Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Patterns of physical activity progression in patients with COPD

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Statistical analysis full version

We compared the baseline characteristics of patients with follow-up vs those lost to follow-up by descriptive statistics and obtained p-values using mixed logistic regression models with random intercepts for study.

We identified cluster groups (physical activity patterns) using k-means (1), a hypothesis-free method that allowed grouping patients based on the baseline level, the final level and the change in daily step count. We used the Calinski-Harabasz stopping rule to decide the number of clusters (2). To characterize the patterns, we described physical activity and physical activity experience variables according to the cluster groups and compared baseline to follow-up values by paired t-tests. Because both studies used the same data collection methods main results are based on the pooled dataset and corrected for study.

To assess determinants of physical activity progression patterns, we first compared subjects' characteristics by physical activity patterns and obtained p-values from mixed logistic regression models with random intercepts for study and city area to account for possible heterogeneity in unmeasured characteristics related to study and city area. Then we built a multivariable multinomial regression model using the generalized linear latent and mixed model, with also random intercepts for study and city (3). Model building combined step-forward and backward algorithms, and determinants were included in the final model if: (i) they related to the outcome with a p-value <0.05; or (ii) they modified (>10% change in regression coefficient) the estimates of the remaining variables in the model (4). We tested goodness of fit of the final model.

We performed the following sensitivity analyses: (1) to investigate a possible difference between the two study samples, cluster analysis and description of resulting patterns

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

was performed separately for both samples; (2) to test whether the observed patterns were due to changes in wearing time, we tested the association between the change in daily step count and the change in wearing time overall and per pattern; (3) to rule out a relevant effect of pulmonary rehabilitation on the physical activity patterns we repeated the clustering after excluding patients included in pulmonary rehabilitation programs at baseline and/or during follow-up.

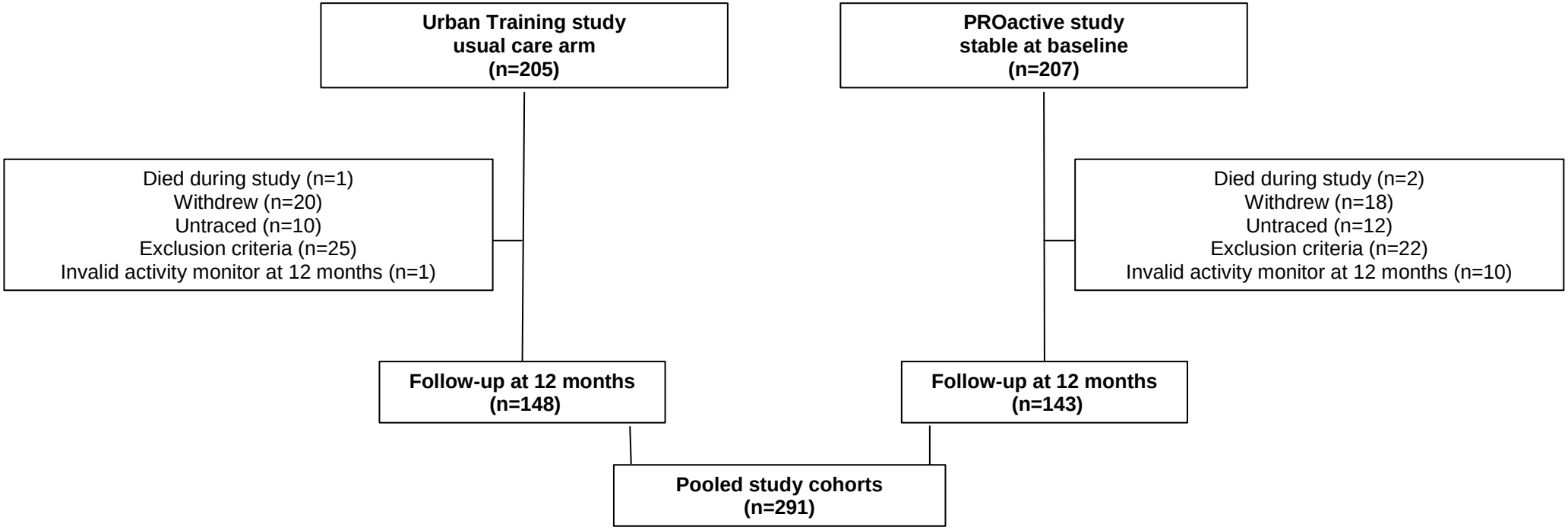
We estimated that the available sample size (n=291), fixed by the primary objectives of the original studies, was sufficient to identify physical activity patterns using cluster k-means, as our ratio of number of subjects to number of variables (291/3 = 97) was much higher than the 0.01 often used for the same analysis in other contexts (5,6).

