German Pharmacoepidemiological Research Database (GePaRD)

Ulrike Haug & Tania Schink

DOI
10.1007/978-3-030-51455-6_8

Published in
Databases for Pharmacoepidemiological Research

Document version
Accepted manuscript

Online publication date
12 January 2021

Corresponding author
Ulrike Haug

Citation
Haug U, Schink T. German Pharmacoepidemiological Research Database (GePaRD). In: Sturkenboom M, Schink T, editors. *Databases for Pharmacoepidemiological Research*. Cham: Springer International Publishing; 2021. p. 119-24.

German Pharmacoepidemiological Research Database (GePaRD)

Ulrike Haug, Tania Schink

1. Database description

1.1 Introduction

The German Pharmacoepidemiological Research Database **(**GePaRD) is an administrative database based on claims data from statutory health insurance providers in Germany. It was set up for research on the utilization and safety of drugs and vaccines in the real-world setting, but is also used for other purposes such as the utilization of screening tests. In Germany, about 90% of the general population are covered by statutory health insurance (Statista 2020). Membership in statutory health insurance is compulsory but there are exceptions, e.g., for persons with a very high income and for civil servants. These persons can choose private health insurance, i.e., they belong to the 10% of the population not covered by statutory health insurance. Around 75% of these higher-income patients, however, remain voluntary members of statutory health insurance. The health insurance system in Germany is characterised by uniform access to all levels of care and free choice of providers. An advantage of data from German health insurance providers is the stability of their membership which makes long-term follow-up studies feasible. In a pilot database of more than 3.5 million insurants from three statutory health insurance providers, membership was stable in about 75% of all subjects over four years (Pigeot and Ahrens 2008).

1.2 Database characteristics

GePaRD is based on claims data from four statutory health insurance providers and currently (as of January 2020) includes information on about 25 million persons who have been insured with one of the participating health insurances since 2004 or later. Per data year, there is information on approximately 17% of the German population and all geographical regions of the country are represented. The four health insurance providers participating in GePaRD are not equal in size; two of them cover the majority of persons in GePaRD. In previous analyses, the data has been shown to be representative with respect to drug prescriptions (Fassmer and Schink 2014; Schink and Garbe 2010). All persons insured with one of the four health insurance providers contributing to GePaRD entered the database on January 1, 2004 or at the start of insurance with the respective provider (whichever came first) and are followed up until the end of insurance or death. Usually, the database is updated annually and the lag time is about two years.

1.3 Available data

GePaRD contains demographic information such as year of birth, sex, and region of residence as well as information on hospitalizations, outpatient visits, and outpatient drug prescriptions.

Information on hospitalizations includes the date of admission, the admission diagnosis, diagnostic and surgical/medical procedures during the hospital stay, the discharge date, main and secondary discharge diagnoses, and the reason for discharge (incl. death).

Outpatient data include diagnoses as well as outpatient diagnostic and therapeutic procedures and services. Physicians in the outpatient setting are expected to code the disease(s) for which they treat their patients once per quarter (GBA 2019; KBV 2011). Additionally, it is mandatory in the outpatient setting to code the diagnostic certainty. This coding differentiates between “confirmed”, “suspected”, “status post”, and “excluded” diagnoses. Outpatient diagnosis codes are thus available on a quarterly basis only. However, given that an exact date is available for outpatient visits, the diagnosis can be assigned to the date of the visit if there was only one outpatient visit in the respective quarter, i.e., the exact date of diagnosis can partly be determined indirectly.

Hospital and outpatient diagnoses are coded using the International Classification of Diseases, version 10 in the German Modification (ICD-10-GM) with at least four digits; diagnostic and surgical/medical procedures are coded using the Operations and Procedures Coding System (OPS) and outpatient treatment/diagnostic procedures as well as immunizations are coded using claim codes for outpatient services and procedures (*Einheitlicher Bewertungsmaßstab*, EBM).

GePaRD contains information on all drugs prescribed by physicians that were dispensed in a pharmacy and were reimbursed by the health insurance provider. Information on drugs is coded based on the German modification of the Anatomical Therapeutic Chemical (ATC) Classification System. Information on drugs that are purchased over the counter (OTC) is not available in the database. Furthermore, there is no information on medication administered in the hospital, but there are a few exceptions regarding expensive drugs (e.g., monoclonal antibodies).

