

DATA QUALITY

Abstract 1401

NEW CHALLENGES AND ISSUES WITH THE RECORD LINKAGE OF AUSTRIAN HEALTH INSURANCE DATA OF DIFFERENT SOURCES AND THE NEED FOR A RECORD LINKAGE FRAMEWORK

Authors:

Barbara Glock¹, Florian Endel², Gottfried Endel³, Niki Popper⁴
¹*Vienna University of Technology, Institute for Analysis and Scientific Computing*, ²*Institute for Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstrasse 8, 1040 Vienna, Austria*, ³*Main Association of Austrian Social Security Organisations, Kundmannngasse 21, 1031 Vienna, Austria*, ⁴*dexhelpp, Neustiftgasse 57-59, 1070 Vienna, Austria*

Presenting Author: **Barbara Glock**Corresponding author
email : **barbara.glock@dwh.at**

Due to data privacy issues, routinely collected data of different sources is *pseudonymized* (e.g. MBDS minimum basic data set from the Federal Ministry of Health, which up to 2015 even don't have a personal identifier). This makes statistical analysis for decision support and health care planning very difficult, since no statements on patient pathways can be made. *Data from insurance carriers (FoKo) is event based*: each time a hospital reports to the insurance carrier a new data entry is generated. To enable efficient, significant and quality assured data analysis for patient centred assertions a record linkage of these episodes is required.

For historical data a linkage has been done before, but there are **some challenges for the new data sets**: in MBDS data for the whole of Austria is available, but in FoKo only data for Lower Austria; in FoKo a hospital stay may be split off in more data entries, due to intermediate reportings from the hospital; reported diagnoses of split episodes may differ from each other; and many more. We propose as first step of the **record linkage** to determine the already quality assured **matching variables** (e.g. birth year, gender), together with a *quality check* for same episodes (equal in all variables) with different person-IDs. Then a **sequencing** for split off episodes in FoKo, which are then joined, can be established. The minimum on variables that identify the data entry uniquely is determined. Finally, the record linkage is done, by firstly checking for matches in all chosen variables. If a unique match exists, the episodes are matched. Then an **iterative process** starts where *different variables are varied*. **Quality checks** after each run are included and this process can be applied more than once on the still remaining unmatched episodes.

The results of this record linkage are on the one hand the linked episodes for further statistical analyses and on the other hand *statements on the quality of the given data*, that are established through the preparing phases and also in the matching phases of this process.

Once the record linkage is done for existing data, the work is not finished yet, since more data will be available soon. This **pleads for the establishment of a record linkage framework**, for which the authors of this paper are preparing semi-automated fundamentals within the *K-Project dexhelpp in COMET – Competence Centers for Excellent Technologies, funded by BMVIT, BMWGJ and transacted by FFG*.