Abstract LR

To generate evidence needed to address major challenges facing healthcare systems, health services researchers are often confronted with heterogenous populations and multivariate measurements. Cluster analysis is a statistical approach used to summarize complex data by identifying common patterns and grouping individuals based on shared characteristics. It provides a data-driven classification for the population under study and allows for relating the homogeneous profiles thus obtained with relevant covariates.

Despite its usefulness, cluster analysis involves several pitfalls that require careful consideration. In particular, the sample dependence of the results remains under-studied in the context of distance-based clustering. This work addresses the issue by applying a standard cluster analysis to two health services research datasets, critically appraising the process, and developing a new method to assess the impact of sampling error on the findings.

The thesis presents three studies, each corresponding to a published article. The first identifies six typical trajectories of healthcare utilization among 2271 community-dwelling older adults and examines their relationship with sociodemographic and health factors, thus revealing specific associations between high utilization clusters and manifestations of frailty and functional ageing. The second study uncovers five distinct patterns of working conditions and experiences among 1707 healthcare professionals and relates them with the intent to stay in various healthcare professions. The third study develops a bootstrap and multilevel model procedure to account for the sampling uncertainty in a standard cluster analysis, focusing on the use of a typology in regression models.

The empirical applications of clustering featured in the thesis provide data-driven insights for health services resource planning and policy development. Additionally, common methodological limitations of clustering studies are discussed and a novel procedure to strengthen their robustness proposed. Our work establishes a framework to improve the reliability of cluster analysis findings and contributes to advancing its use in health services research and beyond.