Outpatient drug data include the dates of the prescription and dispensation, the number of prescribed packages, the specialty of the prescribing physician, and the central pharmaceutical number of the drug. Based on the central pharmaceutical number, information on the generic and brand name of the drug, packaging size, strength, the defined daily dose (DDD), and further pharmaceutical information (e.g., route of administration) is linked to GePaRD.

If lab tests and physical exams were performed, the related information including the date is available in the database provided that they are reimbursable. The results of these examinations or lab tests are not available, but can partly be derived indirectly if specific ICD-10 diagnoses or treatments are coded subsequently to the test or the exam.

There is no lifestyle information in GePaRD. Certain subgroups that have developed diseases due to an unhealthy lifestyle may be identified through diagnosis codes (e.g., obesity, liver diseases due to alcohol abuse) or specific treatments. There is also an ICD-10 code for heavy smoking but it is expected that this information is only in the database if the person was treated for this condition. The socioeconomic status can be approximated through information on the educational level for the majority of persons in GePaRD.

There is information in GePaRD allowing for research on drug safety during pregnancy. Based on the respective codes, algorithms to identify pregnancies and classify pregnancy outcomes (Mikolajczyk et al. 2013; Wentzell et al. 2018), to estimate the beginning of pregnancy (Schink and Haug 2019) and to link mothers with their offspring (Garbe et al. 2011) have been developed.

Linkage of GePaRD to other data sources such as cancer registries is possible for specific research questions if approved by the health insurance providers and the responsible authorities but the required data flow is complex due to strict regulations for data protection (see section 1.5). It is not possible to access patient records for case validation or to contact patients, e.g., for the collection of bio-samples unless the health insurance provider is willing to do this within a research project approved by an ethics committee.

1.4 Strengths and limitations

A major strength of the GePaRD database is its large sample size allowing, for example, investigation of rare exposures and outcomes. In addition, millions of individuals can be followed up over a long period of time given that only a minority of persons in Germany switches between health insurance providers. In the outpatient setting, the data cover the care provided by general practitioners and specialists. While in databases recording only the prescription of drugs it is uncertain whether prescriptions were actually filled, GePaRD only contains information on drugs that were actually dispensed, i.e., this part of primary non-adherence is not an issue in GePaRD. In terms of drug safety in pregnancy, it is advantageous that the beginning of pregnancy can be estimated very precisely and that the majority of newborns can be followed up for many years, i.e., outcomes occurring later in life can also be studied.

There are limitations inherent to the fact that GePaRD is based on claims data. This includes the lack of information on lifestyle factors, lab values, and other measurements (e.g., lung function), overall frailty, the severity of diseases, cause of death, and OTC medication. Particularly in the outpatient setting, miscoding or unspecific coding of diseases may occur, i.e., algorithms combining different types of information are typically applied to define cases. Furthermore, information on the prescribed daily dose is not available in GePaRD; thus the

dose and intended duration have to be estimated and sensitivity analyses have to be performed to check the robustness of the results.

1.5 Validation

As mentioned before, direct validation by linking GePaRD to other data sources for specific research questions is possible but the required data flow is complex due to strict regulations for data protection. Such linkages have successfully been conducted, e.g., to validate the vital status and the date of death in GePaRD (Langner et al. 2019; Ohlmeier et al. 2016).

Other options typically used to plausibilize algorithms developed for the definition of diseases include the comparison of prevalences and incidences with other data sources (indirect validation), e.g., disease registries (Czwikla et al. 2017). Furthermore, for certain outcomes such as acute liver injury, the positive predictive values of algorithms developed based on claims have been determined by applying the algorithm to data from a hospital information system (Timmer et al. 2018). Finally, there is the option of reviewing a random sample of patient profiles including all codes available in GePaRD by clinical experts blinded to the classification of the algorithm.

1.6 Governance and ethical issues

Although GePaRD is maintained by BIPS, the data contained in GePaRD are legally still owned by the respective statutory health insurance providers. Access to the database is granted only to BIPS employees in the context of approved research projects. It is not permitted to give third parties access to the data.

In Germany, the utilization of health insurance data for scientific research is regulated by the Code of Social Law. All involved health insurance providers as well as the Federal Office for Social Security (formerly German Federal (Social) Insurance Office) and the Senator for Science, Health, and Consumer Protection in Bremen as their responsible authorities have to approve the use of GePaRD data for specific research questions. Informed consent for studies based on claims data is required by law unless obtaining consent appears unacceptable and would bias results, which is typically the case in pharmacoepidemiological studies. According to the Ethics Committee of the University of Bremen studies based on GePaRD are exempt from institutional review board review.

