Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

A Randomized Trial of Long-term Oxygen for COPD with Moderate Desaturation
by
Long-term Oxygen Treatment Trial Research Group*

Supplement

Table of Contents   **Page**

Outcomes reported
The LOTT Protocol specified 17 hypotheses, 14 of which are addressed in this manuscript. The outcomes of maintenance of nutritional status and risk of cardiovascular disease will be addressed in secondary publications. Neurocognitive assessment was not included in the LOTT data collection schedule so that hypothesis cannot be addressed.

Methods for assessing resting and exercise desaturation
Resting and exercise desaturation were assessed with the Masimo Radical 7® pulse oximeter (Irvine, CA); the sampling rate was once every second and once every 2 seconds for the resting and exercise assessments, respectively. Resting saturation was calculated as the mean of the acceptable quality samples obtained in the last 5 minutes of a 6-minute test session; the coefficient of variation of the samples had to be ≤2.5%. Exercise desaturation was assessed during the 6-minute walk; desaturation was rated severe if any mean of 30 consecutive samples (≥20 having acceptable quality) was <80%, and was rated moderate if 5 consecutive samples (all having acceptable quality) were <90% and severe exercise desaturation was not present.

Methods for ascertainment of the primary outcome
Vital status was determined by clinic report and review of the Social Security Master Death File[1]. Hospitalizations and COPD exacerbations were ascertained by self-report supplemented by medical records review.

Randomization process
The randomization schedule was stratified by regional clinical center with randomly permuted blocks of sizes 2, 4, and 6. The data system generated the treatment assignment only if the electronic checks for conformance with the eligibility criteria were passed.

Additional details on sample size calculation
For the No LTOT group, we assumed a 33% annual hospitalization rate in those with, and a 10% annual hospitalization rate in those without, a COPD hospitalization in the prior year, and a 7% annual mortality rate in those with, and a 6% rate in those without, a COPD hospitalization in the prior year. We also assumed that 50% of all enrolled patients would have had a COPD hospitalization in the prior year, yielding an estimated 28% annual composite event rate. These assumptions were the consensus of the investigators after reviewing available data[2-8] and considering differences between the populations studied and the proposed population. We assumed that the primary outcome would be obtained on 99% of patients.

Additional details on statistical analysis
Side-by-side histograms were used to compare the treatment groups on changes from baseline in measured outcomes. P-values for the treatment group comparisons on the ranked changes were determined by Wilcoxon rank-sum test.

Results are reported as mean ± standard deviation (SD) except where noted. Bonferroni corrections were used to determine the P-value required for statistical significance for secondary and other outcomes,[9] as specified in each table or figure caption. Analyses were conducted in SAS (Cary, NC), STATA (College Station, TX), or R (https://www.r-project.org/).

Comparison of self-reported oxygen use to measured oxygen use
All patients randomized to the LTOT group and using stationary oxygen concentrators were asked to report meter reading and date read approximately every 2 months for the duration of their follow-up. Because of problems with small dials, inaccessible dials, and change outs from one concentrator to another due to malfunction, dissatisfaction or other reason, data were not always available or, if available, not always usable (e.g., date of change out provided but starting or ending reading not provided, or reading provided but not the date of the reading). For these reasons, the analysis on self-reported versus measured oxygen usage was limited to 100 patients using stationary concentrators who were able to provide usable data for at least 4 months (3 usage readings); the mean number of usage estimates was 12.5 compared to mean number of self-reports of 10. The analyses presented used all available data during each patient's follow-up. The 100 patients included in the analysis constitute 27% of the LTOT group and 35% of the 286 stationary concentrator users. The patients included in this adherence analysis were similar to the LTOT patients who were not included in gender, minority

status, education level, MMRC score, GOLD COPD score, and exacerbation history. However, the LTOT patients included in this adherence analysis were more likely to have exercise desaturation only and less likely to have resting desaturation only than the LTOT patients who were not included. Figure S2 presents data relating to validation of the self-report oxygen usage estimates.

Members of the LOTT Research Group

Office of the Chair of the Steering Committee, University of Alabama, Birmingham: William C. Bailey, M.D.
Regional clinical centers:
Brigham and Women's Hospital: Anne L. Fuhlbrigge; M.D., Ernestina Sampong.
    Associated sites: **Boston Medical Center**: Karin Sloan, M.D.; Ashley Wagner; Susan Anderson.
                          Boston VA: Marilyn Moy, M.D.; Osarenoma Okunbor.

Cleveland Clinic: James K. Stoller, M.D., M.S. (Principal Investigator); Scott Marlow, R.R.T., Yvonne Meli, R.N., Richard Rice, R.R.T., M.Ed. (Study Coordinators); Loutfi S. Aboussouan, M.D., Robert Castele, M.D., Joseph Parambil, M.D., Sumita Khatri, M.D., Aman Pande, M.D., Joe Zein, M.D., Thomas Olbrych, M.D. (Co-Investigators).
    Associated sites: **Crouse Medical Practice**: Stephan Alkins, M.D.; Christine Jocko, M.A.
                          Cleveland Clinic Florida: Franck Rahaghi, M.D., M.H.; Jean Barton, M.B.A.

Denver Health and Hospital Authority: Richard K. Albert, M.D.; Jennifer Underwood.
    Associated sites: **National Jewish Health:** Barry Make, M.D., F.A.C.P., F.C.C.P., F.A.A.C.V.P.R.; Jennifer Underwood.

Duke University: Neil MacIntyre, M.D.; John Davies.

Kaiser Foundation Hospitals: Thomas Stibolt, M.D.; Richard Mularski, M.D.; Allison Naleway, Ph.D.; Sarah Vertrees.

Los Angeles Biomedical Research Institute at Harbor - UCLA Medical Center: Richard Casaburi, Ph.D., M.D.; Janos Porszasz, M.D., Ph.D.; Peggy Walker, R.R.T.; Renee Indelicato.
    Associated sites: **Loma Linda VA**: Lennard Specht, M.D.; Kathleen Ellstrom, Ph.D., R.N.; Jamie Portillo, R.R.T.
                          City of Hope National Medical Center: David Horak, M.D.; Brian Tiep, M.D.; Mary Barnett, R.N.

Ohio State University: Philip Diaz, M.D.; Janice Drake; Mahasti Rittinger; Rachael Compton, Scott Miller.
    Associated site: **University of Cincinnati**: Ralph J. Panos, M.D.; Laura A. Lach, B.H.S.

Temple University: Gerard Criner, M.D.; Carla Grabianowski, B.S.N., R.N., C.C.R.P; Francis Cordova, M.D.; Parag Desai, M.D.; Samuel Krachman, D.O.; James Mamary, M.D.; Nathaniel Marchetti, M.D.; Aditi Satti, M.D.; Eileen Mumm, C.R.N.P.; Michelle Vega-Olivo, C.R.N.P.; Jenny Hua; Vanna Tauch; Lii-Yoong Criner, R.N., C.C.R.C.; Michael Jacobs, Pharm.D.; Peter Rising, M.S.
    Associated sites: **Geisinger Institute**: Paul Simonelli, M.D.; Michele Mitchell, B.S.N., R.N., C.C.R.C.
                          Louisiana State University: Matthew Lammi, M.D.; Connie Romaine, M.S.N., A.P.R.N.-N.P.-C.
                          Institute for Respiratory and Sleep Medicine: Howard Lee, M.D.; Mary Ianacone, D.O.
                          University of Maryland: Steven Scharf, M.D., Ph.D.; Wanda Bell-Farrell.
                          Buffalo VA: M. Jeffery Mador, M.D.; Ayesha Rahman, M.S.
                          Respiratory Specialists: Mumtaz Zaman, M.D.; Lisa Hill L.P.N., C.R.C.; Alec Platt, M.D.

University of Alabama: J. Allen Cooper, Jr., M.D.; Kathleen Harrington, Ph.D., M.P.H.; Mark Dransfield, M.D.; Patti Smith, R.N.; Donald Davis.
    Associated sites: **Birmingham VA:** J. Allen Cooper, Jr., M.D.; Patti Smith, R.N.
                          North Florida/South Georgia VA: Peruvemba Sriram, MD; Katherine Herring.

University of Michigan: Steven E. Gay, M.D.; Fernando Martinez, M.D., MS; Meilan Han, M.D.; Kelly Rysso; Catherine Meldrum, Ph.D., R.N., M.S., C.C.R.C.
    Associated sites: **Beaumont Hospital**: K. P. Ravikrishnan, M.D.; Daniel Keena, M.D.; Jennifer DeRidder, R.N.; Beth Kring, C.N.M., C.C.R.C.
        San Antonio VA: Antonio Anzueto, M.D.; Alex Aguilera; Timothy Houlihan, R.N.
        Spectrum Health: Reda Girgis, M.D.; Jennifer Cannestra, R.N., B.S.N.

University of Pittsburgh: Frank Sciurba, M.D.; Benjamin Kelly.

University of Utah: Richard E. Kanner, M.D.; Mary Beth Scholand, M.D.; G. Martin Villegas; Judy Carle.

University of Washington: David H. Au, M.D., M.S.; Edmunds Udris, M.P.H.
    Associated sites: **Harborview Medical Center**: Randall Curtis, M.D., M.P.H.
        VA Puget Sound HCS: David Au, M.D., M.S.; Laura C. Feemster, M.D, M.S.; Richard Goodman, M.D.; Brianna Moss, B.S.; Lynn Reinke, Ph.D., A.R.N.P.; Edmunds Udris, M.P.H.
        University of Washington Medical Center: Moira Aitken, M.D.; Bruce Culver, M.D.

Washington University: Roger D. Yusen, M.D., M.P.H.; Mario Castro, M.D., M.P.H.; Brigitte Mittler, B.A.; Jeanne Heaghney, R.N.
    Associated sites: **Pulmonary Consultants/Christian Hospital**: Myron Jacobs, M.D.
        University of Illinois at Chicago: Min Joo, M.D., M.P.H.; Nina Bracken, A.P.N.
        Suburban Lung Associates: Edward Diamond, M.D.; Mary K. Joseph, Ph.D.
        University of California, San Diego: Xavier Soler, M.D., Ph.D.; Arianna Villa, B.S., R.R.T.
        Central Florida Pulmonary Group: Daniel Layish, M.D.

Biospecimen Repository, Channing Division of Network Medicine, Brigham and Women's Hospital: Edwin Silverman, M.D., Ph.D.; Roxanne Kelly, B.S., M.B.A.; Daniel Cossette, B.S.

Data Coordinating Center, Johns Hopkins University: James Tonascia, Ph.D.; Patricia Belt, Amanda L. Blackford, Sc.M.; Betty Collison; John Dodge; Michele Donithan, M.H.S.; Cathleen Ewing; Rosetta Jackson; K Patrick May, M.S.; Jill Meinert; Steven Piantadosi, M.D., Ph.D.; Girlie Reyes, B.S.; David Shade, J.D.; Michael Smith, B.S.; Alice L. Sternberg, Sc.M.; Mark Van Natta, M.H.S., Laura Wilson, Sc.M.; Annette Wagoner; Robert Wise, MD; Katherine P. Yates, Sc.M.

Centers for Medicare and Medicaid Services: Rosemarie Hakim, Ph.D.

National Heart, Lung, and Blood Institute: Antonello Punturieri, M.D., Ph.D.; Julie Bamdad, M.S.E.; Thomas Croxton, Ph.D., M.D.; Joanne Deshler; Pamela McCord-Reynolds; Mario Stylianou, Ph.D.; Gail Weinmann, M.D (DSMB executive secretary).

Data and Safety Monitoring Board: Gordon Bernard, M.D. (chair; Vanderbilt University); James Anderson, Ph.D. (2007-2015; Frontier Science); Bernard Lo, M.D. (2007-2013; University of California, San Francisco); Andrew Ries, M.D., M.P.H. (2007-2014; University of California, San Diego); Stuart Stoloff, M.D. (University of Nevada); Byron Thomashow, M.D. (Columbia University); Barbara Tilley, Ph.D. (University of Texas); Kevin Weiss, M.D. (Accreditation Council of Graduate Medical Education).

