

Access, use, and challenges of claims data analyses in Germany

Sarah Neubauer¹ · Kristine Kreis¹ · Mike Klora¹ · Jan Zeidler¹

Published online: 22 November 2016
© Springer-Verlag Berlin Heidelberg 2016

Worldwide, information infrastructures have developed rapidly towards big data over the past few years. The evolution of powerful computers has made it possible to improve the data exchange process within the healthcare system [1]. This technological trend has also triggered new and powerful possibilities in health service-related research (e.g., claims data analyses), because it now allows the analysis of more complex and larger data sets.

In this editorial, we focus on claims data analyses, because they are an important and powerful source of information to support the decision-making processes of healthcare stakeholders, researchers, and policy-makers regarding various aspects of the healthcare system. A few years ago, the statutory health insurance (SHI) companies in Germany made their data available for research purposes, because they realized the comprehensive potential of these data for both rational allocation of resources and for health services research to optimize healthcare provisions. The German government increasingly supports and funds claims data research, thus confirming the growing relevance of such research. Furthermore, recently, this research field developed a sophisticated methodological framework for comprehensively analyzing and interpreting the data in the context of various research questions. The significant increase in the number of claims data studies (and frequency of policy-relevant research) confirms this trend [2, 3] (see Fig. 1).

Claims data studies mainly use the information stored in data warehouses of the SHI. As a form of administrative data, primarily collected for billing and reimbursement purposes, claims data belong to the category of secondary data. These data are transmitted directly from healthcare providers to SHI funds, which include information on (all) cross-sector contacts between the insured and the healthcare system in their database. Compared with clinical trials, claims data analyses could reflect real-life healthcare provisions. Further advantages of this data source are the cost-efficient generation of data and a typically large study population compared with randomized controlled trials, which researchers often view as the gold standard for providing information on treatment efficacy and safety in a clinical setting. These data also help include groups that are typically more difficult to observe through primary data collection (e.g., children, severely ill individuals, dementia patients, or residents of nursing homes).

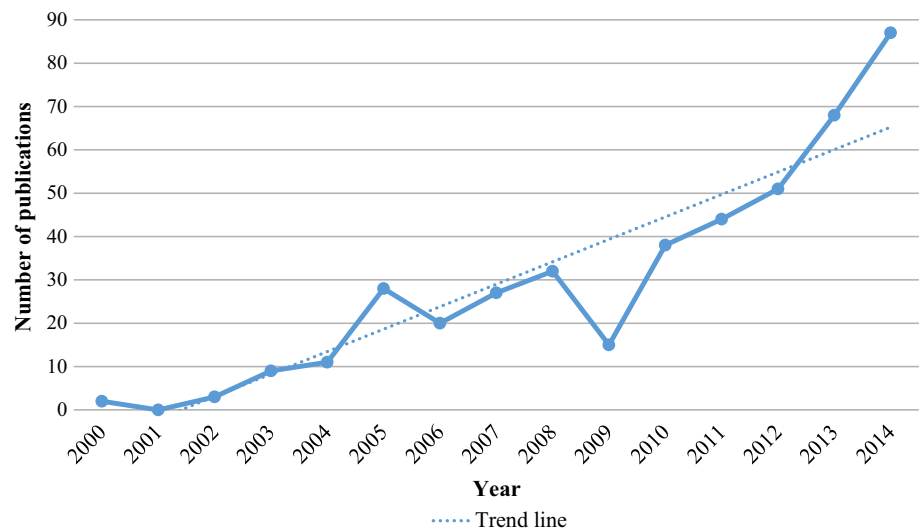
Over the last 15 years, almost one-third of the published claims data studies have focused on health economics [2]. Furthermore, due to its advantages, researchers have previously used this data source for health services research on treatment patterns, under- and overtreatment of specific diseases, epidemiology studies, and as an input factor for modeling studies [4]. The aim of all these studies was to increase transparency and efficiency of the healthcare system by analyzing real-world evidence.

Moreover, the latest national call for proposals confirms that there is both an increase in the importance of and positive development in claims data research. The German government and the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA) established the ‘Innovation Fund’, which will provide funding of 75 million euros per year for projects dealing with health service research over a period of 4 years. In such projects, it is

✉ Sarah Neubauer
sn@cherh.de

¹ Center for Health Economics Research Hannover (CHERH),
Gottfried Wilhelm Leibniz Universität Hannover, Otto-
Brenner-Str. 1, 30159 Hannover, Germany

Fig. 1 Number of publications dealing with German claims data on SHIs over time. Source: based on Kreis et al. [2]



recommended to involve at least one health insurance fund. A recently published report stated that the ‘use and combination of administrative data to improve care’ was the second most popular subject field for project proposals [5]. Furthermore, the current policy reforms of the European Union regarding data protection regulations for member states have identified the potential of this valuable data source and, in contrast to commercial use, do not restrict the requirements for research purposes.

