**Statistical analyses.**

Hierarchical clustering on principal components was performed as described by Lê et al. (1). This method combines multiple correspondence analysis, hierarchical clustering and partitional clustering to reveal the resemblances and differences between individuals from a multidimensional standpoint.

The first step was a multiple correspondence analysis (MCA) of the dataset (2). To address whether our dataset was adapted to MCA, we first tested its suitability for structure detection using Bartlett's test of sphericity (3). p-value was < 0.001, indicating that MCA was very likely to be useful despite the small number of patients.

Correspondence analysis was thus performed on the indicator matrix of the variables (individuals × variables matrix). Associations between variables were highlighted by calculating the chi-square distance between the different categories of variables and between the individuals. These associations were then represented graphically in a factorial map. In this representation, individuals were considered in a high dimensional Euclidean space. The first axis of the factorial map was the dimension that best described variance of the data, the second axis the second dimensions that best described the variance of the data, and so forth. The number of axes retained for analysis was determined by calculating modified eigenvalues. An Euclidean subspace of lower dimensions that preserves the distances between individuals as much as possible was then determined (i.e. the principal component). The principal component was set to minimize the overfitting effect (the production of an analysis that corresponds too closely to the particular set of the data, and may therefore fail to fit additional data or predict future observations reliably). Overfitting may occur when many parameters are estimated with respect to the number of observed values. This issue is especially important when the structure of the dataset is low, meaning that the relationships between variables may not be strong. To reduce overfitting, it is necessary to reduce the number of dimensions for the imputation step in order to estimate less parameters; however it is important not to remove too many components since information can be lost. Since the number of dimensions affects the prediction of the imputed values and the estimation of the axes and components, the mean square error of prediction (MSEP) appears to be a well-fitted criterion to select the number of components (4). In this perspective, we used K-fold cross-validation, 5% of missing values were inserted at random in the data matrix and predicted with a MCA model using 0 to 10 dimensions. This process was repeated 1000 times. The number of components leading to the smallest MSEP was retained. In summary, MCA can be viewed as a “denoising” method which separates signal and noise in the dataset: the ﬁrst dimensions extract the essential of the information while the last dimensions are restricted to noise, with the threshold between information and noise set to minimize the overfitting effect.

The second step was a hierarchical clustering onto the principal component. A hierarchical tree was built using Ward’s criterion (5). This criterion is based on the Huygens theorem which allows to decompose the total inertia (total variance) in between-group and within-group variance (Total inertia = Between inertia + Within inertia). Ward’s method consists in aggregating clusters such that the growth of within-inertia is minimum (in other words minimizing the reduction of the between-inertia) at each step of the algorithm. The within-inertia characterizes the homogeneity of a cluster. The hierarchy is represented by a dendrogram which is indexed by the gain of within-inertia.

Finally, in a third step, partitioning (setting the number of clusters) was performed. The hierarchical tree can be considered as a sequence of nested partitions from the one in which each individual is a cluster to the one in which all the individuals belong in the same cluster. Partitioning was chosen by studying the criteria Δ(Q) / Δ(Q + 1), Δ(Q) being the between-inertia increase when moving from Q − 1 to Q clusters. The number Q which minimized this criterion was retained (6). Of note, this method generally yields a number of clusters that generally corresponds to that expected from examining the general shape of the factorial map and the dendrogram.

Qualitative variables are presented as counts and percentages and were compared between groups using the unconditional z-pooled test as recommended by Lydersen and al.(7) Quantitative variables are presented as median (range) and were compared using the Mann-Whitney U test (when comparing two groups) or the Kruskall Wallis test (when comparing more than two groups). Correction for multiple comparisons was performed using the Bonferroni-adjusted *p* value method. Confidence intervals for the ratio of two binomial proportions were calculated using the Miettinen-Nurminen asymptotic score method as recommended by Lydersen et al. (7) and Fagerland et al. (8). All statistical tests were two-tailed. A *p* value <0.05 was considered statistically significant.

All analyses were performed using R software version 3.2.2 R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (URL https://www.R-project.org/.) Hierarchical clustering on principal components was performed using FactoMineR, An R Package for multivariate analysis developed by Lê et al. (1) available at https://cran.r-project.org/web/packages/FactoMineR/index.html.

The analyses were performed with the scientific and technical support of François Severac, MD, PhD, expert in biostatistical analysis (Groupe Méthodes en Recherche Clinique, Service de Santé Publique, Hôpitaux Universitaires de Starsbourg; Strasbourg, France).

**References for the statistical analysis:**

1. Lê S, Josse J, Husson F. FactoMineR : An *R* Package for Multivariate Analysis. J Stat Softw. 2008;25:1.

2. Greenacre M, Blasius J, Blasius J. Multiple Correspondence Analysis and Related Methods. Chapman and Hall/CRC; 2006.

3. Tobias S, Carlson JE. Brief report: bartlett’s test of sphericity and chance findings in factor analysis. Multivar Behav Res. 1969;4:375–7.

4. Josse J, Chavent M, Liquet B et al. Handling Missing Values with Regularized Iterative Multiple Correspondence Analysis. J Classif. 2012;29:91–116.

5. Ward JH. Hierarchical Grouping to Optimize an Objective Function. J Am Stat Assoc. 1963;58:236–44.

6. Husson F, Josse J, Pagès J. Principal component methods-hierarchical clustering-partitional clustering : why would we need to choose for visualizing data ? Technical Report – Agrocampus. Applied Mathematics Department, September 2010. http://www.agrocampus-ouest.fr/math/

7. Lydersen S, Langaas M, Bakke Ø. The exact unconditional z-pooled test for equality of two binomial probabilities: optimal choice of the Berger and Boos confidence coefficient. J Stat Comput Simul. 2012. 82:1311-1316

8. Fagerland M, Lydersen S, Laake, P. Statistical Analysis of Contingency Tables. Chapman and Hall/CRC; 2017.