Figure S1. CONSORT diagram for the Long-term Oxygen Treatment Trial.
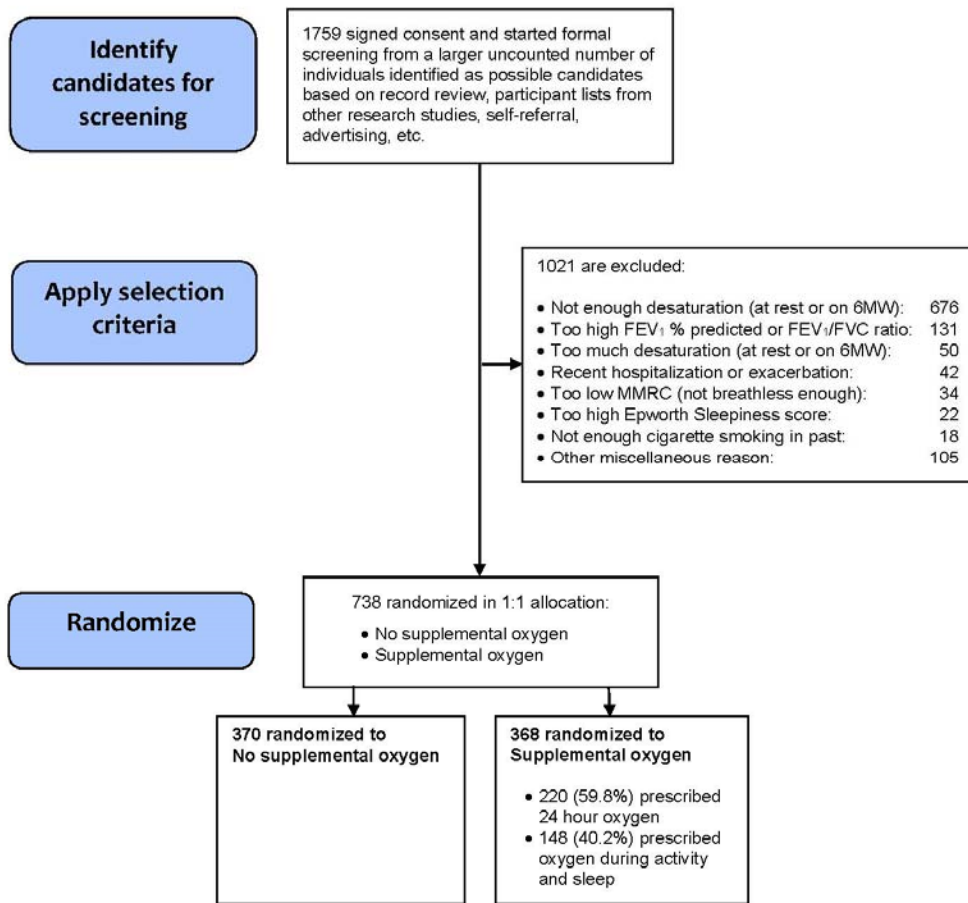
Notes: 6MW = 6-minute walk; FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity

The Modified Medical Research Council (MMRC) dyspnea score is a single item scale that is completed by the patient; the score ranges from 0 to 4, with higher score indicating greater breathlessness. MMRC=0 was exclusionary in LOTT[10,11].

The Epworth Sleepiness Scale is an 8-item scale that measures general daytime sleepiness and is completed by the patient; the total score ranges from 0 to 24, with higher scores indicating greater daytime sleepiness. Score ≥16 was exclusionary in LOTT[12].

Figure S2. Validation of self-report oxygen usage estimates. Usage estimates from stationary concentrator meter readings were available for 100 LTOT patients (each patient had at least 2 usage assessments by meter reading [at least 3 meter readings] and more than half of the patient's usage assessments were plausible [0-24 hours/day]); 12.5 ± 5.0 assessments per patient). Panel A is a plot of mean usage assessment by concentrator versus the mean self-report of stationary oxygen use (mean of all self-reports; 10 ± 4.2 reports per patient). Panel B is a Bland-Altman plot of the difference (self-report – meter) versus the mean ([self-report + meter] /2) and shows a significant (P<0.001) linear relationship of decreasing difference (and subsequent reversal) with increasing mean. The regression equation (standard error) was $Y = 3.5(0.7) + 0.3(0.06) * X$. The 95% limits of agreement on the slope were $\pm$ 5.9 hours. The Shapiro-Wilk test for normality of the residuals was not rejected (P=0.20).

(A)

Regression Line: Y(Meter)=-1.4 + 1.1*X(Self Report)
Prob(Slope=1)=0.17
Correlation = 0.83

(B)

Regression Line: Y(Difference)=3.5-0.3*X(Mean)
Prob(Slope=0)<0.001
95% Limits of Agreement on Regression Line = ±5.9
P-value from Shapiro-Wilk Test of Normality of Residuals = 0.20
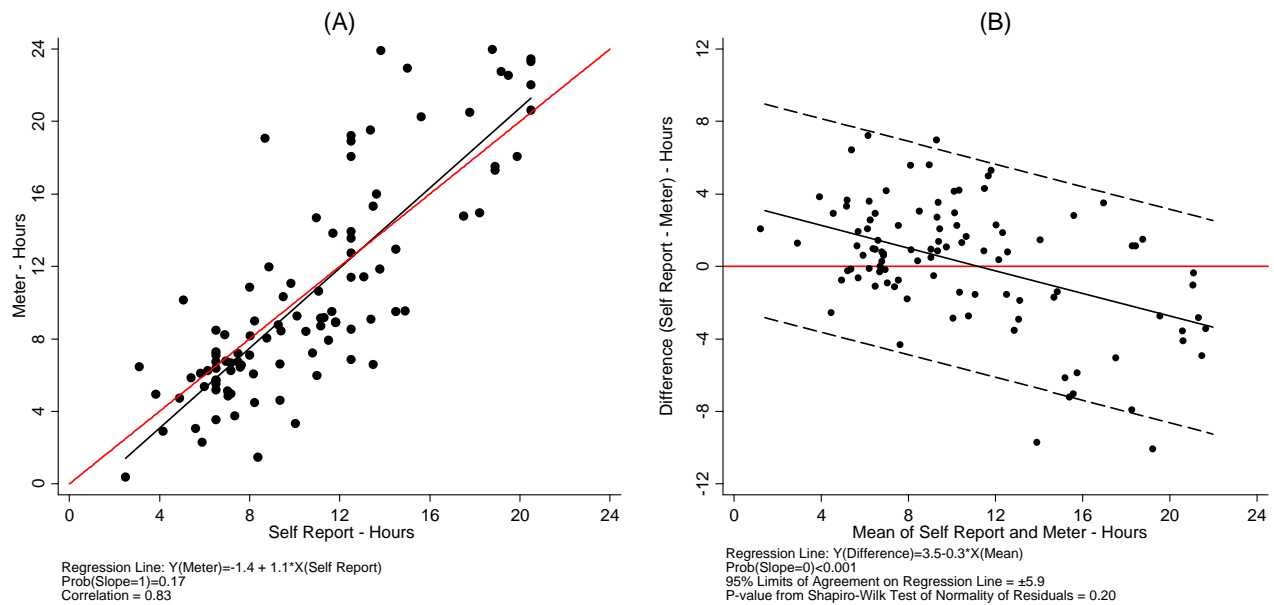
Figure S3. Kaplan –Meier analyses of secondary outcomes:
- A) Time to first COPD exacerbation (median follow-up of 11.4 months; 114 LTOT patients and 111 No LTOT patients were censored as of their date death (if no COPD exacerbation prior to death) or as of the date of their last interview (if alive and no COPD exacerbation);
- B) Time to first hospitalization for COPD exacerbation (median follow-up of 24.3 months; 243 LTOT patients and 238 No LTOT patients were censored as of their date of death (if no COPD hospitalization prior to death) or as of the date of their last interview (if alive and no COPD hospitalization);
- C) Time to first hospitalization for COPD exacerbation or death, whichever came first (median follow-up of 24.3 months; 201 LTOT patients and 205 No LTOT patients who neither died nor had a COPD hospitalization were censored as of the date of their last interview);
- D) Time to first hospitalization for reason other than COPD exacerbation (median follow-up of 18.7 months; 181 LTOT patients and 179 No LTOT patients were censored as of their date of death (if no non COPD hospitalization prior to death) or as of the date of their last interview (if alive and no non COPD hospitalization).

The hazard ratios and 95% confidence limits were calculated from Cox regression models with LTOT (solid line) versus No LTOT (dashed line) as the single model variable; P-values were derived from logrank tests. This was an intention-to-treat analysis. P<0.0125 (0.05/4) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
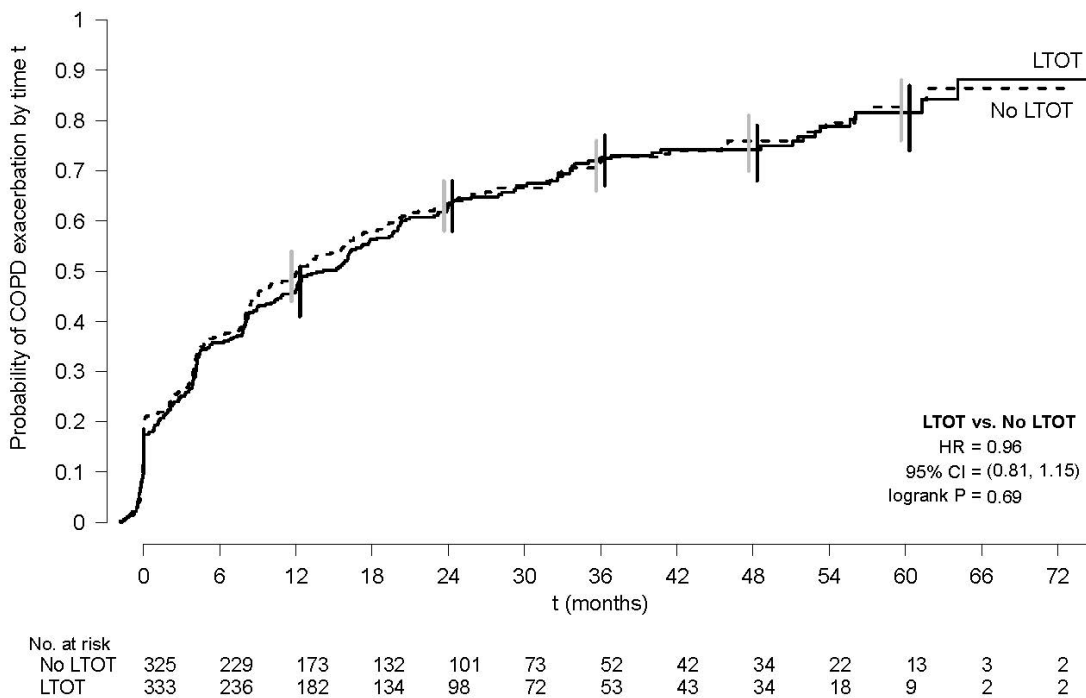
A) Time to first COPD exacerbation

No. at risk

No LTOT	325	229	173	132	101	73	52	42	34	22	13	3	2
LTOT	333	236	182	134	98	72	53	43	34	18	9	2	2

B) Time to first hospitalization for COPD exacerbation

LTOT

No LTOT

LTOT vs. No LTOT
HR = 0.9
95% CI = (0.71, 1.15)
logrank P = 0.41

No. at risk													
No LTOT	370	335	292	248	208	160	128	103	87	64	40	13	2
LTOT	368	350	304	264	219	176	142	109	86	57	37	15	2

C) Time to first hospitalization for COPD exacerbation or death

No. at risk													
No LTOT	370	335	292	248	208	160	128	103	87	64	40	13	2
LTOT	368	350	304	264	219	176	142	109	86	57	37	15	2

D) Time to first hospitalization for reason other than COPD exacerbation
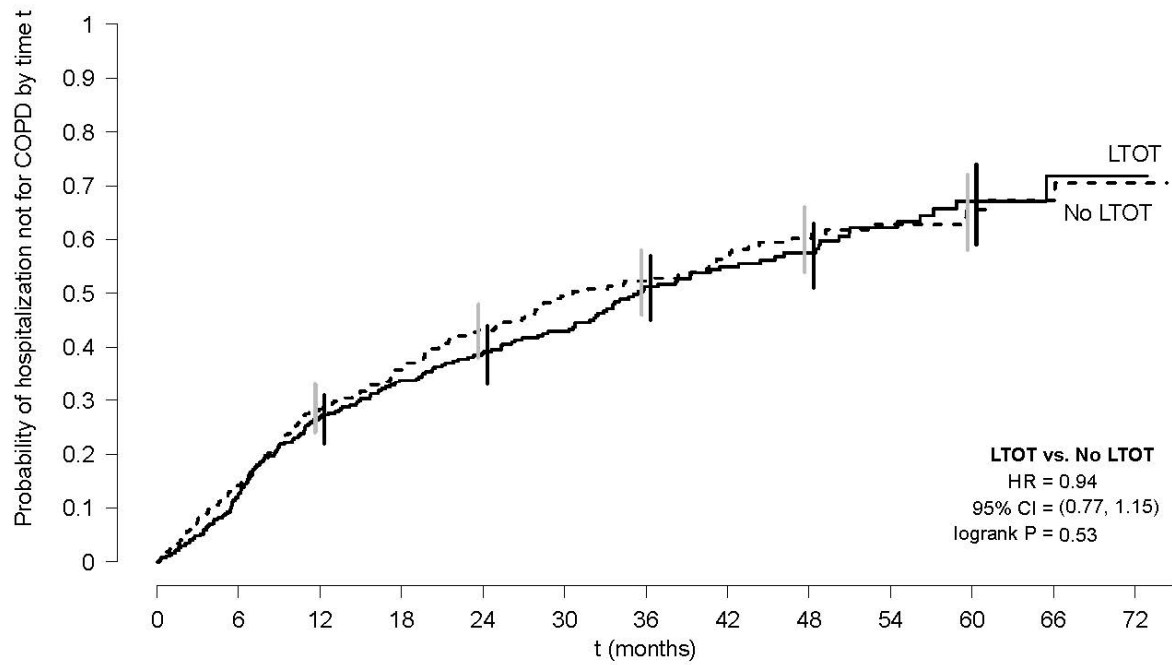
Figure S4. Histograms of changes from baseline to 4, 12, 16, 24, 36, and 48 months after randomization:

      A) Quality of Well-Being Scale (QWB) mean daily score
      B) St. George's Respiratory Questionnaire (SGRQ) total score
      C) SF-36 Physical Component Summary (PCS) score
      D) SF-36 Mental Component Summary (MCS) score
      E) Hospital Anxiety and Depression Questionnaire (HADS) anxiety score
      F) HADS depression score
      G) Pittsburgh Sleep Quality Index (PSQI) total score
      H) Post-bronchodilator FEV_1 (L)
      I) Room air 6-minute walk distance (feet)
      J) Room air resting oxygen saturation (SpO_2, %).

