OVERCOMING THE LIMITATIONS OF CLAIMS DATA: LINKAGE OF CLAIMS DATA WITH A SECONDARY DATA SOURCE IN GERMANY

Xcenda®
AmerisourceBergen

Jacob C, Günther J, Borchert K, Krinke K-S, Braun S
Xcenda GmbH, Hannover, Germany

BACKGROUND

- Claims data analyses offer several strengths, like high actuality of data and health resource utilization under real-life conditions independent from a predefined study purpose.

- Furthermore, claims data provide the full picture of reimbursed healthcare costs from a statutory health insurance perspective, at least in the German setting.[1]

- Nevertheless, researchers have to face certain limitations such as a delay in clinical data or patient-reported outcomes.[2]

- To overcome these limitations, claims data can be linked to other primary or secondary data sources.

OBJECTIVE

- The aim of this study was to give an overview of opportunities and challenges of linking claims data with other data sources in Germany.

METHODS

- A systematic literature search was performed in the electronic database PubMed.

- All publications available until June 2017 using data linkage between German claims data and further primary or secondary data sources were included in this study by searching for "link*" in combination with "sickness fund", "health insurance", "claims", and "German*" with * as a wildcard for multiple further characters.

- Study focus, study periods of the applied data, claims and further data sources, and the reasons for linking data sources were analyzed by two independent researchers.

RESULTS

- The search resulted in 92 studies, of which 23 were included in the analysis after abstract screening.

- An increasing application of data linkage with claims data can be observed over recent years.

- In 2006 (year of the first publication), only one study was published; but there was a significant peak in 2017 (until June), when six studies were published (**Figure 1**).

Figure 1. Year of Publication of the Identified Studies
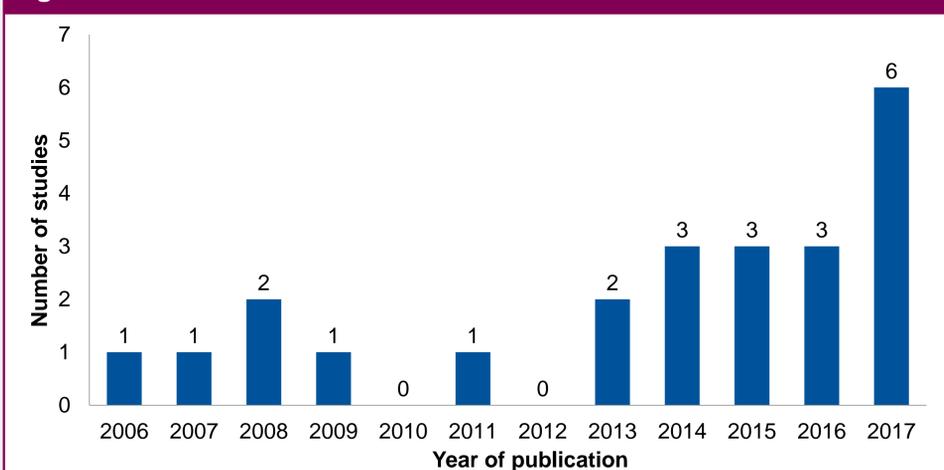
- Claims data that were linked in the identified studies came from various German statutory health insurances. Data from the "Techniker Krankenkasse" (TK) and "Allgemeine Ortskrankenkasse" (AOK) group (AOK Bremen, AOK Hessen, AOK Nordost, AOK Nordwest) were used most commonly (**Figure 2**).

- At the time of publication, the claims data used were between two and eight years old, with most of the data being four years old (n=8 studies).

- Ten of the identified studies linked claims data to primary data and 13 studies linked them to another secondary data source.

- Primary data were collected via surveys or prospective studies. Secondary data sources were registries (n=6), data of hospital information systems (n=2), official statistics (n=1), drug dispensation data (n=2), data from the health insurances' medical service (n=1), and data on a prevention program (n=1).