1.7 Administrative information

The database is maintained by BIPS and funded by own resources.

Contact details

Organization/affiliation:	Leibniz Institute for Prevention Research and Epidemiology – BIPS GmbH Achterstrasse 30, 28359 Bremen, Germany
Administrative Contact:	Heike Gerds gepard@leibniz-bips.de +49 421 218-56860
Website:	https://www.bips-institut.de/en/research/gepard.html

References

Czwikla J, Jobski K, Schink T (2017) The impact of the lookback period and definition of confirmatory events on the identification of incident cancer cases in administrative data. BMC Med Res Methodol 17 (1):122. doi:10.1186/s12874-017-0407-4

Fassmer A, Schink T (2014) Repräsentativität von ambulanten Arzneiverordnungen in der German Pharmacoepidemiological Research Database (GePaRD). Paper presented at the 9. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi), Ulm, 17.-20. September 2014

Garbe E, Suling M, Kloss S et al. (2011) Linkage of mother-baby pairs in the German Pharmacoepidemiological Research Database. Pharmacoepidemiol Drug Saf 20 (3):258-264. doi:10.1002/pds.2038

GBA (2019) Richtlinie des Gemeinsamen Bundesausschusses zu Untersuchungs- und Behandlungsmethoden der vertragsärztlichen Versorgung (Richtlinie Methoden vertragsärztliche Versorgung). Bundesanzeiger Nr 48 (S 1 523) vom 9 März 2006, zuletzt geändert am 20 Juni 2019 veröffentlicht im Bundesanzeiger (BAnz AT 04092019 B2) in Kraft getreten am 5 September 2019

KBV (2011) Ambulante Kodierrichtlinien – Definition der Behandlungsdiagnose. Dtsch Arztebl International 108 (4):165-167

Langner I, Ohlmeier C, Zeeb H et al. (2019) Individual mortality information in the German Pharmacoepidemiological Research Database (GePaRD): a validation study using a record linkage with a large cancer registry. BMJ Open 9 (7):e028223. doi:10.1136/bmjopen-2018-028223

Mikolajczyk RT, Kraut AA, Garbe E (2013) Evaluation of pregnancy outcome records in the German Pharmacoepidemiological Research Database (GePaRD). Pharmacoepidemiol Drug Saf 22 (8):873-880. doi:10.1002/pds.3467

Ohlmeier C, Langner I, Garbe E et al. (2016) Validating mortality in the German Pharmacoepidemiological Research Database (GePaRD) against a mortality registry. Pharmacoepidemiol Drug Saf 25 (7):778-784. doi:10.1002/pds.4005

Pigeot I, Ahrens W (2008) Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. Pharmacoepidemiol Drug Saf 17 (3):215-223. doi:10.1002/pds.1545

Schink T, Garbe E (2010) Assessment of the representativity of in-patient hospital diagnoses in the German Pharmacoepidemiological Research Database. Abstracts of the 26th International Conference on Pharmacoepidemiology & Therapeutic Risk Management. August 19-22, 2010. Brighton, United Kingdom. Pharmacoepidemiol Drug Saf 19 (s1):S178-S179. doi:10.1002/pds.2019

Schink T, Haug U (2019) Estimation of the beginning of pregnancies ending in stillbirths. Pharmacoepidemiol Drug Saf 28 (s2):236. doi:10.1002/pds.4864

Statista (2020) Anzahl der Mitglieder und Versicherten der gesetzlichen und privaten Krankenversicherung in den Jahren 2013 bis 2019. https://de.statista.com/statistik/daten/studie/155823/umfrage/gkv-pkv-mitglieder-und-versichertenzahl-im-vergleich/. Accessed 3 Jan 2020

Timmer A, Kappen S, de Sordi D et al. (2018) Validity of hospital ICD-10 GM codes to identify acute liver injury. Abstract of the 34th International Conference on Pharmacoepidemiology & Therapeutic Risk Management, 22-26 August 2018.Prague, Czech Republic. Pharmacoepidemiol Drug Saf, vol 27. doi:10.1002/pds.4629

Wentzell N, Schink T, Haug U et al. (2018) Optimizing an algorithm for the identification and classification of pregnancy outcomes in German claims data. Pharmacoepidemiol Drug Saf 27 (9):1005-1010. doi:10.1002/pds.4588