Analyses are limited to patients whose visit window for the specified follow-up time had closed as of the end of the trial. If the measure was completed, the change in the measure from baseline to the specified follow-up time was calculated by subtracting the baseline value from the value at the specified follow-up time and the change was then scored from 1 to 10. Except for the QWB, which is anchored in death (score for death=0), scores were not imputed for those who had died and deaths are ranked below the worst change category. In the analyses including those who were alive and missed the assessment, such patients were ranked above the deaths but below the worst change category. The P-values compare LTOT versus No LTOT distributions of changes and were derived from Wilcoxon rank-sum tests on the scores. P-values labeled "without missing" exclude the patients who were alive and missed the assessment; P-values labeled "with missing" include the patients who were alive and missed the assessment (shown in the bar labeled "missing"). Readers should interpret the P-values with caution since the difference between the treatment groups in the proportion missing may make a "with missing" P-value statistically significant. The degree to which the distribution is shifted to the upper right of the chart indicates the degree of relative benefit of LTOT over No LTOT. This is an intention-to-treat analysis.

A) Change in Quality of Well-Being Scale (QWB) mean daily score (range 0-1, higher score indicates better quality of life, MID=0.03)[13,14]. P<0.008 (0.05/6 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
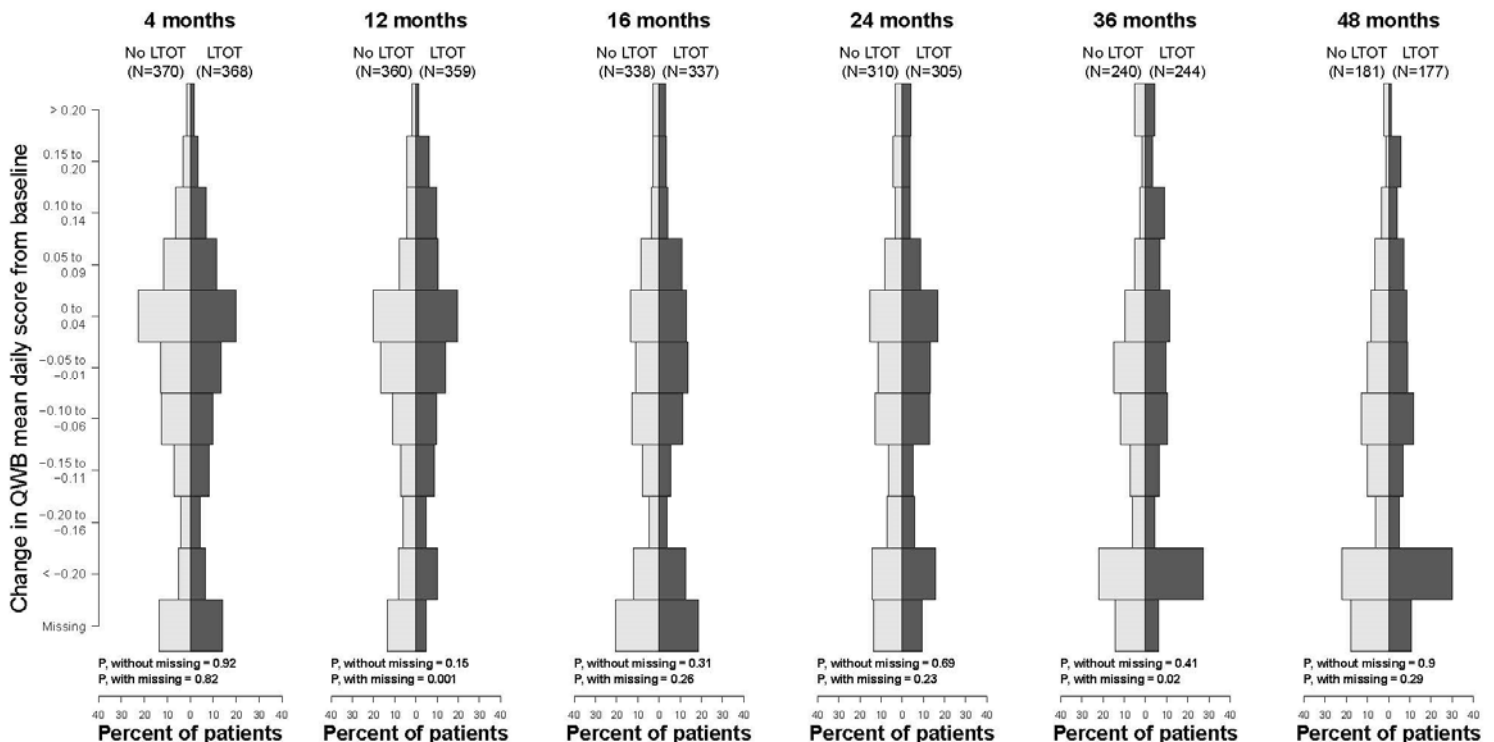
B) Change in St George's Respiratory Questionnaire (SGRQ) total score (range 0-100, higher score indicates worse better health-related quality of life, MID=4)[15,16]. P<0.008 (0.05/6 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
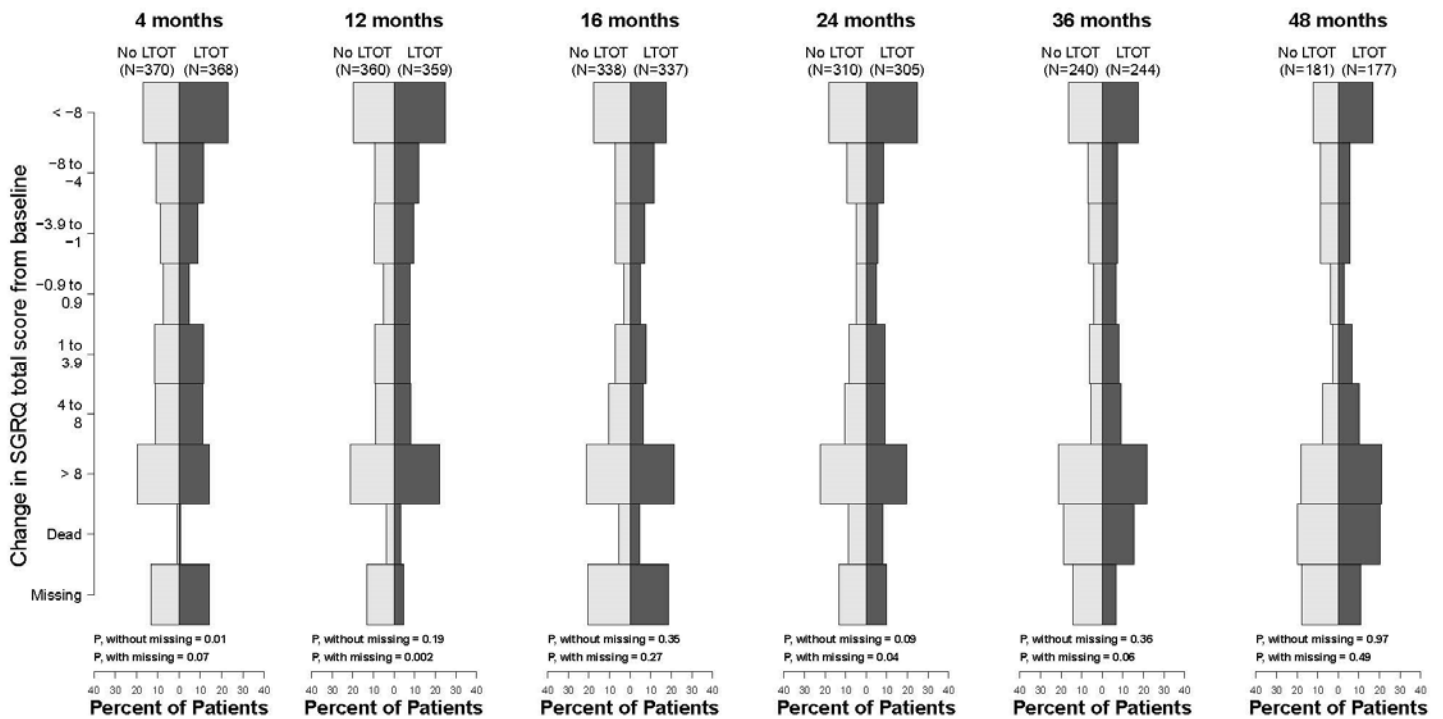
C) Change in SF-36 Physical Component Summary score (range 0-100, higher score indicates better function, MID=5)[17]. P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
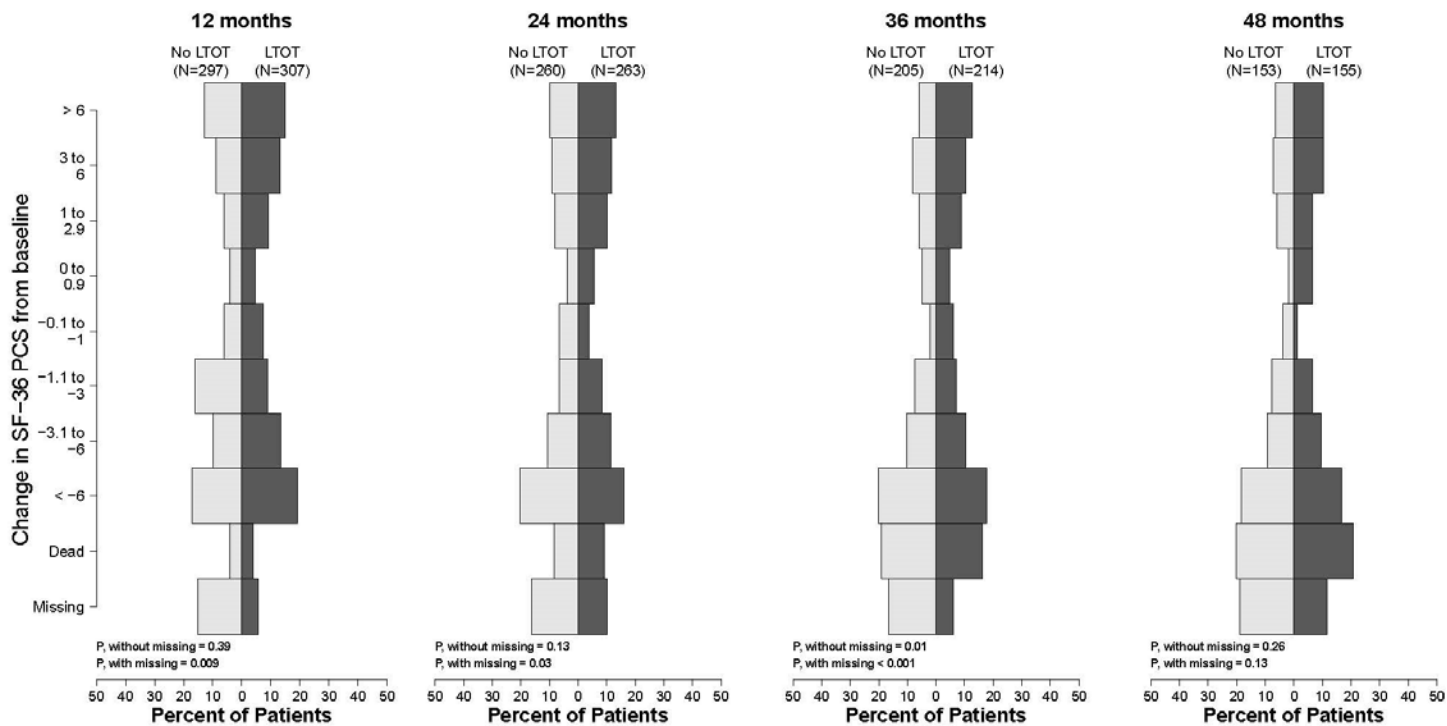
D) Change in SF-36 Mental Component Summary score (range 0-100, higher score indicates better function, MID=5)[17]. P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
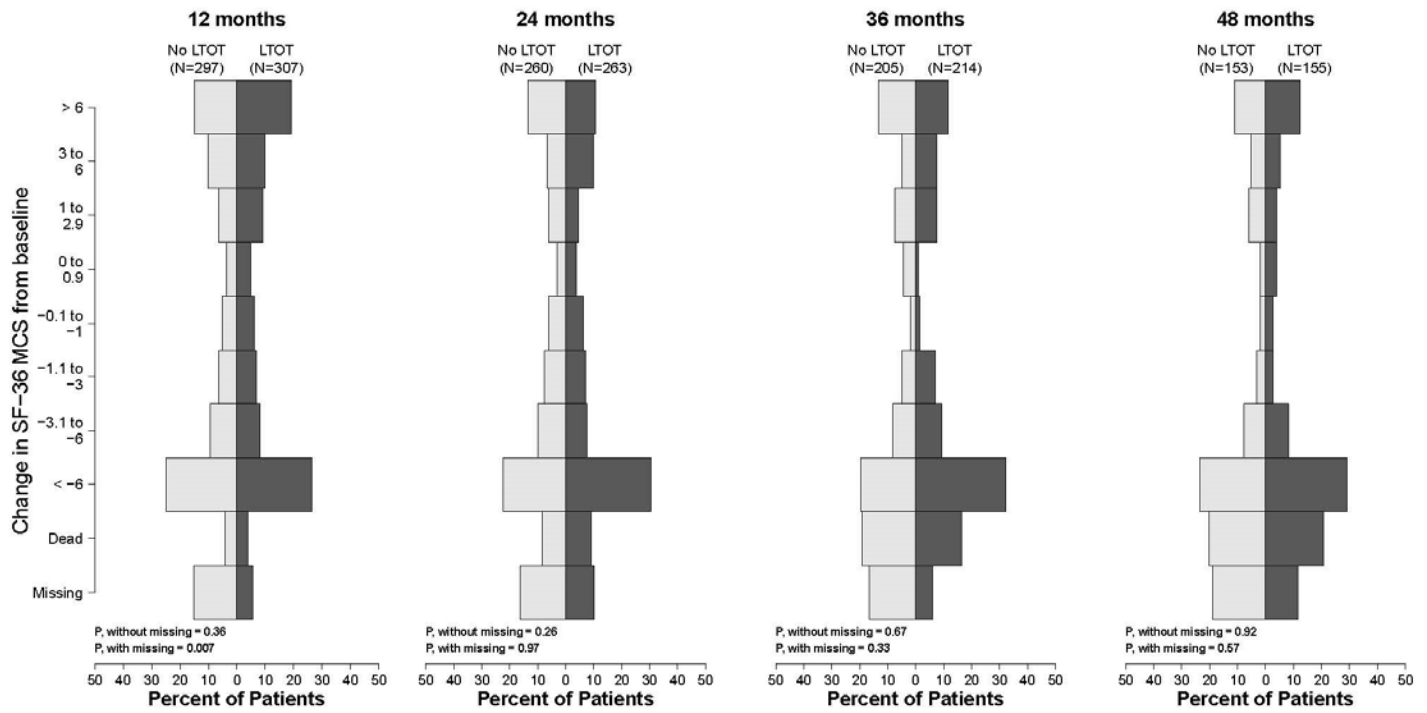
E) Change in Hospital Anxiety and Depression Scale (HADS) anxiety score (range 0-21, higher score indicates greater anxiety, MID=1.5)[18,19]. P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
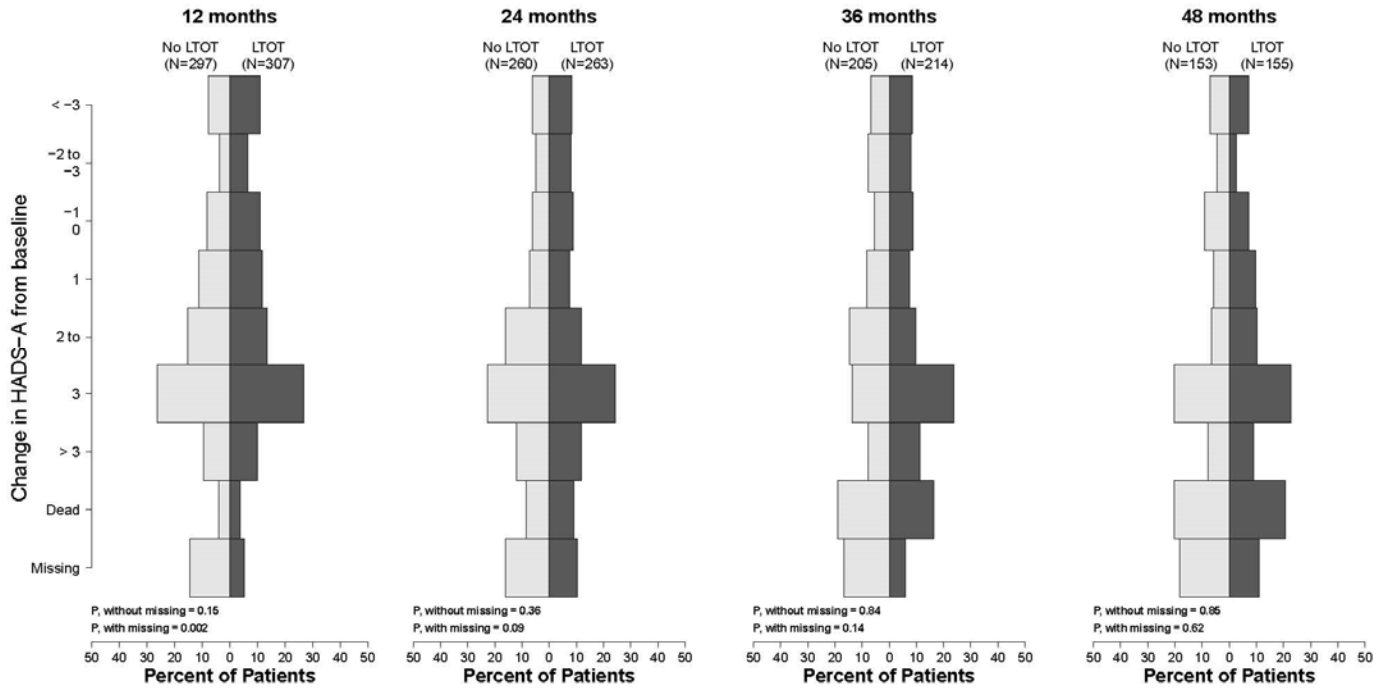
F) Change in Hospital Anxiety and Depression Scale (HADS) depression score (range 0-21, higher score indicates greater depression, MID=1.5)[18,19]. P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
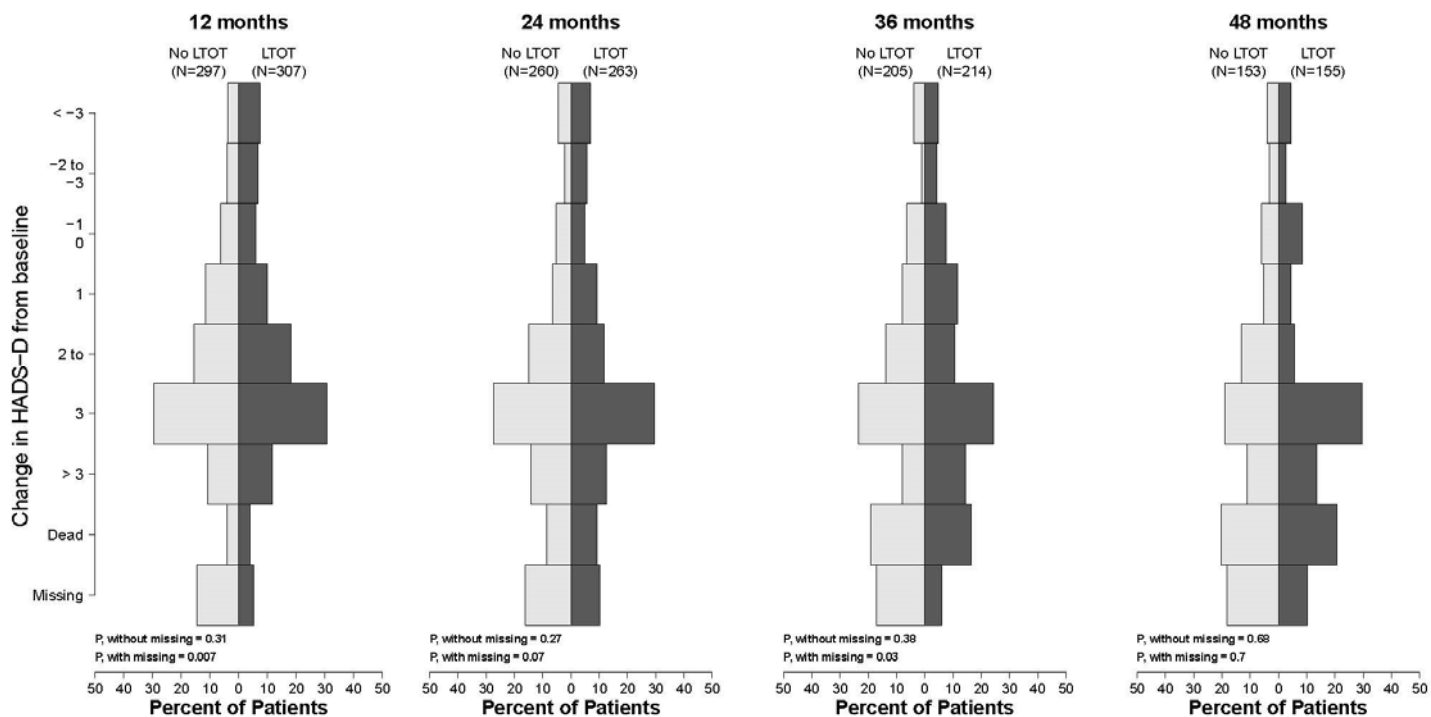
G) Change in Pittsburgh Sleep Quality Index (PSQI) total score (range 0-21, higher score indicates worse sleep quality)[20]. P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].