Figure 2. Claims Data Sources

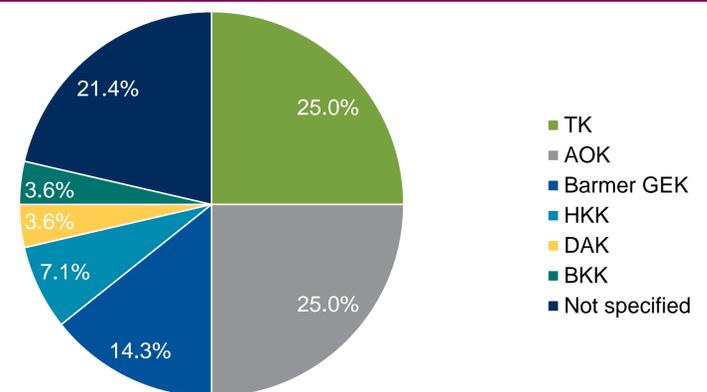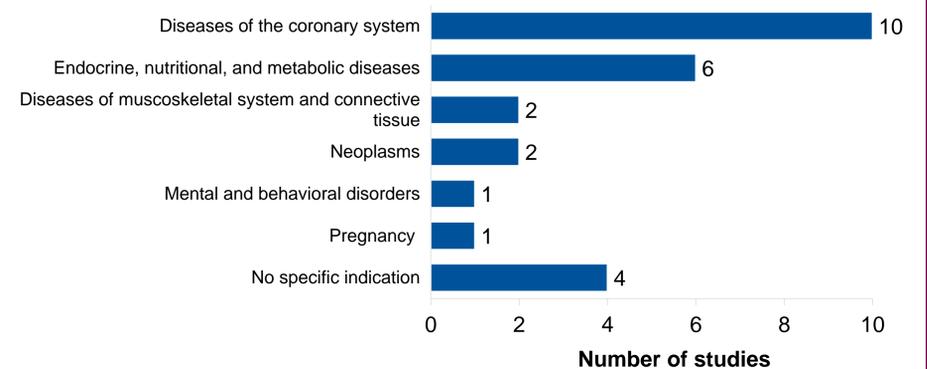Legend: TK, AOK, Barmer GEK, HKK, DAK, BKK, Not specified

Figure 3. Research Topics of the Included Studies

- Most studies had disease-specific research objectives. Almost 40% of studies (n=10) focussed on diseases of the coronary system and 23% (n=6) focussed on endocrine, nutritional, and metabolic diseases. Others focussed on neoplasms (n=2), mental and behavioral disorders (n=2), or pregnancy (n=1) (**Figure 3**).

- Data linkage was mostly motivated by the ability of linking missing or unavailable information to the claims database or by comparing claims data with other data sources.

CONCLUSIONS

- Data linkage constitutes a promising opportunity to overcome limitations of claims data research, although the application of data linkage in German healthcare studies is a rarely utilized approach until recently.

- It provides the opportunity to extend claims data in terms of data on health-related quality of life, sociodemographic characteristics, clinical variables, or indication-specific survey instruments, for example.

- Most of the studies used primary data to close information gaps of claims data; however, several challenges had to be addressed to leverage this opportunity.

- Overall, data linkage is methodically and technically challenging, and the right linkage method - out of a pool of methods - has to be identified that meets the requirements of the data sources.

- A challenge when linking claims data to a secondary data source is that data have to be of high quality in order to link persons or personalized data completely and correctly. Moreover, a key variable is needed that must be identical in both data sources and should be valid and reliable.[3]

- A challenge when linking claims data to a primary data source is that statutory health insurances often have strict criteria regarding the sampling of patients, mailing of questionnaires, and storage of data to secure privacy and confidentiality.[4]

- Besides, results of survey data may be affected by differences in the respondents' survey-response behavior. It is possible that certain groups of patients are not willing to participate in a survey (e.g., very ill individuals).[5]

REFERENCES

1. Neubauer S, Kreis K, Klora M, Zeidler J. Access, use, and challenges of claims data analyses in Germany. *Eur J Health Econ.* 2017;18:533-536. 2. Tyree PT, Lind BK, Lafferty WE. Challenges of using medical insurance claims data for utilization analysis. *Am J Med Qual.* 2006;21(4):269-275. 3. March S, Iskenius M, Hardt J, et al. Methodische Überlegungen für das Datenlinkage von Primär- und Sekundärdaten im Rahmen arbeitsepidemiologischer Studien. *Bundesgesundheitsbl.* 2013;56:571-578. 4. Roettger J, Bluemel M, Engel S, et al. Exploring health system responsiveness in ambulatory care and disease management and its relation to other dimensions of health system performance (RAC) – study design and methodology. *Int J Health Policy Manag.* 2015;4(7):431-437. 5. Roettger J, Bluemel M, Busse R. Selective enrollment in disease management programs for coronary heart disease in Germany – an analysis based on cross-sectional survey and administrative claims data. *BMC Health Serv Res.* 2017;17:246.