weekdays. Weekend days were used only for the presentation of daily physical activity hourly patterns. All values are represented by absolute values. Values relative to peak exercise capacity were not presented as measurement of maximal exercise capacity was not available. For the clustering of patients, a set of relevant variables were generated after stratifying averages of physical activity measures according to different criteria (i.e., intensity, duration, period of the day, frequency and quantity, or the combination of these criteria; Table XVI in the appendix).

4.3.2 Statistical analyses

Continuous variables were expressed as median (interquartile range), as most variables presented non-normal distribution. Categorical variables were expressed as absolute and/or relative frequency. Mann-Whitney U test or Kruskal-Wallis test (post hoc Dunn; significant if $P<0.05$) was used for comparing continuous variables, while the chi-square test was used for categorical variables. The influence of seasons on daily physical activity measures was minimal (Table XVII in the appendix) and therefore this was not taken into consideration throughout the analyses. Spearman coefficient was used to investigate correlations, when appropriate. The area under each hourly pattern, named as the Area Under the Curve (AUC), was calculated and presented with its 95% confidence intervals in order to quantitatively represent time-varying averages of the hourly patterns. $P<0.01$ was considered significant and all statistical analyses were performed using SPSS 17.0 (SPSS, Chicago, Illinois, USA) or GraphPad Prism 5 (GraphPad Software, La Jolla, California, USA).

Cluster analysis was adopted to identify subgroups with distinct physical activity profiles. Firstly, Principal Component Analysis (PCA) was used to compress the information contained in the high-dimensional feature set (180 dimensions) to a lower subspace (three dimensions) that is both convenient for data visualization and able to account for the desired variance of the data (set to 60%). The features were initially standardized using z-scores. Secondly, a k-means clustering algorithm with automatic selection of the number of clusters was applied to the three principal components to separate the subjects into groups with distinct characteristics. The algorithm selects the number of clusters in a way that the corresponding clustering results are the most stable under small perturbations of the input dataset [58]. The normalized mean over pairwise clustering distances was used as instability measure [58]. Feature extraction, PCA and cluster analysis were performed using Matlab R2012b (Mathworks Inc., USA).

4.4 Results

4.4.1 General characteristics

In total, 1001 patients with COPD were analysed (Table IV). The majority of patients were men, had a normal-to-overweight BMI, moderate-to-severe degree of airflow limitation, and only a small proportion used long-term oxygen therapy (LTOT). Compared to female subjects, male subjects were slightly older (67 (62 – 73) versus 65 (59 – 71) years; $P<0.0001$) and had higher BMI (26.5 (23.3 – 29.9) versus 24.5 (21.1 – 28.6) $kg \cdot m^{-2}$; $P<0.0001$), but no differences were found in FEV_1, modified Medical Research Council (mMRC) grades, or Global Initiative for