LOTT-PrimaryOucomePaper-Appendix.docx   10/4/2016 1:55 PM
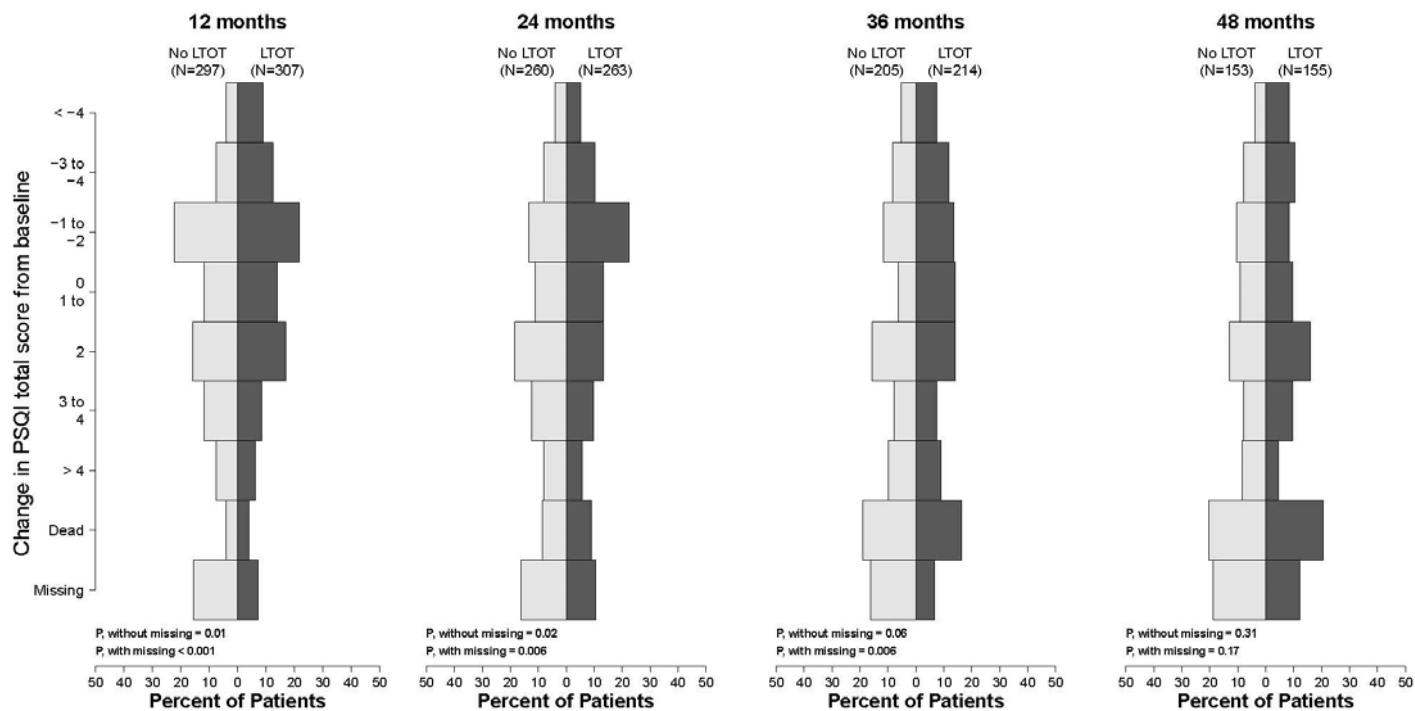
H) Change in post-bronchodilator FEV$_1$ (L). P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
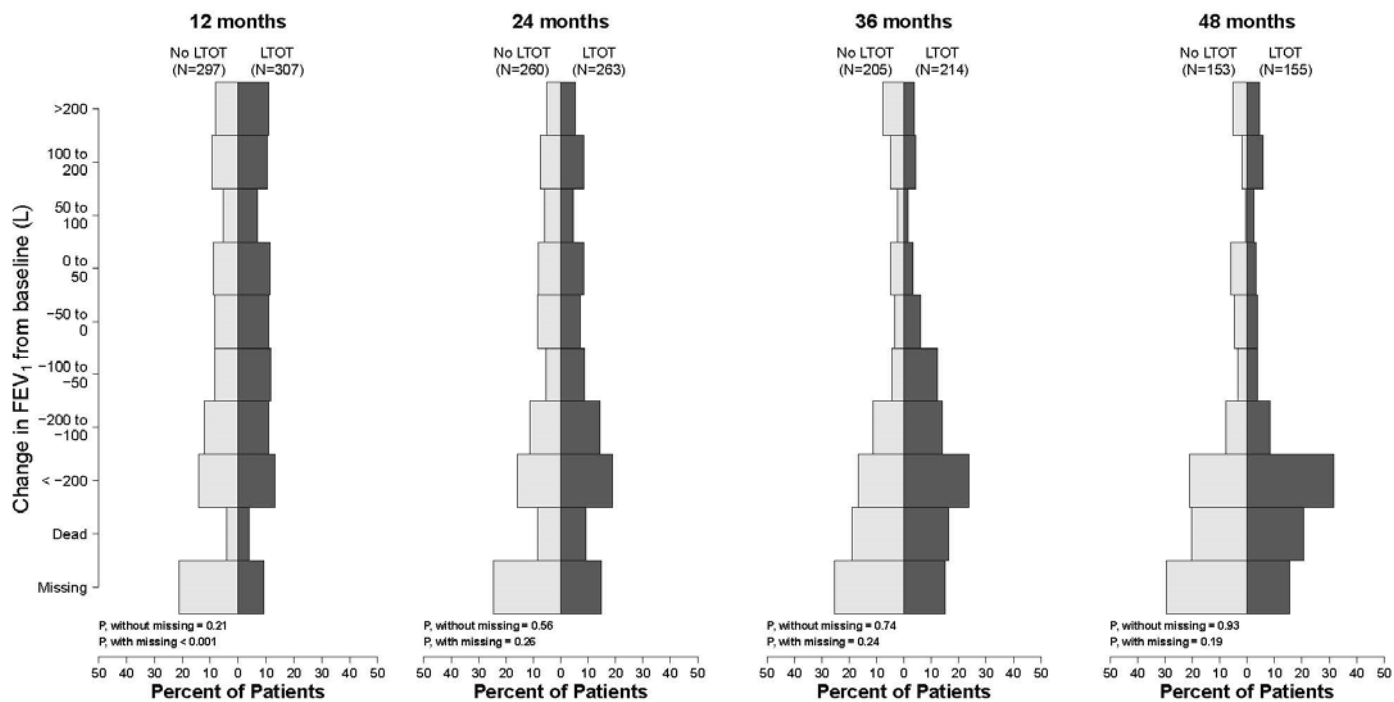
I) Change in room air 6-minute walk distance (feet). P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
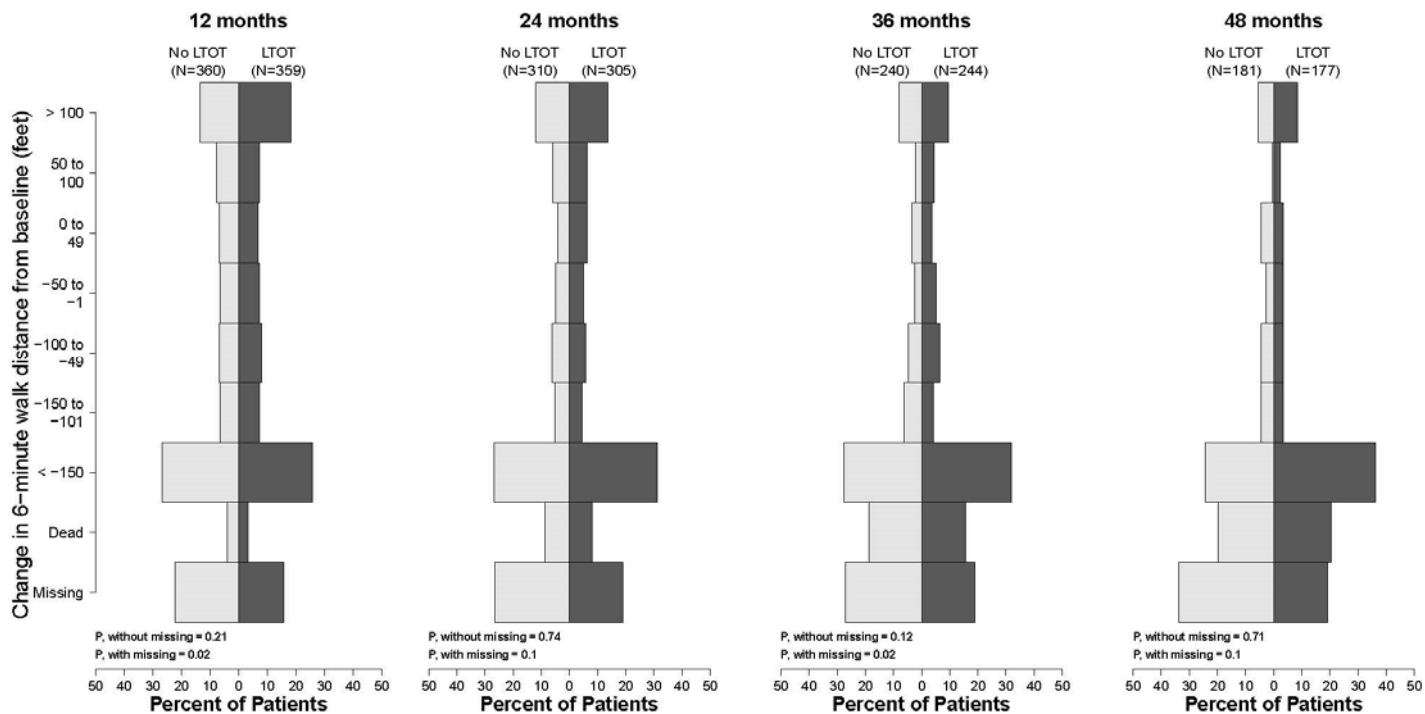
J) Change in room air resting oxygen saturation (SpO$_2$, %). P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].
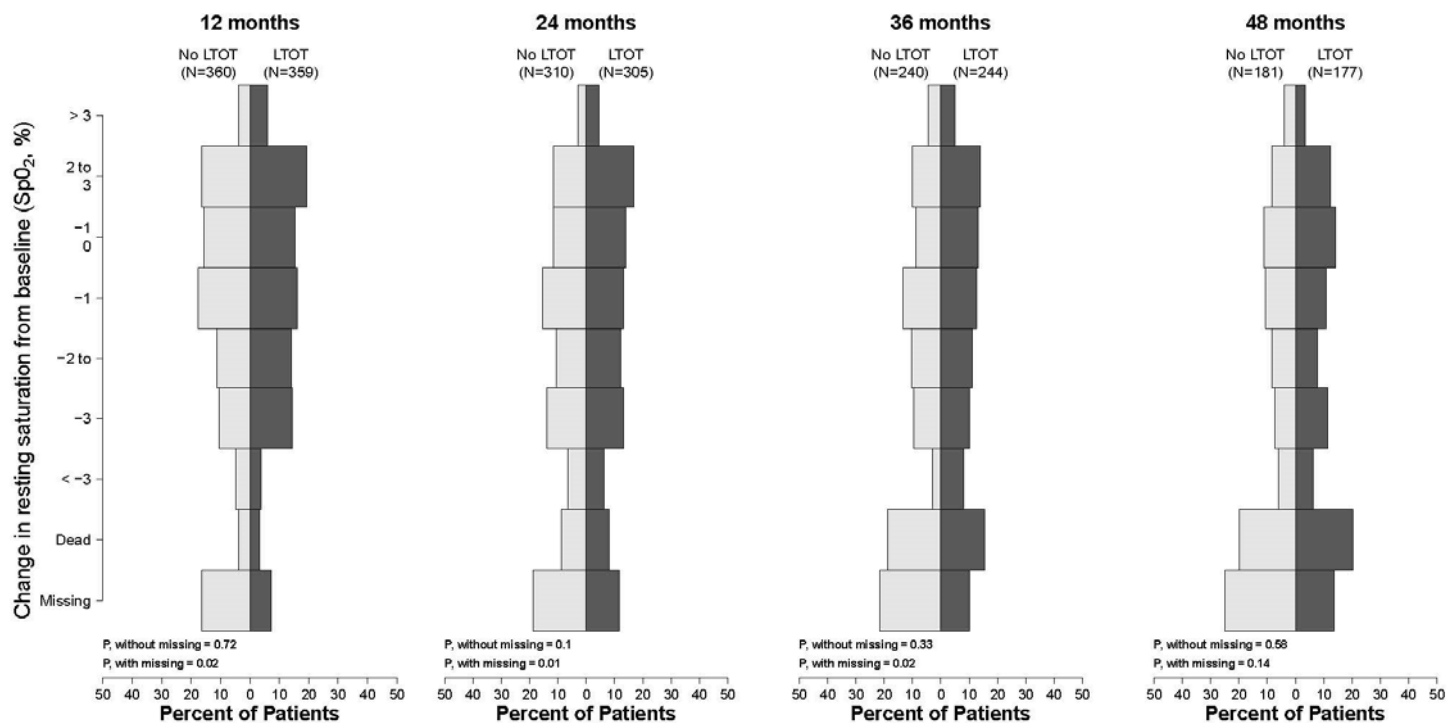
Table S1. Patient selection criteria

