Cluster analysis was adopted for identifying subgroups with distinct PA measures. Firstly, Principal Component Analysis (PCA) was used to perform dimensionality reduction.

PCA transforms the data into a subset of linearly uncorrelated variables called principal components so that the variance of the data in the low-dimensional representation is maximized. The components are constructed in such 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 orthogonal to (i.e. 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 practical for data visualization.

Secondly, a k-mean clustering algorithm with automatic selection of the number of clusters was applied to the 3 dimensional principal components space to partition the patients into groups with distinct characteristics. The algorithm selects the number of clusters in such 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 instability measure (ref).

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