paul, MN). These instruments were purchased by the study sponsor, Pfizer Inc., for the purpose of validating instruments being considered to measure PFT end points in clinical drug trials. Each instrument was set up and maintained according to manufacturer specifications. All PFT instruments were powered on continuously for the duration of the study. Instruments were calibrated or calibration checked according to manufacturer specifications on test days.

Study Population

Nonsmoking male and female subjects (20 to 65 years old) with no history of respiratory diseases or symptoms who passed a screening history and physical examination were eligible for inclusion. Women who were pregnant, lactating, or not using adequate contraception were excluded. Additional exclusion criteria were as follows: history of smoking in the 2 years prior to screening or a life-long total of >5 pack-years; history of recreational drug use in the year prior to the study; recent eye surgery; treatment with any asthma medications or corticosteroids (except nasal corticosteroids for allergic rhinitis); any respiratory tract ailment in the 6 weeks prior to the study; inability to perform acceptable quality PFT at screening; recent blood donation; or any health condition that would, in the judgement of the investigators, interfere with the study. Lung function exclusion criteria were as follows: FVC or $FEV_1 > 120\%$ or $< 70\%$ of predicted,[6] or $FEV_1/FVC < 70\%$, or DLCO $>120\%$ or $< 70\%$ of predicted.[6,7] Upper and lower exclusion limits were used to minimize regression to the mean.

Assessments

At the initial screening visit, the following were obtained: a medical history and physical examination including measurements of height and weight, and measurements of spirometry and DLCO using clinical laboratory instruments. Subjects were required to have negative alcohol breathalyzer (AlcoMate CA 2000 Digital Alcohol Detector; Wookyung Tech; Incheon City, South Korea) results at screening to continue study participation.[8] Blood samples for hemoglobin concentrations were obtained at screening and weeks 12 and 24 and used to adjust DLCO values to a standard hemoglobin concentration according to ATS recommendations.[9] Room temperature and barometric pressure were recorded on test days. Adverse events were collected at each study visit.

PFT and spirometry were performed on each subject at three time points: baseline (0 to 2 weeks), 3 months (12 to 14 weeks), and 6 months (24 to 26 weeks). All PFT was performed at the LDS hospital (Salt Lake City, UT) by two experienced technicians according to ATS standards.[9,10] DLCO washout and sample volumes were fixed on the Collins CPL and were not adjustable in the Morgan, Jaeger, or Medical Graphics devices. Only on the SensorMedics Vmax were DLCO alveolar samples adjusted when thought necessary in the judgement of the technician performing the test. A computer-generated randomization scheme was utilized to ensure that each subject was tested on each instrument in a unique sequence. At each testing interval, subjects were tested on each instrument twice in a 2-week period with a restriction that they could only be tested on one instrument per day. On each test day, subjects were required to complete three acceptable FVC and three acceptable DLCO maneuvers on the selected instrument. The largest measured FEV_1 from the three acceptable trials was recorded as the FEV_1 value for that test day. All three acceptable DLCO measurements were recorded for each test day.

*From LDS Hospital and University of Utah (Drs. Jensen and Crapo, and Ms. Howell), Salt Lake City, UT; and Pfizer Global Research and Development (Drs. Teeter, England, White, and Pickering), Groton, CT.
The study was sponsored by Pfizer Inc.
Drs. Jensen and Crapo are consultants for Pfizer Inc, and both received royalties from Hans Rudolph Company. Drs. Teeter, England, and Pickering, and Ms. White are employed by Pfizer Inc. Ms. Howell has no conflict of interest to disclose.
Manuscript received August 11, 2006; revision accepted February 16, 2007.

Correspondence to: Robert L. Jensen, PhD, Pulmonary Laboratory, LDS Hospital and University of Utah, Eighth Ave and C St, Salt Lake City, UT 84143; e-mail: ldrjens1@ihc.com
DOI: 10.1378/chest.06-1999