___

Inclusion (all must be met)
- COPD-dominated lung disease
- Age at least 40 years
- At least 10 pack-years cigarette smoking history
- Modified Medical Research Council (MMRC)* dyspnea score ≥ 1 (short of breath when hurrying on the level or walking up a slight hill)
- Post-bronchodilator FEV_1 / FVC < 0.70
- Post-bronchodilator FEV_1 ≤70% of the predicted normal value or > 70% of the predicted normal value and Study Physician determines that there is radiologic evidence of emphysema
- Resting SpO_2 89-93% (moderate resting hypoxemia) **OR** resting SpO_2 94% or greater and desaturation during exercise defined as SpO_2 below 90% for at least 10 seconds during the 6-minute walk test (normal resting saturation but hypoxemia with exercise)
- Medicare Part A and Part B beneficiary, insurance willing to pay costs of treatment and costs of study procedures and visits, or willing to self-pay costs
- Approval by study physician for randomization to either treatment group
- No exacerbation requiring antibiotics or new/ increased dose of systemic corticosteroids in the 30 days prior to screening
- At least 30 days post-discharge from an acute care hospital for COPD or other condition prior to screening
- If patient regularly uses supplemental oxygen prior to screening, all of the following must be met before randomization:
  - Patient agrees to stop using supplemental oxygen if randomized to no supplemental oxygen
  - Patient's physician agrees in writing to rescind order for supplemental oxygen if patient is randomized to no supplemental oxygen
  - Patient must not use supplemental oxygen for the 4 calendar days prior to randomization and must report that he/she had no problems doing without the oxygen
- Signature of written contract agreeing not to smoke while using supplemental oxygen

Exclusion (none may be met)
- COPD exacerbation requiring antibiotics, new or increased dose of systemic corticosteroids, or oxygen treatment after screening starts and prior to randomization (chronic use of corticosteroids while health is stable is not exclusionary)
- New prescription of supplemental oxygen after screening starts and before randomization
- Thoracic surgery or other procedure in the 6 months prior to evaluation likely to cause instability of pulmonary status
- Non-COPD lung disease that would affect oxygenation or survival
- Epworth Sleepiness Scale† score greater than 15
- Desaturation below 80% for at least 1 minute during the 6-minute walk
- Disease or condition expected to cause death or inability to perform procedures for the trial or inability to comply with therapy within 6 months of randomization, as judged by study physician
- Participation in another intervention study

___

COPD = chronic obstructive pulmonary disease; FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity.
*The Modified Medical Research Council (MMRC) dyspnea score is a single item scale that is completed by the patient; the score ranges from 0 to 4, with higher score indicating greater breathlessness[10,11]. MMRC=0 was exclusionary in LOTT.
†The Epworth Sleepiness Scale is an 8-item scale that measures general daytime sleepiness and is completed by the patient; the total score ranges from 0 to 24, with higher scores indicating greater daytime sleepiness[12]. Score ≥16 was exclusionary in LOTT.