Due to the small proportion of missing data (<2% of total data), we used a complete case strategy and reported missing data in the table footnotes.

All analyses were conducted using Stata/SE 14.2 (StataCorp, College Station, TX, USA).

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Figure S1 Flow of participants through the study.



Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Table S1 Patient characteristics at baseline for all patients (Urban Training and PROactive study, n=412) and for patients with 12-month follow-up vs lost-to follow-up.

	All patients	Follow-up	Lost-to follow-up	p-value[b]
	n = 412 (100%)	n = 291[a] (71%)	n = 121[a] (29%)	
Sociodemographic				
Age (years)	68±8	68±8	68±8	0.745
Sex (men)	316 (77)	237 (81)	79 (65)	**0.001**
Current smoker (yes)	77 (19)	52 (18)	25 (21)	0.508
Pack-years	58±41	58±41	60±41	0.684
Education, high school or higher	236 (57)	168 (58)	68 (56)	0.746
Interpersonal				
Living with a partner[c]	291 (71)	216 (74)	75 (63)	**0.016**
Active worker[d]	47 (11)	36 (12)	11 (9)	0.342
Grandparenting[e]	89 (43)	67 (45)	22 (39)	0.389
Dog walking [e]	26 (13)	20 (14)	6 (11)	0.566
Environmental				
Recruitment season				
Spring	55 (13)	35 (12)	20 (16)	0.161
Summer	82 (20)	58 (20)	24 (20)	
Fall	218 (53)	154 (53)	64 (53)	
Winter	57 (14)	44 (15)	13 (11)	
Urban vulnerability index (from 0 -lowest to 1 – highest)[e,f]	0.642±0.178	0.637±0.175	0.655±0.186	0.514
Clinical				
FEV_1 (% predicted)	57.7±18.9	58.6±19.3	55.6±17.9	0.140
FEV_1/FVC ratio	0.51±0.13	0.51±0.13	0.51±0.13	0.699
Airflow limitation severity (post-bronchodilator FEV_1)				
GOLD 1: Mild ($FEV_1 \geq 80\%$ predicted)	52 (13)	39 (13)	13 (11)	0.259
GOLD 2: Moderate ($50\% \leq FEV_1 < 80\%$ predicted)	207 (50)	147 (51)	60 (50)	
GOLD 3: Severe ($30\% \leq FEV_1 < 50\%$ predicted)	125 (30)	88 (30)	37 (30)	
GOLD 4: Very Severe (FEV_1 <30% predicted)	28 (7)	17 (6)	11 (9)	
6MWD (meters)	461±109	477±103	421±111	**<0.001**
CAT score (0–40)	13.3±7.5	12.9±7.6	14.2±7.3	0.094
CCQ score (0-6)	1.59±0.98	1.55±0.98	1.69±0.98	0.172
C-PPAC amount score (0-100)	67.8±16.9	69.0±15.8	64.2±19.5	**0.024**
C-PPAC difficulty score (0-100)	77.9±14.9	78.4±14.5	76.3±16.0	0.269
C-PPAC total score (0-100)	72.8±13.6	73.7±12.8	70.3±15.4	**0.044**
mMRC score (0-4)	1.4±1.0	1.3±0.9	1.7±1.1	**<0.001**
Any COPD exacerbation with hospital admission in previous 12 months	49 (12)	34 (12)	15 (13)	0.781
BMI (kg/m^2)	27.7±5.2	27.6±4.6	28.1±6.3	0.306
FFMI (kg/m^2)	18.8±3.2	19.0±3.0	18.4±3.5	0.086
Cardiovascular disease[g]	240 (59)	176 (60)	64 (54)	0.212
Ischemic heart disease[g]	40 (10)	29 (10)	11 (9)	0.823
Diabetes mellitus[g]	73 (18)	51 (18)	22 (18)	0.817
LABA or LAMA, alone	56 (14)	41 (14)	15 (13)	0.686
Inhaled corticosteroid with LABA and/or LAMA	256 (63)	179 (62)	77 (65)	0.557
Pulmonary rehabilitation at baseline	25 (6)	15 (5)	10 (8)	0.233
Knowledge of baseline PA	24 (6)	19 (7)	5 (4)	0.348
Psychological				
Anxiety (HAD-A, 0-21)	5±4	5±4	6±4	0.117
Depression (HAD-D, 0-21)	4±3	4±3	4±4	0.210
Physical activity				
Step count (steps/day)	6415±3678	6720±3667	5682±3613	**0.010**
Time in moderate-to-vigorous physical activity (≥3 METs; min/day)	95.8±45.9	99.4±45.3	87.0±46.2	**0.013**
Intensity during walking (m/s^2)	1.84±0.31	1.86±0.31	1.80±0.30	**0.050**
Sedentary time (h/day)	10.53±1.93	10.53±1.94	10.52±1.92	0.961

