**Statistical Analysis**

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) was used for comparing continuous variables, whilst the chi-square test was used for categorical variables. Spearman coefficient was used to investigate correlations, when appropriate. *P*<0.05 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). Details about sample size calculation can be found in the online supplement.

Cluster analysis was adopted for identifying subgroups with distinct PA profiles. Firstly, Principal Component Analysis (PCA) was used to perform dimensionality reduction. PCA transforms the data into a subset of linearly uncorrelated variables so that the variance of the data in the low-dimensional representation is maximized. The subset of variables are called principal components and constructed in a way that the first principal component has the largest possible variance, and each subsequent component has the highest variance possible under the constraint that it is uncorrelated with the preceding components. In this study, the high-dimensional feature set (180 dimensions) was projected to a 3 dimensional space of principal components, which is practical for data visualization. Secondly, a k-means clustering algorithm with automatic selection of the number of clusters was applied to the 3 dimensional principal components space 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. The normalized mean over pairwise clustering distances was used as an instability measure.([1](#_ENREF_1)) The features were first standardized using z-scores. Feature extraction, PCA and clustering analyses were performed using Matlab R2012b (Mathworks Inc., USA).

5 vs Cluster 4.

1. von Luxburg U. Clustering stability: An overview. *Foundations and trends in machine learning* 2010;2:235-274.