Table S2. Data collection schedule

			Follow-up								
	BL	RZ		Year 1			Year 2			Year 3*	
Months from randomization	-2	0	4	8	12	16	20	24	28	32	36
Type of visit†	C	C	T,M	T	C	T,M	T	C	T	T	C
Core (all participants, all sites)											
History	X	X	X	X	X	X	X	X	X	X	X
Room air resting oximetry	X				X			X			X
Room air 6MW with oximetry	X				X			X			X
Ambulatory dosing‡		X			X			X			X
FEV$_1$, FVC	X										
Height	X										
Weight and pretibial pitting edema	X				X			X			X
Hemoglobin and hematocrit	X										
Cotinine (if not using nicotine products)	X				X						
MMRC dyspnea scale§	X				X			X			X
Epworth Sleepiness Scale¶	X										
Quality of Well-Being Scale‖	X		M		X	M		X			X
St. George's Respiratory Questionnaire**	X		M		X	M		X			X
Expanded (participants at selected sites)											
SF-36 quality of life scale††	X				X			X			X
Hospital Anxiety and Depression Scale‡‡	X				X			X			X
Pittsburgh Sleep Quality Index§§	X				X			X			X
FEV$_1$, FVC					X			X			X
Alpha 1-antitrypsin level and phenotype	X										

Notes: BL = baseline; RZ = randomization; C=clinic; T=telephone; M=mail; 6MW=6 minute walk; FEV$_1$ = forced expiratory volume in one second; FVC = forced vital capacity; MMRC = Modified Medical Research Council.

*Years 4, 5, and 6 follow the same pattern of data collection.

† Table does not show adherence contact schedule.

‡ For participants randomized to supplemental oxygen.

§The MMRC dyspnea scale is a single item scale that is completed by the patient; the score ranges from 0 to 4, with higher score indicating greater breathlessness[10,11]. MMRC=0 was exclusionary in LOTT.

¶The Epworth Sleepiness Scale is an 8-item scale that measures general daytime sleepiness and is completed by the patient; the total score ranges from 0 to 24, with higher scores indicating greater daytime sleepiness[12]. Score ≥16 was exclusionary in LOTT.

‖The Quality of Well Being scale is a 77-item quality of life questionnaire completed by the patient. The scoring is anchored in death (score 0). The average daily score ranges from 0 to 1, with higher scores indicating better quality of life; Minimum Important Difference (MID)=0.03[13,14].

**The St. George's Respiratory Questionnaire is a 51-item questionnaire on the health-related quality of life with regard to respiratory symptoms that is completed by the patient; the total score ranges from 0 to 100, with higher scores indicating worse health-related qualify of life; MID=4[15,16].

††The SF-36 is a 36-item quality of life scale that is completed by the patient. The Physical Component Summary (PCS) score ranges from 0 to 100, with higher scores indicating better function; MID=5. The Mental Component Summary (MCS) score ranges from 0 to 100, with higher scores indicating better function; MID=5[17].

‡‡The Hospital Anxiety and Depression Scale (HADS) is a 14-item scale that is completed by the patient. The anxiety score ranges from 0 to 21, with higher scores indicating greater anxiety; MID=1.5 and the depression score ranges from 0 to 21, with higher scores indicating greater depression; MID=1.5[18,19].

§§Pittsburgh Sleep Quality Index (PSQI) is a 24-item scale completed by the patient and partner. The total score ranges 0-21, with higher scores indicating worse sleep quality[20].

Table S3. Screening and randomization by clinical center

Regional Clinical Center	Clinical center name	No. screened	No. rz'd by site	No. rz'd by RCC (RCC site + associate satellites)
Brigham and Women's Hospital	Brigham and Women's Hospital	25	5	48
	Boston VA	75	25	
	Boston Medical Center	40	18	
Cleveland Clinic	Cleveland Clinic	58	33	42
	University Hospitals	4	1	
	Valley Care Health	2	0	
	Crouse Medical	5	4	
	Cleveland Clinic Weston, Florida	13	4	
Denver Health	Denver Health	0	0	6
	National Jewish	8	4	
	University of Colorado	4	2	
	University of Iowa	2	0	
Duke University	Duke University	45	28	28
Kaiser Foundation	Kaiser Foundation	67	9	9
Los Angeles Biomedical Research Institute	Los Angeles Biomedical Research Institute	21	8	55
	Loma Linda VA	98	44	
	City of Hope	7	3	
Ohio State University	Ohio State University	101	47	97
	Cincinnati VA	115	50	
	University of Kentucky	8	0	
Temple University	Temple University	108	63	139
	Geisinger Health System	42	11	
	Louisiana State University	6	2	
	Institute for Respiratory and Sleep Medicine	71	45	
	University of Maryland	8	5	
	Buffalo VA	27	8	
	Respiratory Specialists	13	5	
University of Alabama at Birmingham	University of Alabama at Birmingham	72	33	84
	Birmingham VA	79	35	
	North Florida/South Georgia VA	30	16	

Regional Clinical Center	Clinical center name	No. screened	No. rz'd by site	No. rz'd by RCC (RCC site + associate satellites)
University of Michigan	University of Michigan	75	22	43
	Beaumont Hospital	29	6	
	Henry Ford Hospital	6	2	
	Ann Arbor VA	2	0	
	San Antonio VA	37	13	
	Spectrum Health	2	0	
University of Pittsburgh	University of Pittsburgh	38	21	22
	Pulmonary Partners	1	1	
University of Utah	University of Utah	140	40	40
University of Washington	University of Washington	13	4	71
	Puget Sound VA	157	66	
	Harborview	1	1	
Washington University	Washington University	52	24	54
	Christian Hospital	15	8	
	University of Illinois at Chicago	4	2	
	Suburban Lung Associates	6	4	
	Central Florida Pulmonary Group	3	1	
	University of California, San Diego	24	15	
Total		**1759**	**738**	**738**

Table S4. Characteristics of randomized patients at enrollment

	No LTOT (N=370)	LTOT (N=368)
Demographic		
Age (years)	69.3 ± 7.4	68.3 ± 7.5
Male	276 (75%)	266 (72%)
Hispanic ethnicity	4 (1%)	10 (3%)
Race†		
African American	34 (9%)	46 (12%)
Caucasian	328 (89%)	311 (85%)
Other	11 (3%)	17 (5%)
Married or widowed	236 (64%)	234 (64%)
Medicare coverage	273 (74%)	268 (73%)
Enrolled at Veterans Administration site	129 (35%)	128 (35%)
Clinical		
Pack-years of tobacco cigarette smoking	60.8 ± 31.1	62.0 ± 34.7
Missing, n	4	0
Current tobacco cigarette smoker	92 (25%)	110 (30%)
Body mass index (kg/m^2)	28.3 ± 6.5	28.9 ± 6.5
Ever used home oxygen	113 (31%)	107 (29%)
Currently using home oxygen	59 (16%)	55 (15%)
COPD exacerbation in 3 months prior to screening	75 (20%)	63 (17%)
Hospitalized for COPD exacerbation in year prior to screening‡	28 (8%)	35 (11%)
Missing, n	39	41
Symptoms/health status		
Quality of Well-Being Scale mean daily score§	0.56 ± 0.13	0.56 ± 0.13
St. George's Respiratory Questionnaire total score¶	50.2 ± 17.1	49.8 ± 18.7
Modified Medical Research Council dyspnea score‖		
1	103 (28%)	97 (26%)
2	101 (27%)	114 (31%)
3	132 (36%)	136 (37%)
4	34 (9%)	21 (6%)

	No LTOT (N=370)	LTOT (N=368)
Epworth Sleepiness Scale total score**		
0-5	173 (47%)	157 (43%)
6-10	143 (39%)	137 (37%)
11-15	53 (14%)	72 (20%)
> 15	1 (<1%)	2 (1%)
Physiology		
Oxygen desaturation type qualifying patient for enrollment		
Resting only	60 (16%)	73 (20%)
Exercise only	171 (46%)	148 (40%)
Resting and exercise	139 (38%)	147 (40%)
Room air SpO_2 at rest (%)		
All	93.5 ± 1.9	93.3 ± 2.1
Resting only	92.3 ± 0.8	92.4 ± 0.9
Exercise only	95.2 ± 1.2	95.4 ± 1.4
Resting and exercise	91.9 ± 1.2	91.7 ± 1.1
Room air nadir SpO_2 during 6-minute walk (%)††		
< 86	85 (29%)	86 (29%)
86-88	103 (36%)	105 (36%)
> 88	102 (35%)	101 (35%)
Missing, n††	80	76
GOLD lung function category		
$FEV_1/FVC \geq 0.7$	4 (1%)	4 (1%)
I	12 (3%)	12 (3%)
II	110 (30%)	128 (35%)
III	180 (49%)	176 (48%)
IV	64 (17%)	47 (13%)
Missing, n	0	1
Post bronchodilator FEV_1 percent predicted (%)	46 ± 17	47 ± 16
Missing, n	0	1
Post bronchodilator FVC percent predicted (%)	75 ± 19	77 ± 20
Missing, n	0	1
Post bronchodilator FEV_1/FVC ratio	0.46 ± 0.12	0.47 ± 0.13
Missing, n	0	1
Distance walked in 6 minutes (feet)‡‡	1027 ± 337	1062 ± 313
Missing, n	4	6

	No LTOT (N=370)	LTOT (N=368)
Co-morbid conditions (self-reported)		
Anemia	56 (15%)	64 (17%)
Cardiovascular disease§§	114 (31%)	121 (33%)
Depression	128 (35%)	126 (34%)
GERD or stomach ulcer	160 (43%)	163 (44%)
Hypertension	233 (63%)	221 (60%)
Sleep apnea	82 (22%)	83 (23%)
Using continuous positive airway pressure device for sleep apnea	52 (63%)	59 (71%)
BODE¶¶		
0-2	77 (21%)	87 (24%)
3-4	123 (33%)	155 (42%)
5-6	126 (34%)	87 (24%)
7-10	40 (11%)	33 (9%)
Mean ± SD	4.2 ± 2.0	3.9 ± 1.8
Missing, n	4	6

*Plus-minus values are observed mean ± SD. The two treatment groups were similar on characteristics except for the BODE index, which tended to be lower in the LTOT group compared to the No LTOT group ((P=0.007 for difference in categorized distribution; P=0.02 for difference in means). GOLD denotes Global initiative for chronic Obstructive Lung Disease; FEV_1 denotes forced expiratory volume in 1 second; FVC denotes forced vital capacity.

†Patient may select more than one race group.

§The Quality of Well Being scale is a 77-item quality of life questionnaire completed by the patient. The scoring is anchored in death (score 0). The average daily score ranges from 0 to 1, with higher scores indicating better quality of life. Minimum Important Difference (MID)=0.03[13,14]

¶The St. George's Respiratory Questionnaire is a 51-item questionnaire on the health-related quality of life with regard to respiratory symptoms that is completed by the patient; the total score ranges from 0 to 100, with higher scores indicating worse health-related qualify of life. MID=4[15,16].

ⅠThe Modified Medical Research Council Dyspnea Score is a single item scale that is completed by the patient; the score ranges from 0 to 4, with higher score indicating greater breathlessness[10,11]. MMRC=0 was exclusionary in LOTT.

**The Epworth Sleepiness Scale is an 8-item scale that measures general daytime sleepiness and is completed by the patient; the total score ranges from 0 to 24, with higher scores indicating greater daytime sleepiness[12]. Score ≥16 was exclusionary in LOTT.

††Nadir SpO_2 is the 10th lowest SpO_2 observed during the 6-minute walk. 10 patients did not complete the 6-minute walk; nadir SpO_2 could not be calculated for 146 patients due to loss or a technical issue with their enrollment oximetry data file.

‡‡10 patients did not complete the 6-minute walk at baseline due to leg amputation (2), arthritis, neuropathy, use of a wheelchair (3), peripheral artery disease, sciatica, and shortness of breath. To convert feet to meters, divide by 3.28.

§§Cardiovascular disease includes angina, coronary artery disease, coronary artery revascularization, myocardial infarction, heart failure, and congestive heart failure

¶¶ The BODE index is a scoring system incorporating body mass index, airflow obstruction (post bronchodilator FEV_1 percent predicted), dyspnea (MMRC score), and exercise (6-minute walk distance); higher score indicates greater risk of mortality[21].

Table S5. Supplemental oxygen flow setting prescribed to LTOT patients (N=368) for ambulation.

Supplemental oxygen flow setting	At RZ	12 months	24 months	36 months	48 months
Median	2	2	2	2	2
(25th %-tile, 75th %-tile)	(2, 2)	(2, 2)	(2, 3)	(2, 3)	(2, 3)
(10th %-tile, 90th %-tile)	(2, 3)	(2, 4)	(2, 4)	(2, 4)	(2, 4)
No. of patients with data	357	286	218	165	105

Table S6. Primary outcome (time to death or first hospitalization, whichever comes first) for those randomized to LTOT versus those randomized to No LTOT in subgroups of patients defined at baseline. This is an intention-to-treat analysis. P<0.0007 (0.01/14) was considered statistically significant using a Bonferroni adjustment for multiple comparisons[9].