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Notes: Data are presented as n (%), mean±SD.

[a]Some variables have missing values, as follows. Follow-up: 1 in education, 1 in living with a partner, 1 in CAT total, 1 in CCQ score, 31 in C-PPAC scores, 1 in any COPD exacerbation with hospital admission in previous 12 months, 26 in FFMI, 3 in LABA or LAMA, alone, 3 in inhaled corticosteroid with LABA and/or LAMA, 3 in HAD anxiety and depression; Lost-to follow-up: 1 in living with a partner, 1 in 6MWD, 33 in C-PPAC scores, 3 in any COPD exacerbation with hospital admission in previous 12 months, 5 in FFMI, 2 in ICD10 codes: I00 to I99 for Cardiovascular diseases; I20 to I25 for Ischemic heart disease, E14 for Diabetes mellitus, 3 in LABA or LAMA, alone, 3 in inhaled corticosteroid with LABA and/or LAMA, 1 in HAD depression.

[b]p-value from mixed logistic regression models with random effects for study (Urban Training and PROactive), due to small numbers random effects for city area were not applied.

[c]marital status: living with a partner vs single, widowed or divorced.

[d]working status: active worker (working full-time or part-time) vs. unemployed, housework or retired.

[e]only available for Urban Training.

[f]The urban vulnerability index is a measure of socioeconomic status at the census tract level that combines demographic, economic, residential and subjective indicators, and ranges from lowest [0] to highest [1] level of neighborhood vulnerability.

[g]ICD10 codes: I00 to I99 for cardiovascular diseases; I20 to I25 for ischemic heart disease, E14 for diabetes mellitus.

Abbreviations: FEV_1: forced expiratory volume in 1 second; FVC: forced vital capacity; GOLD: Global Initiative for Chronic Obstructive Lung Disease; 6MWD: 6-min walking distance; CAT: COPD Assessment Test; CCQ: Clinical COPD Questionnaire; C-PPAC: Clinical visit—PROactive Physical Activity in COPD (higher numbers indicate a better score); mMRC: modified Medical Research Council; BMI: body mass index; FFMI: fat free mass index; LABA: long-acting $beta_2$-agonists; LAMA: long-acting anti-muscarinics; HAD-A: Hospital Anxiety and Depression scale - Anxiety; HAD-D: Hospital Anxiety and Depression scale – Depression; MET: metabolic equivalent of task.

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Table S2 Physical activity and physical activity experience variables at baseline and at 12-month follow-up, overall and by PA progression pattern (*Inactive*, *Active Improvers* and *Active Decliners*).

