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 claims data 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).

- 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 Hannover, AOK Nordwest, AOK Nordost) 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).

CONCLUSIONS

- 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.

REFERENCES