Genealogical information from co-insurance networks in pseudonymized administrative claims data in Austria

Florian Endel, *Vienna University of Technology, Vienna*

Introduction Routinely collected administrative claims data from the Austrian health and social insurance system is available for research in the GAP-DRG database. It is operated by Vienna University of Technology on behalf of the Main Association of Austrian Social Security Institutions. GAP-DRG holds pseudonymized information on reimbursement of prescriptions, inpatient and ambulatory outpatients contacts of almost all 8 million inhabitants. Genealogical information and family relationships are not directly available in the database. In this project, it is indirectly deduced, analyzed and integrated into GAP-DRG. This project is part of the K-Project dexhelpp in COMET and funded by BMVIT, BMWGJ and transacted by FFG.

Methods Co-insurance of relatives as spouses, children and close family members is encoded in the reimbursement information of GAP-DRG. These relationships between two persons are used to extract networks representing individuals who are associated with each other by co-insurance. Persons are classified as children, parents, in a relationship or single based on thorough data analysis and applying rules originating from qualitative descriptions of family structures in Austria. Additional data as the direction of the graph, representing the dependence of one partner on another and weights of edges holding information on e.g. difference in age is included. Visualization and common methods from graph theory are utilized to extract more details about data quality, social structure of the insured population and also limitation of the data and applied approach.

Results Depending on quality requirements, there are around 2.000.000 persons in the final dataset on co-insurance. In addition to the estimation of genealogical information, new insights into the database and especially data quality are acquired (e.g. persons older than 120 years could be identified as miscoded children due to their dependence on their parents). Networks of related persons allow in-depth analysis and informative visualizations. New quality issues were identified and missing information on e.g. the socio-economic status could be imputed or corrected. Furthermore, the estimated personal information enables novel research questions and. Due to the stepwise procedures, the implemented approach can be directly adapted to new data or particular projects.

Discussion Although solid and promising results have been obtained, additional analysis and concrete limitation have to be discussed. The quality and interpretation of co-insurance networks might vary over time, region and data source (e.g. social insurance institution). Because relationships are derived from co-insurance, couples not depending on each other directly or indirectly by a common child cannot be detected. As a result, the identification of parents is of a higher quality. External validation, verification of the methodology and its application have to be discussed.

Conclusion Genealogical information and networks of co-insurance can be estimated using administrative data. The presented method is straightforward and flexible but also pointed out limitations of the data collection and its quality. Previous knowledge about GAP-DRG and its general quality and trustworthiness could be verified. Summarizing, the newly acquired information on relationships and the extracted networks of co-insurance are interesting on their own and are expected to are the basis of novel data analysis and research.