	N	Events	Rate/ 100 p-y	HR (95% CI)	P*	Interaction P†
All patients						
No LTOT	370	250	36.4	1.0		
LTOT	368	248	34.2	0.94 (0.79, 1.12)	0.52	
LTOT prescription type						
No LTOT	370	250	36.4	1.0		
LTOT during sleep and exercise	148	102	37.9	1.05 (0.83, 1.32)	0.68	0.18
24-hour LTOT	220	146	32.1	0.88 (0.72, 1.08)	0.23	
Desaturation qualifying for LOTT						
At rest only						
No LTOT	60	38	34.4	1.0		
LTOT	73	50	32.6	0.96 (0.63, 1.47)	0.86	
On exercise only						
No LTOT	171	119	39.3	1.0		0.99
LTOT	148	102	37.9	0.95 (0.73, 1.24)	0.71	
At rest and on exercise						
No LTOT	139	93	34.0	1.0		
LTOT	147	96	31.8	0.95 (0.72, 1.27)	0.74	
Age (years)						
65-70						
No LTOT	211	132	31.7	1.0		
LTOT	238	162	35.3	1.11 (0.88, 1.40)	0.38	0.03
71 or older						
No LTOT	159	118	43.6	1.0		
LTOT	130	86	32.4	0.75 (0.57, 0.99)	0.04	
Race						
Minority						
No LTOT	41	30	44.2	1.0		
LTOT	55	41	38.8	0.86 (0.53, 1.37)	0.52	0.77
White						
No LTOT	328	219	35.5	1.0		
LTOT	311	206	33.6	0.95 (0.78, 1.15)	0.58	
Gender						
Male						
No LTOT	276	190	39.1	1.0		
LTOT	266	178	33.9	0.87 (0.71, 1.07)	0.19	0.15
Female						
No LTOT	94	60	29.9	1.0		
LTOT	102	70	35.1	1.15 (0.82, 1.63)	0.42	

	N	Events	Rate/ 100 p-y	HR (95% CI)	P*	Interaction P†
Current cigarette smoker						
Yes						
No LTOT	92	64	39.9	1.0		
LTOT	110	77	38.6	0.96 (0.69, 1.33)	0.79	0.81
No						
No LTOT	278	186	35.4	1.0		
LTOT	258	171	32.6	0.93 (0.75, 1.14)	0.47	
COPD exacerbation in 3 months prior to enrollment						
Yes						
No LTOT	75	57	51.1	1.0		
LTOT	63	38	28.9	0.58 (0.39, 0.88)	0.01	0.007
No						
No LTOT	295	193	33.6	1.0		
LTOT	305	210	35.4	1.07 (0.88, 1.30)	0.52	
Nadir SpO_2 during 6-minute walk‡						
< 86%						
No LTOT	85	53	31.8	1.0		
LTOT	86	53	33.5	1.10 (0.75, 1.63)	0.62	
86% - 88%						
No LTOT	103	69	37.9	1.0		0.62
LTOT	105	69	35.0	0.92 (0.66, 1.28)	0.62	
> 88%						
No LTOT	102	71	42.4	1.0		
LTOT	101	70	36.6	0.88 (0.63, 1.23)	0.45	
Pre bronchodilator FEV_1						
< 41% predicted						
No LTOT	168	115	39.2	1.0		
LTOT	169	121	35.9	0.93 (0.72, 1.20)	0.55	0.60
≥ 41 predicted						
No LTOT	162	107	32.4	1.0		
LTOT	179	114	32.9	1.00 (0.77, 1.31)	0.97	
BODE score§						
0-4						
No LTOT	200	122	28.8	1.0		
LTOT	242	153	31.0	1.08 (0.85, 1.37)	0.53	0.16
5-10						
No LTOT	166	127	49.4	1.0		
LTOT	120	91	41.3	0.84 (0.64, 1.11)	0.22	

	N	Events	Rate/ 100 p-y	HR (95% CI)	P*	Interaction P†
QWB mean daily score¶						
< 0.55 (below median)						
No LTOT	185	132	44.2	1.0		
LTOT	177	120	33.9	0.77 (0.60, 0.99)	0.04	
≥ 0.55 (at or above median)						0.03
No LTOT	185	118	30.4	1.0		
LTOT	191	128	34.6	1.15 (0.89, 1.48)	0.28	
SF-36 PCS score§						
< 33 (below median)						
No LTOT	148	98	39.0	1.0		
LTOT	158	108	34.7	0.90 (0.68, 1.18)	0.45	
≥ 33 (at or above median)						0.42
No LTOT	154	104	32.4	1.0		
LTOT	156	105	34.1	1.08 (0.82, 1.42)	0.58	
Body mass index (kg/m^2)						
< 25.1						
No LTOT	135	98	43.1	1.0		
LTOT	109	77	35.1	0.82 (0.61, 1.11)	0.21	
25.1 – 30.8						
No LTOT	116	72	30.1	1.0		0.94
LTOT	133	92	38.5	1.28 (0.94, 1.75)	0.12	
>30.8						
No LTOT	119	80	36.3	1.0		
LTOT	126	79	29.7	0.82 (0.60, 1.12)	0.21	
History of anemia						
Yes						
No LTOT	56	42	41.3	1.0		
LTOT	64	46	42.0	1.0 (0.66, 1.53)	>0.99	
No						0.70
No LTOT	314	208	35.6	1.0		
LTOT	304	202	32.9	0.93 (0.76, 1.12)	0.43	

*Logrank test for difference in the primary outcome between groups within the specified stratum.

†P-value for Wald chi square test of interaction between treatment group and subgrouping factor on the primary outcome, except for oxygen prescription groups, where the P-value tests for difference between patients prescribed oxygen during sleep and exercise versus those prescribed 24-hour oxygen.

‡Nadir SpO$_2$ is the 10th lowest SpO$_2$ observed during the 6-minute walk. Data from oximetry during the 6-minute walk could not be analyzed for 146 patients; 10 patients did not complete the 6-minute walk.

§The BODE index is a scoring system incorporating body mass index, airflow obstruction (post bronchodilator FEV$_1$ percent predicted), dyspnea (MMRC score), and exercise (6-minute walk distance); higher score indicates greater risk of mortality[21].

¶The Quality of Well Being scale is a 77-item quality of life questionnaire completed by the patient. The scoring is anchored in death (score 0). The average daily score ranges from 0 to 1, with higher scores indicating better quality of life. Minimum Important Difference (MID)=0.03[13,14].

§ The SF-36 is a 36-item quality of life scale that is completed by the patient. The Physical Component Summary (PCS) score ranges from 0 to 100, with higher scores indicating better function; MID=5[17].

Table S7. Time to death or first hospitalization, whichever occurred first, for those reporting supplemental oxygen use versus those reporting not using supplemental oxygen. This is an as treated analysis. P<0.025 (0.05/2) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].

	Not using supplemental oxygen	Using supplemental oxygen
Definition 1:		
Using supplemental oxygen includes patients:		
- Randomized to 24-hour oxygen and reporting use of **at least 16 hours** per day averaged over follow-up		
- Randomized to oxygen during sleep and exercise and reporting use of **at least 8 hours** per day averaged over follow-up		
- Randomized to no oxygen and reporting use of **at least 8 hours** per day averaged over follow-up		
Not using includes all other patients		
No. of patients	496	242
No. of events	328	170
Rate per 100 person-years	35.2	35.6
Hazard ratio (Using versus Not using)	1.02	
(95% CI)	(0.85, 1.23)	
P*	0.84	
Definition 2:		
Using includes patients:		
- Randomized to 24-hour oxygen and reporting use of **at least 16 hours** per day averaged over follow-up		
- Randomized to oxygen during sleep and exercise and reporting use of **at least 16 hours** per day averaged over follow-up		
- Randomized to no oxygen and reporting use of **at least 16 hours** per day averaged over follow-up		
Not using includes all other patients		
No. of patients	604	134
No. of events	405	93
Rate per 100 person-years	35.1	36.1
Hazard ratio (Using vs. Not using)	1.03	
(95% CI)	(0.82, 1.29)	
P*	0.80	

Notes:
    CI denotes confidence interval.
    For composite event analysis, patients who have neither death nor hospitalization are censored as of their last interview.
    *Logrank test

Table S8. Comparison of LOTT design assumptions to observed data

	Assumed	Observed (95% CI)
Crossover rates		
Overall % of No LTOT group receiving oxygen treatment	11.7%*	8.7% (6.6, 11.2)
Overall % of LTOT group stopping oxygen treatment	3.1%*	2.2% (1.3, 3.6)
Event rates		
Composite event rate in No LTOT group	28%	36.4% (32.0, 41.2)
Hospitalizations/yr in those with recent COPD exacerbation	33%	73.1 (46.8, 108.7)
Hospitalization/yr in those without recent COPD exacerbation	10%	32.2 (27.7, 37.3)
Mortality in those with a recent COPD exacerbation	7%	5.3% (1.7, 12.2)
Mortality in those without a recent COPD exacerbation	6%	5.6% (4.2, 7.3)
Population composition at trial entry		
Patients with recent COPD hospitalization	50%	9.6%

*By 2012, it had become evident that the treatment group crossover rates assumed when the trial was designed (originally 21%, No LTOT receiving oxygen treatment; 50% LTOT stopping oxygen treatment) were much greater than the observed crossover rates; therefore, in March 2012, the LOTT DSMB approved a revised sample size calculation of 737 patients based on the crossover rates observed to that date, 11.7%, No LTOT receiving oxygen treatment and 3.1%, LTOT group stopping oxygen treatment.

Table S9. Frequency and rate of hospitalization, COPD exacerbation, hospitalization for COPD exacerbation, and hospitalization for other than COPD exacerbation. This is an intention-to-treat analysis. P<0.0125 (0.05/4) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].

	No LTOT (N=370)	LTOT (N=368)	P*
All cause hospitalization			
Patients with no hospitalization	36%	38%	0.77
Patients with exactly 1 hospitalization	24%	22%	
Patients with 2 or more hospitalizations	40%	40%	
Total number of hospitalizations	651	685	
Rate of hospitalization per 100 person-years	56.2	56.9	
Rate ratio	1.01		0.81
(95% CI)	(0.91, 1.13)		
COPD exacerbation			
Patients with no COPD exacerbation	35%	34%	0.91
Patients with exactly 1 COPD exacerbation	22%	22%	
Patients with 2 or more COPD exacerbations	43%	45%	
Total number of COPD exacerbations	785	880	
Rate of COPD exacerbation per 100 person-years	67.7	73.1	
Rate ratio	1.08		0.12
(95% CI)	(0.98, 1.19)		
Hospitalization for COPD exacerbation			
Patients without hospitalization for COPD exacerbation	64%	66%	0.83
Patients with exactly 1 hospitalization for COPD exacerbation	21%	20%	
Patients with 2 or more hospitalizations for COPD exacerbation	14%	14%	
Total number of hospitalizations for COPD exacerbation	259	265	
Rate of hospitalization for COPD exacerbation	22.4	22.0	
Rate ratio	0.99		0.86
(95% CI)	(0.83, 1.17)		
Hospitalization for other than COPD exacerbation			
Patients without hospitalization for other than COPD exacerbation	48%	49%	0.24
Patients with exactly 1 hospitalization for other than COPD exacerbation	26%	21%	
Patients with 2 or more hospitalizations for other than COPD exacerbation	25%	29%	
Total number of hospitalizations for other than COPD exacerbation	392	420	
Rate of hospitalization for other than COPD exacerbation	33.8	34.9	
Rate ratio	1.03		0.66
(95% CI)	(0.90, 1.18)		

Notes:

CI denotes confidence interval.

For time to event analysis, living patients who do not have the event are censored as of their last interview; deceased patients who did not have the event prior to death are censored as of their date of death.

*Fisher's exact test for difference in frequency distribution or difference in rate ratio.

Table S10. Changes from baseline to 4, 12, 16, 24, 36, and 48 months after randomization in Quality of Well-Being Scale (QWB) mean daily score, St. George's Respiratory Questionnaire (SGRQ) total score, SF-36 Physical Component Summary (PCS) score, SF-36 Mental Component Summary (MCS) score, Hospital Anxiety and Depression Questionnaire (HADS) anxiety score, HADS depression score, Pittsburgh Sleep Quality Index (PSQI) total score, post-bronchodilator FEV_1 (L), room air 6-minute walk distance (feet), and room air resting oxygen saturation (SpO_2, %), and development of severe resting and exercise desaturation during follow-up, those randomized to LTOT group versus those randomized to No LTOT. Analyses are limited to patients whose visit window for the specified follow-up time had closed as of the end of the trial. Changes were calculated by subtracting the baseline value from the value at the specified follow-up time. For the QWB, patients who died were assigned a score of 0 on the questionnaire for that follow-up time. This is an intention-to-treat analysis.