	All				***Inactive***				***Active Improvers***				***Active Decliners***			
	n = 291[a]				**n = 173[a] (59%)**				**n = 49[a] (17%)**				**n = 69[a] (24%)**			
	Baseline	**Follow-up**	**Change**	**p-value[b]**	**Baseline**	**Follow-up**	**Change**	**p-value[b]**	**Baseline**	**Follow-up**	**Change**	**p-value[b]**	**Baseline**	**Follow-up**	**Change**	**p-value[b]**
Step count (steps/day)	6720 ±3667	6474 ±3772	-246 ±2420	0.084	4621 ±1757	4134 ±1817	-487 ±1201	**<0.001**	7727 ±3275	11105 ±3330	3378 ±2203	**<0.001**	11267 ±3009	9051 ±2897	-2217 ±2085	**<0.001**
Time in MVPA (≥3 METs; min/day)	99.4 ±45.3	96.1 ±47.5	-3.3 ±29.1	**0.052**	74.5 ±25.9	68.2 ±26.6	-6.3 ±16.7	**<0.001**	114.8 ±47.2	152.7 ±47.1	37.9 ±26.8	**<0.001**	150.9 ±32.3	125.8 ±33.0	-25.0 ±26.0	**<0.001**
Intensity during walking (m/s^2)	1.86 ±0.31	1.82 ±0.31	-0.04 ±0.19	**<0.001**	1.78 ±0.27	1.73 ±0.27	-0.05 ±0.19	**<0.001**	1.90 ±0.34	1.95 ±0.32	0.05 ±0.17	0.062	2.03 ±0.30	1.95 ±0.30	-0.08 ±0.17	**<0.001**
Sedentary time (h/day)	10.53 ±1.94	10.46 ±1.98	-0.08 ±1.82	0.463	11.03 ±2.00	10.98 ±2.15	-0.05 ±2.10	0.744	10.29 ±1.73	9.59 ±1.54	-0.70 ±1.25	**<0.001**	9.46 ±1.39	9.76 ±1.29	0.30 ±1.17	**0.037**
C-PPAC amount score (0-100)	67.6 ±15.9	66.7 ±16.0	-0.9 ±12.7	0.305	61.5 ±14.2	60.3 ±14.7	-1.2 ±12.4	0.255	73.6 ±12.0	82.3 ±10.3	8.7 ±11.2	**<0.001**	83.6 ±10.3	77.1 ±9.4	-6.5 ±11.0	**<0.001**
C-PPAC difficulty score (0-100)	77.2 ±14.3	77.7 ±13.6	0.4 ±10.6	0.547	74.4 ±14.2	74.8 ±13.0	0.4 ±10.0	0.673	80.3 ±13.8	84.5 ±11.3	4.3 ±11.3	**0.054**	84.4 ±12.5	82.5 ±14.0	-1.9 ±11.5	0.295
C-PPAC total score (0-100)	72.4 ±12.6	72.2 ±12.6	-0.2 ±9.2	0.717	68.0 ±11.7	67.5 ±11.5	-0.4 ±8.5	0.562	76.9 ±9.5	83.4 ±8.8	6.5 ±8.0	**<0.001**	84.0 ±8.1	79.8 ±9.6	-4.2 ±9.7	**0.008**
Wearing time (h/day)	14.73 ±1.56	14.52 ±1.63	-0.21 ±1.67	**0.035**	14.68 ±1.72	14.35 ±1.82	-0.33 ±2.01	**0.034**	14.77 ±1.58	15.02 ±1.59	0.25 ±0.77	**0.028**	14.82 ±1.10	14.59 ±0.93	-0.23 ±1.03	0.064

Notes: Data are presented as mean±SD. For C-PPAC variables means and p-values are reported for patients with data at baseline and follow-up.
[a]C-PPAC variables have 87 missing values: 38 in *Inactive*, 21 in *Active Improvers*, and 28 in *Active Decliners*.
[b]paired t-test.
Abbreviations: MVPA: moderate-to-vigorous physical activity; MET: metabolic equivalent of task; C-PPAC: Clinical visit—PROactive Physical Activity in COPD (higher numbers indicate a better score).

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Table S3 Step count (mean steps/day) at baseline and at 12-month follow-up as well as selected variables at baseline, by cluster groups (physical activity progression patterns) identified by k-means, performed separately for the Urban Training and the PROactive study.

