Supplement A. Converting patient subtypes into patient groups

Latent class analysis (LCA) and latent Dirichlet allocation (LDA) can both be seen as decomposition methods: They reduce patients, represented by a number of symptoms, to the probability that they belong to some class (LCA) or to a distribution of topics (LDA). For LCA, it is standard among LCA implementations to group patients by the class they are most likely to be in. This allows an easy interpretation. It is also standard to perform additional analyses to assess fit, such as by analyzing how many patients are likely to be in multiple groups.

LDA is less standardized. For our purposes, to stay close to other machine learning methodologies, we used k-means clustering. Other methods would have worked as well. However, unlike the groups generated through LCA, LDA clusters do not lend themselves to as easy of an interpretation—there is no inherent correspondence between clusters and the subtypes, since k-means can be used to generate an arbitrary number of clusters. To translate patient groups to subtypes, we calculated the average mixture for each group and associated groups based upon the average mixture. For our particular data, there was a one-to-one correspondence between the average mixtures and the subtypes with respect to which subtype was in the majority.

That said, this need not be the case—perhaps the average mixture for one cluster is an even split of subtypes. There are many different routes one could take in response. For example, it would also make sense to define subtypes based upon the patient groups instead of vice versa. However, there is the a priori question of what the end goal is. Our goal was to provide descriptive data, as is done in many latent variable model studies. However, for many machine learning tasks, the goal is prediction, in which case subtypes can be evaluated instead based upon accuracy metrics (area under the curve, specificity, sensitivity, F1-scores, etc.), in which case, generating patient groups may not be of interest.