Covariance matrices of fMRI scans yield information about a person. Attempts have been made to establish the similarity between two scans of the same person at different points in time. To find the similarity, previous works have calculated the correlation between matrix entries for groups at two time points. If the highest correlated matrix of person x at time 1 is that of person x at time 2, then this algorithm is considered to have correctly identified person A. The *id rate* of this method is defined as the percent of subjects correctly matched between time points.

There are two points to consider in this definition. First, the id rate may be dependent on the size of the sample studied. As an extreme example, in a sample of one person scanned at two time points, the id rate would always be 100%. For large samples, id rates will likely degrade due to the increased number of data points. Id rate is sample dependent, and it is difficult to compare id rate measures across different samples. Using subscripts to denote the measured correlation matrix at scan index, id rate may be defined as

where is a chosen measure of similarity between two correlation matrices.

Second, it is not clear that correlation between matrices is the best measure of similarity. Matrices can be compared in many ways. An optimal similarity measure will pay less attention to those variations seen on repeated scans of the same person and will put more emphasis on those features that are different between subjects.

To solve the first problem, we seek an alternative measure of identifiability that is independent of sample size. More formally, if there is a sample A with scans of a cohort at two time points, and the similarity measure *f* for A equals *s*, then the expected similarity value for a random subset of A should also equal *s* (so long as the subset contains at least two subjects).

Note that id rate does not necessarily satisfy this property. Subsets of a sample will, on average, have higher id rates due to having fewer comparisons. As an alternative, we propose the *pairwise id rate*, which is the rate at which a randomly chosen individual *x* will be correctly identified when compared against another randomly chosen individual . Pairwise id rate is

This definition leads into the second problem, which is determining the appropriate function s for calculating similarity. Previous studies of id rate have used the correlation between fc matrix elements to calculate s. Here we will discuss some alternatives.

**Similarity measures**

**Metrics**

A metric is a function between pairs of points that makes explicit the idea of closeness. Multiple metrics may be applied to the same space, and each metric may endow the space with a different set of properties. For instance, consider the Earth’s north and south poles. When calculating the distance between these points, one must first determine what paths are available. If paths through the center of the Earth are allowed, then the standard Euclidean metric, where the shortest path is a straight line, gives the distance, which is about 8,000 miles. However, if paths are restricted to the surface, then the shortest path is any of the lines of longitude, and the distance is about 12,000 miles. Alternatively, one may use the difference in latitude as a metric, in which case the distance is 180 degrees.