Respiratory Medicine (2013) **107**, 1409–1416

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/rmed

Quality and reproducibility of spirometry in COPD patients in a randomized trial (UPLIFT®)

W. Janssens [a,*], Y. Liu [b], D. Liu [c], S. Kesten [d], D.P. Tashkin [e], B.R. Celli [f], M. Decramer [g]

[a] *University Hospital Leuven, Leuven, Belgium*
[b] *University of North Carolina at Chapel Hill, NC, USA*
[c] *Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, USA*
[d] *Cytori Therapeutics, Inc., San Diego, CA, USA*
[e] *David Geffen School of Medicine, UCLA, Los Angeles, CA, USA*
[f] *Brigham and Women's Hospital, Boston, MA, USA*
[g] *University Hospital, Katholieke Universiteit Leuven, Leuven, Belgium*

Received 30 October 2012; accepted 24 April 2013
Available online 25 May 2013

KEYWORDS
Forced expiratory volume;
Forced vital capacity;
Placebo-controlled;
Tiotropium

Summary
Background: This study explores spirometry quality and reproducibility in the Understanding Potential Long-term Impacts on Function with Tiotropium (UPLIFT®) trial.
Methods: Four-year, randomized, double-blind, placebo-controlled, multicenter trial in 5993 patients with chronic obstructive pulmonary disease. Within-test variability of pre- and post-bronchodilator forced expiratory volume in 1 s (FEV_1) was compared across study visits. Between-test variability of best pre- or post-FEV_1 values between two visits 6 months apart was compared at the start, middle and end of the trial.
Results: Three or more acceptable maneuvers were obtained in 93% of visits. Within-test variability of pre- and post-FEV_1 (mean standard deviation: 0.092 and 0.098 L) decreased during the trial. Between-test variability also decreased: pre-FEV_1 (visit 3–5 = 0.141 ± 0.138 L; visit 9–11 = 0.129 ± 0.121 L; visit 17–19 = 0.121 ± 0.122 L); post-FEV_1 (0.139 ± 0.140, 0.126 ± 0.123, 0.121 ± 0.122 L, respectively), and was dependent on age, sex, smoking status and disease stage, but not on bronchodilator response or study treatment.

* Corresponding author. Respiratory Division, Department of Pneumology, University Hospital Leuven, Herestraat 49, 3000 Leuven, Belgium. Tel.: +32 1634 6800; fax: +32 1634 6803.
E-mail addresses: wim.janssens@uzleuven.be, wim.janssens@med.kuleuven.be (W. Janssens).

0954-6111/$ - see front matter
http://dx.doi.org/10.1016/j.rmed.2013.04.015

Conclusion: Spirometry quality in UPLIFT® was good and improved during the trial. Between-test variability across patient subgroups suggests that relevant cut-offs for individual disease monitoring are difficult to establish.
Trial registration number: NCT00144339.

Introduction

International guidelines recognize spirometry as the gold standard for diagnosing, categorizing and monitoring disease progression in chronic obstructive pulmonary disease (COPD).[1] Within a spirometry session, a valid test is defined by three technically acceptable and reproducible forced expiratory volume in 1 s (FEV_1) and forced vital capacity (FVC) maneuvers.[2,3] The largest FVC and the largest FEV_1 should be recorded after examining the data from all of the acceptable time/volume curves, even if they do not come from the same curve; reproducibility is defined as the difference between the highest and second-highest FVC and FEV_1 from acceptable curves.[2] The spirometry standardization paper of the American Thoracic Society (ATS) and the European Respiratory Society (ERS) states that a variance of $\leq$0.15 L in the two largest values of FEV_1 or FVC meets reproducibility criteria (or $\leq$0.1 L, where FVC is $\leq$1 L).[3] Earlier ATS criteria allowed for $\leq$0.2 L difference.[2] If these criteria are not met in three maneuvers, up to a total of eight maneuvers should be attempted to meet the minimal requirements. Standardization is essential to obtain the "true" spirometric values within one session, particularly for COPD trials evaluating small differences in lung function as primary outcomes.

Between-test variability may be greater than within-test variability, owing to technical differences in the testing procedures (including equipment variability), or intra-patient factors such as the degree of airway obstruction, changes in bronchomotor tone, diurnal variation in FEV_1, baseline FEV_1 levels or bronchodilator reversibility.[4–7] Other factors that may affect test performance include smoking, medication use or recent illness.[2,3] Previous data have shown mean between-test differences of 0.1 $\pm$ 0.1 L for FEV_1.[7–9] Confounding variability in spirometry must be minimized in COPD trials, as there may be little difference in lung function decline over time between active treatment and control groups.[5,10,11] Additionally, between-test variance should be considered when using spirometry for individual patient monitoring. Yet there are no specific recommendations for thresholds defining a clinically relevant difference in FEV_1 between spirometry sessions.[6,7]

The Understanding Potential Long-term Impacts on Function with Tiotropium (UPLIFT®) trial was a 4-year, randomized, double-blind, placebo-controlled study conducted in 487 study centers in 37 countries.[11] To ensure high-quality spirometry, diurnal variation was minimized, all sites were provided with the same training and electronic data transmission provided study staff with automated, real-time feedback on the quality and reproducibility of the measurements via a centralized quality assurance review.[5] Therefore, the UPLIFT® database represents a unique opportunity to perform post hoc evaluation of quality and reproducibility of spirometric measurements in COPD patients.

Methods

Study design and population

UPLIFT® was a 4-year, randomized, placebo-controlled clinical trial of tiotropium in 5993 patients with COPD (ClinicalTrials.gov number, NCT00144339.).[11] Details of the UPLIFT® study design can be found in the online depository. All patients gave written informed consent. The study was approved by local ethical review boards and conducted in accordance with the Declaration of Helsinki.

Spirometry assessments

To standardize spirometry, all sites were provided with identical spirometry systems (KoKo® Spirometer, Quantum Research Inc., Louisville, CO, USA) with customized, study-specific software. All technicians performing pulmonary function testing received identical, detailed training and were required to show proficiency in using the equipment and performing technically acceptable pulmonary function tests, before testing study patients. After each test was performed, the spirometry software gave immediate feedback to the technician to show whether the maneuver met ATS acceptability and reproducibility standards. All data were stored electronically.[5] Details of the spirometry data quality assurance methods used are in the online depository.

To reduce diurnal variation in FEV_1, spirometry was performed at approximately the same time in the morning at screening (baseline, visit 1), on randomization to treatment (day 1, visit 2), on day 30 (visit 3) and then every 6 months (at visits 5, 7, 9, 11, 13, 15, 17 and 19) until the end of treatment. A final test was performed at the 30-day post-treatment follow-up visit (end of trial).

Post-bronchodilator spirometry was performed at the 90-min time point. Detailed information on the spirometry protocol can be found in the online depository. FEV_1 and FVC measurements were obtained in triplicate following slow vital capacity measurement. The best FEV_1 and FVC values of three attempts meeting ATS reproducibility criteria ($\leq$5% or $\leq$0.2 L, current at the time of trial design)[2] were recorded for the data set.

Statistical analysis

All analyses were restricted to visits with at least three acceptable maneuvers.

Within-test variability of pre- and post-bronchodilator FEV_1 was assessed by using the following model at each visit: Y_{ik} = mean $FEV_1 + b_i + e_{ik}$ (where Y_{ik} = acceptable maneuvers, ik = kth maneuver of the ith subject, b_i = random individual effect accounting for between-