However, despite these advantages, SHI claims data also have specific limitations [6, 7] based on the specific purposes (i.e., billing and reimbursement) of collecting the data. A common restriction of retrospective analyses with German claims data is that detailed clinical data (e.g., disease activity, severity grades of a disease, symptom scores, clinical test results, quality of life data, and documentation of prescribed doses, i.e., ‘days of supply’) are mostly unavailable.

Moreover, one of the biggest challenges for healthcare scientists in Germany is to obtain equal access to claims data for research purposes. For historical reasons, SHI funds differ in the quantity and composition of their members (e.g., socio-economic composition of the insured persons), and not all SHI funds operate nationwide, reducing the comparability of the studied population across SHI members or even the entire German population. Therefore, the choice of the data source could influence the study population and validity of the results. Moreover, access is often only possible with good relations with a health insurance fund, which does not provide equal chances to all researchers. Therefore, to provide equal access opportunities and increase the power and quality of data analyses, many scientific associations in Germany have requested to merge the data of all health insurance funds into one database.

As a first important step towards a merged database, the German Institute of Medical Documentation and Information (DIMDI) implemented a nationwide data pool in 2014. Based on the financial allocation mechanisms for social insurance funds (morbidity-oriented risk structure compensation scheme, hereinafter Morbi-RSA), which should decrease selection risk, this official database includes aggregated healthcare data from SHI funds. According to the German Social Security Code (§§303a to 303e SGB V), only institutions such as health insurance funds, the Federal Joint Committee G-BA, representations of patients, and service providers, as well as institutions for research and healthcare reporting, may use these aggregated data from the DIMDI data pool [8]. In contrast, single health insurance funds provide data at the individual level. Nevertheless, these are not open for commercial use. As the DIMDI data pool includes individual data for approximately 70.65 million German individuals, researchers could use the data for calculating treatment prevalence and performing representative evaluation studies for the publicly insured. Further advantages include analysis options for service providers, low user fees, central point of contact offered by the data pool, and the validation by the prior Morbi-RSA verification.

Thus far, the DIMDI Institute has received approximately 53 applications; of these, the institute granted delivery of aggregated results to 18 applications, and 29 are currently undergoing the review process. The institute rejected or researchers withdrew four applications for technical reasons. As of June 2016, two publications based on the DIMDI data pool have been published [9, 10].

Despite these improvements, in Germany, the data pool still has limitations. This concerns, for example, the limited amount of variables compared with the variables from single SHI funds. Additionally, the original data

transmitted from the SHI funds for the Morbi-RSA do not include specific demographic information, such as insurance start and end (entry/leave) date, type of insurance, and date of death. Some of these are strongly aggregated data, so that some variables contain reduced information (e.g., only for a month or year). Moreover, there are restrictions on information on regional codes (e.g., place of residence and zip code), medical aids, rehabilitation, long-term care insurance, outpatient services, and medical procedures and operations. Furthermore, there are no institutional identifiers for hospitals or physicians. The data are available with a delay of approximately 4 years, and the processing time of project proposals is long and unpredictable.

We can conclude that the possibilities to conduct health services research using the DIMDI database are limited. However, the DIMDI data pool is undergoing continuous structural development, and its enhancement can further extend its importance in claims data analyses [2].

In the US or Canada, where claims data research has a longer history, a significant number of claims databases are accessible to scientists (i.e., those from the American Health Maintenance Organisations, the US Government, or Canadian Provinces) [11]. Moreover, certain research centers even provide free assistance to scientists in obtaining claims data (e.g., by assisting in preparation of requests), as well as in analyzing data sources (e.g., by conducting workshops and offering technical support) [12].

Apart from better access opportunities, the advantages of US and Canadian databases when compared with German claims data include larger population size, longer follow-back and follow-up periods, larger amount of variables, and the possibility to link claims data to internal and external data sources. Furthermore, although data from German SHIs and the DIMDI data pool include information solely on treatments financed directly by the SHI, in the US, all payer claims databases are available, including data derived from public and private payers [13, 14].