	Cluster 1					***Cluster 2***					***Cluster 3***				
	n (row%)	**Baseline**	**Follow-up**	**Change**	**p-value[a]**	**n (row%)**	**Baseline**	**Follow-up**	**Change**	**p-value[a]**	**n (row%)**	**Baseline**	**Follow-up**	**Change**	**p-value[a]**
Urban Training															
Step count (steps/day)	**79 (54%)**	6028 ±2176	5125 ±2045	-903 ±1428	**<0.001**	**36 (24%)**	6722 ±2491	10435 ±2293	3713 ±2288	**<0.001**	**33 (22%)**	13144 ±3340	11442 ±3606	-1702 ±2754	**0.001**
Age (years)		69±8					71±9					67±7			
Sex (men)		70 (89)					32 (89)					28 (85)			
FEV_1 (% predicted)		55.2±18.2					62.6±15.5					60.5±17.3			
6MWD (meters)		485±87					519±83					517±68			
mMRC score (0-4)		1.2±0.9					1.1±0.8					0.8±0.7			
PROactive															
Step count (steps/day)	**118 (83%)**	4431 ±1735	4097 ±1891	-334 ±1098	**0.001**	**6 (4%)**	11339 ±1239	14340 ±2593	3001 ±1716	**0.008**	**19 (13%)**	11188 ±1992	8222 ±2685	-2966 ±2166	**<0.001**
Age (years)		68±8					65±9					63±7			
Sex (men)		89 (75)					4 (67)					14 (74)			
FEV_1 (% predicted)		56.8±20.7					68.6±12.1					69.8±21.7			
6MWD (meters)		435±109					520±130					541±110			
mMRC score (0-4)		1.6±1.0					1.0±0.6					1.0±0.8			

Notes: Data are presented as mean±SD.
[a]paired t-test.
Abbreviations: FEV_1: forced expiratory volume in 1 second; 6MWD: 6-min walking distance; mMRC: modified Medical Research Council.

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Table S4 Correlation between the change in daily step count and the change in wearing time, overall and by PA progression pattern (*Inactive*, *Active Improvers* and *Active Decliners*).

	n (%)	Pearson correlation coefficient	p-value
All patients	291 (100%)	0.090	0.124
Inactive	173 (59%)	0.002	0.981
Active Improvers	49 (17%)	-0.097	0.508
Active Decliners	69 (24%)	0.162	0.184

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Table S5 Step count (mean steps/day) at baseline and at 12-month follow-up, by cluster groups (physical activity progression patterns) identified by k-means, performed separately for all patients (n=291) and excluding patients with rehabilitation (n=270).

	Inactive					***Active Improvers***					***Active Decliners***				
	n (row%)	**Baseline**	**Follow-up**	**Change**	**p-value[b]**	**n (row%)**	**Baseline**	**Follow-up**	**Change**	**p-value[b]**	**n (row%)**	**Baseline**	**Follow-up**	**Change**	**p-value[b]**
Step count (steps/day), all patients	**173 (59%)**	4621 ±1757	4134 ±1817	-487 ±1201	**<0.001**	**49 (17%)**	7727 ±3275	11105 ±3330	3378 ±2203	**<0.001**	**69 (24%)**	11267 ±3009	9051 ±2897	-2217 ±2085	**<0.001**
Step count (steps/day), patients with rehabilitation excluded[a]	**154 (57%)**	4544 ±1763	4003 ±1753	-541 ±1209	**<0.001**	**50 (19%)**	7516 ±3187	10730 ±3345	3213 ±2266	**<0.001**	**66 (24%)**	11206 ±3033	9078 ±2928	-2128 ±1943	**<0.001**

Notes: Data are presented as mean±SD.
[a]patients in pulmonary rehabilitation at baseline and/or follow-up were excluded (n=21).
[b]paired t-test

Document downloaded from http://www.elsevier.es, day 11/02/2026. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

References:

1. Genolini, Christophe Falissard B. KmL: k-means for longitudinal data. Comput Stat 2010;25:317–328.

2. Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat - Theory Methods 1974;3:1–27.

3. Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. J Econom 2005;128:301–23.

4. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken, New Jersey: John Wiley & Sons; 2013.

5. Wang Y, Miller DJ, Clarke R. Approaches to working in high-dimensional data spaces: Gene expression microarrays. Br J Cancer 2008;98:1023–8.

6. Garge NR, Page GP, Sprague AP, Gorman BS, Allison DB. Reproducible clusters from microarray research: Whither? BMC Bioinformatics 2005;6:S10.