	Months after randomization											
	4		12		16		24		36		48	
	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT
Quality of Well-Being Scale mean daily score*												
No. of patients												
Missed and alive	50	52	48	18	69	63	42	29	34	15	32	19
In analysis	320	316	312	341	269	274	268	276	206	229	149	158
BL, mean	0.56	0.56	0.56	0.55	0.56	0.56	0.56	0.56	0.57	0.56	0.57	0.56
Chg. from BL, mean	-0.02	-0.02	-0.04	-0.04	-0.07	-0.06	-0.07	-0.07	-0.13	-0.12	-0.15	-0.16
P†, within group	0.02	0.01	<0.001	0.01	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
P†, between groups‡	0.92		0.73		0.77		0.98		0.68		0.57	
St. George's Respiratory Questionnaire, total score§												
No. of patients												
Missed or dead	52	54	61	28	86	79	67	53	78	51	67	55
In analysis	318	314	299	331	252	258	243	252	162	193	114	122
BL, mean	49.5	49.1	48.3	49.7	48.6	48.8	48.4	48.8	46.9	49.0	46.5	50.0
Chg. from BL, mean	0.6	-2.2	0.6	-1.1	1.4	0.2	1.2	-0.5	0.9	0.3	1.8	1.5
P†, within group	0.33	0.007	0.66	0.12	0.14	0.76	0.14	0.43	0.33	0.33	0.34	0.20
P†, between groups‡	0.003		0.21		0.35		0.25		0.88		0.58	

	Months after randomization											
	4		12		16		24		36		48	
	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT
SF-36, Physical Component Summary score¶												
No. of patients‖												
Missed or dead			45	17			47	31	43	21	54	36
In analysis			241	279			197	214	133	169	94	105
BL, mean			34.6	33.8			34.8	34.0	36.2	34.0	36.4	33.3
Chg. from BL, mean			-1.0	-0.5			-1.7	-0.5	-2.3	-1.0	-3.1	-1.2
P† within group			0.03	0.30			0.46	0.46	<0.001	0.22	0.17	0.17
P†, between groups**			0.71				0.17		0.57		0.64	
SF-36, Mental Component Summary score¶												
No. of patients‖												
Missed or dead			45	17			47	31	43	21	54	36
In analysis			241	279			197	214	133	169	94	105
BL, mean			51.7	50.2			52.3	51.1	52.5	51.2	52.1	51.8
Chg. from BL, mean			-1.7	-1.3			-2.8	-3.7	-1.4	-4.9	-1.8	-5.2
P†, within group			0.02	0.12			<0.001	<0.001	0.09	<0.001	0.02	<0.001
P†, between groups**			0.97				0.24		0.001		0.04	

	Months after randomization											
	4		12		16		24		36		48	
	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT
HADS anxiety score††												
No. of patients‖												
Missed or dead			43	16			47	31	43	21	53	35
In analysis			243	280			197	214	133	169	94	106
BL, mean			5.1	6.2			5.0	6.0	5.1	5.9	5.1	5.7
Chg. from BL, mean			0.3	-0.1			0.7	0.2	-0.1	0.2	0.2	0.5
P†, within group			0.20	0.36			0.02	0.62	0.27	0.42	0.99	0.17
P†, between groups**			0.28				0.07		0.69		0.94	
HADS depression score††												
No. of patients‖												
Missed or dead			43	16			47	31	44	21	53	34
In analysis			243	280			197	214	132	169	94	107
BL, mean			4.9	5.4			4.6	5.1	4.3	4.9	4.6	5.1
Chg. from BL, mean			0.7	0.4			1.1	0.6	0.7	1.0	0.8	1.2
P†, within group			<0.001	0.02			<0.001	0.005	0.01	<0.001	0.02	<0.001
P†, between groups**			0.41				0.39		0.23		0.16	

| | Months after randomization | | | | | | | | | | | |
| | 4 | | 12 | | 16 | | 24 | | 36 | | 48 | |
	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT
Pittsburgh Sleep Quality Index total score‡‡												
No. of patients‖												
Missed or dead			46	22			47	32	42	22	54	37
In analysis			240	274			197	213	134	168	93	104
BL, mean			7.4	8.0			7.3	7.9	7.0	7.8	6.8	7.7
Chg. from BL, mean			0.4	-0.3			0.6	-0.1	0.4	0.1	0.5	-0.4
P†, within group			0.09	0.05			0.53	0.53	0.23	0.98	0.20	0.47
P†, between groups**			0.03				0.02		0.87		0.33	
Post-bronchodilator FEV$_1$ (L)												
No. of patients‖												
Missed or dead			63	28			69	43	61	40	70	42
In analysis			223	268			174	202	115	149	77	99
BL, mean			1.4	1.3			1.4	1.3	1.4	1.3	1.4	1.3
Chg. from BL, mean			-0.024	-0.012			-0.073	-0.079	-0.086	-0.128	-0.100	-0.153
P†, within group			0.009	0.08			<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
P†, between groups**			0.93				0.40		0.04		0.05	

	Months after randomization											
	4		12		16		24		36		48	
	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT
Room air 6-minute walk distance (feet)												
No. of patients												
Missed or dead			93	68			109	82	109	82	97	70
In analysis			267	291			201	223	131	162	84	107
BL, mean			1057	1075			1074	1101	1128	1126	1119	1150
Chg. from BL, mean			-85	-53			-114	-115	-156	-148	-175	-210
P†, within group			<0.001	0.001			<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
P†, between groups**			0.37				0.78		0.68		0.17	
Room air resting oxygen saturation (%)												
No. of patients												
Missed or dead			72	37			85	59	96	60	81	60
In analysis			288	322			225	246	144	184	100	117
BL, mean			93.5	93.3			93.5	93.3	93.3	93.5	93.5	93.4
Chg. from BL, mean			0.2	0.2			-0.3	0.0	0.0	-0.2	-0.1	-0.1
P†, within group			0.19	0.05			0.03	0.92	0.95	0.32	0.62	0.56
P†, between groups**			0.85				0.09		0.76		0.90	

	Months after randomization											
	4		12		16		24		36		48	
	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT	No LTOT	LTOT
Development of severe resting desaturation§§												
No. of patients												
Missed or dead			72	37			85	59	96	60	81	60
In analysis			288	322			225	246	144	184	100	117
No. (%) with severe resting desaturation§§			3 (1%)	7 (2%)			9 (4%)	9 (4%)	6 (4%)	11 (8%)	4 (4%)	4 (4%)
RR for developing severe resting desaturation§§			2.09				0.91		1.43		0.85	
95% CI			(0.54, 8.00)				(0.37, 2.26)		(0.54, 3.79)		(0.22, 3.33)	
P¶¶, **			0.35				>0.99		0.62		>0.99	
Development of severe exercise desaturation‖‖												
No. of patients												
Missed or dead			80	48			95	70	102	75	90	64
In analysis			280	311			215	235	138	169	91	113
No. (%) with severe exercise desaturation‖‖			14 (5%)	12 (4%)			13 (6%)	14 (7%)	8 (6%)	11 (8%)	17 (19%)	10 (11%)
RR for developing severe exercise desaturation‖‖			0.77				0.99		1.12		0.47	
95% CI			(0.36, 1.64)				(0.47, 2.05)		(0.46, 2.71)		(0.23, 0.98)	
P¶¶, **			0.55				>0.99		>0.99		0.06	

BL = baseline; Chg. = change; RR = relative risk, LTOT versus No LTOT; CI = confidence interval; HADS = Hospital Anxiety and Depression Scale; FEV_1 = forced expiratory volume in 1 second

*The Quality of Well Being scale is a 77-item quality of life questionnaire completed by the patient. The scoring is anchored in death (score 0). The average daily score ranges from 0 to 1, with higher scores indicating better quality of life; Minimum Important Difference (MID)=0.03[13,14].

†P-value for t-test

‡P<0.008 (0.05/6 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].

§The St. George's Respiratory Questionnaire is a 51-item questionnaire on the health-related quality of life with regard to respiratory symptoms that is completed by the patient; the total score ranges from 0 to 100, with higher scores indicating worse health-related qualify of life; MID=4[15,16].

¶The SF-36 is a 36-item quality of life scale that is completed by the patient. The Physical Component Summary (PCS) score ranges from 0 to 100, with higher scores indicating better function; MID=5. The Mental Component Summary (MCS) score ranges from 0 to 100, with higher scores indicating better function; MID=5[17].

‖The SF-36, HADS, and PSQI questionnaires at all time points and spirometry in follow-up were completed at a subset of 33 sites; 82% of all No LTOT patients and 85% of all LTOT patients enrolled at these sites.

**P<0.0125 (0.05/4 time points) was considered statistically significant using a Bonferroni adjustment for multiplicity of comparisons[9].

††The Hospital Anxiety and Depression Scale (HADS) is a 14-item scale that is completed by the patient. The anxiety score ranges from 0 to 21, with higher scores indicating greater anxiety; MID=1.5. The depression score ranges from 0 to 21, with higher scores indicating greater depression; MID=1.5[18,19].

‡‡Pittsburgh Sleep Quality Index (PSQI) is a 24-item scale completed by the patient and partner. The total score ranges 0-21, with higher scores indicating worse sleep quality[20].

§§Severe resting desaturation is defined as saturation below 89% during room air resting saturation test.

¶¶P-value for Fisher's exact test

‖‖Severe exercise desaturation is defined as saturation below 80% for at least 1 minute during room air 6-minute walk test.

Table S11. Adverse events reported to be possibly, probably or definitely related to use of supplemental oxygen.

	No. of reports	Reports per 100 person-years
Expected, related events		
Fires related to oxygen use	2	0.08
Burn from smoking around oxygen equipment	3	0.12
Burn from using oxygen equipment around open flame	1	0.04
Burn from liquid oxygen frost	4	0.16
Nosebleed	9	0.35
Tripping/falling over oxygen equipment	23*	0.90
Total no. of expected, related events	**42**	**1.64**
Unexpected, related events		
Blisters, ear pain	3	0.12
Dry eyes	1	0.04
Funny feeling in sinus area	1	0.04
Increased intestinal gas	1	0.04
Headache	2	0.08
Nausea	1	0.04
Total no. of unexpected, related events	**9**	**0.35**
Total no. of all related events	**51**	**1.99**
Total no. of patients ever using supplemental oxygen during follow-up	**490**	
No. (%) reporting at least 1 related adverse event	42 (8.6%)	

*Two of these events involved hospitalization: overnight hospitalization with humerus fracture and 6-day hospitalization with rib fractures.

References

1.   Administration SS. Limited Access Death Master File Manual Batch Query Subscription. Alexandria, VA: National Technical Information Service, United States Department of Commerce.

2.   Gorecka D, Gorzelak K, Sliwinski P, Tobiasz M, Zielinski J. Effect of long-term oxygen therapy on survival in patients with chronic obstructive pulmonary disease with moderate hypoxaemia. Thorax 1997;52:674-9.

3.   Chaouat A, Weitzenblum E, Kessler R, et al. A randomized trial of nocturnal oxygen therapy in chronic obstructive pulmonary disease patients. The European respiratory journal 1999;14:1002-8.

4.   Drummond MB, Blackford AL, Benditt JO, et al. Continuous oxygen use in nonhypoxemic emphysema patients identifies a high-risk subset of patients: retrospective analysis of the National Emphysema Treatment Trial. Chest 2008;134:497-506.

5.   Ringbaek TJ, Viskum K, Lange P. Does long-term oxygen therapy reduce hospitalisation in hypoxaemic chronic obstructive pulmonary disease? The European respiratory journal 2002;20:38-42.

6.   Eaton TE, Grey C, Garrett JE. An evaluation of short-term oxygen therapy: the prescription of oxygen to patients with chronic lung disease hypoxic at discharge from hospital. Respir Med 2001;95:582-7.

7.   Crockett AJ, Moss JR, Cranston JM, Alpers JH. The effects of home oxygen therapy on hospital admission rates in chronic obstructive airways disease. Monaldi archives for chest disease = Archivio Monaldi per le malattie del torace / Fondazione clinica del lavoro, IRCCS [and] Istituto di clinica tisiologica e malattie apparato respiratorio, Universita di Napoli, Secondo ateneo 1993;48:445-6.

8.   Albert RK, Connett J, Bailey WC, et al. Azithromycin for prevention of exacerbations of COPD. N Engl J Med 2011;365:689-98.

9.   Hsu JC. Multiple Comparisons: Theory and Methods. London, U. K.: Chapman and Hall; 1996.

10.   Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988;93:580-6.

11.   Brooks SM. Surveillance for respiratory hazards. ATS News 1982;8:12-6.

12.   Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540-5.

13.   Kaplan RM, Atkins CJ, Timms R. Validity of a quality of well-being scale as an outcome measure in chronic obstructive pulmonary disease. J Chronic Dis 1984;37:85-95.

14.   Kaplan RM. The minimally clinically important difference in generic utility-based measures. Copd 2005;2:91-7.

15.   Barr JT, Schumacher GE, Freeman S, LeMoine M, Bakst AW, Jones PW. American translation, modification, and validation of the St. George's Respiratory Questionnaire. Clin Ther 2000;22:1121-45.

16.   Jones PW. St. George's Respiratory Questionnaire: MCID. Copd 2005;2:75-9.

17.   Ware Jr JE, Kosinski M, Bayliss MS, al e. Comparison of methods for the scoring and statisical analysis of SF-36 health profile and summary measures: Summary of results from the Medical Outcomes Study. Medical Care 1995;33:AS264-AS79.

18.   Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:133-8.

19.   Puhan MA, Frey M, Buchi S, Schunemann HJ. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. Health and quality of life outcomes 2008;6:46.

20.   Spira AP, Beaudreau SA, Stone KL, et al. Reliability and validity of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale in older men. The journals of gerontology Series A, Biological sciences and medical sciences 2012;67:433-9.

21.   Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Eng J Med 2004;350:1005-12.