To extend the usability of claims data, we need equal chances for all researchers. This means less restriction in accessing claims data for all researchers and other health-care stakeholders, longer observation period (extension of lengths of follow-back and follow-up), and higher amount of useful information for research purposes. Furthermore, we recommend that the data be available nationwide and the amount of variables be extended for the DIMDI data pool. The establishment of the DIMDI data pool is an important step in the direction of a comprehensive data source and increased transparency. Recently, both the importance and capability of linking claims data to internal data (medical record databases and registries) and external data have increased. However, the possibility of data linkage with other important data sources (e.g., pension

fund data, laboratory results, and regional data) is also necessary to close the existing information gap.

As long as the government does not enforce development towards a comprehensive all-payer database, with less limitations and accessibility to researchers, data access in Germany will be inferior to that in most other European (and OECD) countries.

References

1. Onukwugha, E.: Big data and its role in health economics and outcomes research: a collection of perspectives on data sources, measurement, and analysis. *Pharmacoeconomics* **34**, 91–93 (2016). doi:10.1007/s40273-015-0378-4
2. Kreis, K., Neubauer, S., Klora, M., Lange, A., Zeidler, J.: Status and perspectives of claims data analyses in Germany—a systematic review. *Health Policy* **120**, 213–226 (2016). doi:10.1016/j.healthpol.2016.01.007
3. Hoffmann, F.: Review on use of German health insurance medication claims data for epidemiological research. *Pharmacoepidemiol. Drug Saf.* **18**, 349–356 (2009). doi:10.1002/pds.1721
4. Swart, E., Schubert, I., Ihle, P., Robra, B.-P.: Expertise zum Thema: Notwendigkeit des Datenzugangs und der Datentransparenz für ärztliche Körperschaften. Expertise für Bundesärztekammer im Rahmen der Förderinitiative zur Versorgungsforschung. http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/Datenzugang-1.pdf (2010). Accessed 16 June 2016
5. Ärzteblatt Konzepte zur Versorgungsforschung konkurrieren um Mittel aus dem Innovationsfonds. <http://www.aerzteblatt.de/nachrichten/67795> (2016). Accessed 29 June 2016
6. Zeidler, J., Braun, S.: 2 Sekundärdatenanalysen. In: Schöffski, O., Graf von der Schulenburg, J.-M. (eds.) *Gesundheitsökonomische Evaluationen*, pp. 243–274. Springer, Berlin (2012)
7. Schubert, I., Ihle, P., Köster, I., Küpper-Nybelen, J., Rentzsch, M., Stallmann, C., Swart, E., Winkler, C.: Daten für die Versorgungsforschung. Zugang und Nutzungsmöglichkeiten. Datengutachten für das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI). <http://www.dimdi.de/static/de/versorgungsdaten/wissenswertes/datengutachten/dimdi-sekundaerdaten-expertise.pdf> (2014). Accessed 17 June 2016
8. German Institute of Medical Documentation and Information: Information system for health care data (data transparency). <http://www.dimdi.de/static/en/versorgungsdaten/index.htm> (2016). Accessed 17 June 2016
9. Tamayo, T., Brinks, R., Hoyer, A., Kuß, O., Rathmann, W.: The prevalence and incidence of diabetes in Germany. *Dtsch. Arztebl. Int.* **113**, 177–182 (2016). doi:10.3238/arztebl.2016.0177
10. Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen. Krankengeld—Entwicklung, Ursachen und Steuerungsmöglichkeiten, Sondergutachten. http://www.svr-gesundheit.de/fileadmin/GA2015/SVR_Sondergutachten_2015_Krankengeld_Druckfassung.pdf (2015). Accessed 15 June 2016
11. Suissa, S., Garbe, E.: Primer: administrative health databases in observational studies of drug effects—advantages and disadvantages. *Nat. Clin. Pract. Rheumatol.* **3**, 725–732 (2007). doi:10.1038/nprheum0652
12. Hennessy, S., Palumbo Freeman, C., Cunningham, F.: US Government claims databases. In: Strom, B.L., Kimmel, S.E., Hennessy, S. (eds.) *Pharmacoepidemiology*, pp. 209–223. Wiley-Blackwell, Oxford (2012)

13. Porter, J., Love, D., Costello, A., Peters, A., Rudolph, B.: All-payer claims database development manual: establishing a foundation for health care transparency and informed decision making. <https://www.apcdouncil.org/manual> (2015). Accessed 15 June 2016
14. Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project (HCUP). <http://www.hcup-us.ahrq.gov/> (2016). Accessed 16 June 2016