Cluster analysis: Cluster analysis is a statistical technique to discover homogeneous subgroups based on a set of measurements. In practice, a cluster analysis is the end product of a series of analytical decisions. This series of analytic decisions typically involve choices about what objects to cluster, what proximity measure to use to determine similarity or dissimilarity among the objects, and what type of clustering algorithm to use. In our data, we aimed to cluster fingers that were affected with DD. The measures of proximity and type of clustering algorithms are explained below.

Measures of proximity: To identify clusters of observations (i.e. combinations of affected fingers) in the data, it is important to know how similar individual observations (i.e. individual affected fingers) are, or how far apart they are. For binary data (a finger is affected with DD or not) several measures of similarity can be used, all based on measures of a 2x2 contingency table. We used Jaccard’s coefficient. This method only gives weight to the similarity of two fingers when DD is present in these fingers. Fingers without DD are ignored in this similarity measure. Besides the proximity between two individual observations (for example affected thumb and affected little finger), it is important to measure the dissimilarity of groups of observations in a cluster analysis. This inter-group proximity is based on the inter-individual proximity. In the complete linkage cluster method, which we used, the inter-group dissimilarity is defined as the largest distance between two individual observations, one from each group. This is also known as furthest-neighbor distance.

Type of clustering algorithm: Clustering algorithms can be classified as hierarchical or non-hierarchical. Hierarchical algorithms are most appropriate for classification when objects are related via some underlying systematic structure. Hierarchical algorithms are further classified according to whether the algorithm proceeds by successively merging individual observations into groups (agglomerative method) or starts with one large cluster and separates the observations into smaller groups (divisive method). Agglomerative methods are most widely used.

In agglomerative hierarchical clustering the clusters are formed in several steps. It starts with all single objects as separate clusters, in our case all five fingers are seen as separate cluster at the beginning of the analysis. Successively these objects are grouped into larger clusters until the final grouping contains all the original objects in one group. A dendogram illustrates which fusions are made in each step of the analysis. For example, our dendogram (Figure 2 in the paper) shows that in the first step the middle finger and ring finger are clustered (and thus are most similar compared to other combinations). In the following step the little finger was added to this first cluster, keeping the thumb and index finger still as single clusters.

With a multivariate ordinal logit model an ordinal logistic regression-like analysis was performed. This model is suitable for categorical data with ordered categories (i.e. Tubiana stage), measured at multiple time points or locations (i.e. five fingers). The model takes into account that observations on one hand could be correlated. In addition, covariates can be included in the statistical model to distinguish in severity level between subgroups of patients. In our paper, we used this model to calculate the correlations on the severity between fingers, corrected for age and gender.

Bootstrapping: method to obtain confidence intervals on the parameter estimates of the multivariate model.

References:

