

# BMJ Open Individual mortality information in the German Pharmacoepidemiological Research Database (GePaRD): a validation study using a record linkage with a large cancer registry

Ingo Langner,<sup>1</sup> Christoph Ohlmeier,<sup>2</sup> Hajo Zeeb,<sup>1,3</sup> Ulrike Haug,<sup>1,3</sup> Oliver Riedel<sup>1</sup>

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

► Prepublication history and additional material for this paper are available online. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2018-028223>).

Received 29 November 2018  
Revised 6 June 2019  
Accepted 11 June 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

<sup>2</sup>Health Services Research, IGES Institut GmbH, Berlin, Germany

<sup>3</sup>High-Profile Research Area Health Sciences, University of Bremen, Bremen, Germany

## Correspondence to

Dr Ingo Langner;  
[langner@leibniz-bips.de](mailto:langner@leibniz-bips.de)

## ABSTRACT

**Objective** Claims data need to be validated to assess their use for epidemiological research. This study aimed to examine the validity of mortality information in the German Pharmacoepidemiological Research Database (GePaRD).

**Design** Validation study, secondary data, medical claims.

**Setting** Claims data of two German nationwide acting statutory health insurance providers (SHIs) contributing data for GePaRD; record linkage with epidemiological cancer registry providing individual official mortality information.

**Participants** All women insured with the two SHIs whose insurance coverage ended in the period 2006–2013 and who were residents of North Rhine Westphalia.

**Measures** Descriptive statistics were used to analyse the performance of the linkage procedure. Further, we calculated measures of agreement between the official and the GePaRD-based vital status and assessed differences between the official and the GePaRD-based date of death.

**Results** Of the 256 111 women of the linkage sample, 25 528 were classified as ‘deceased’ in GePaRD and the others as ‘alive’. Compared with the official data, the GePaRD-based vital status showed a sensitivity of 95.9% and a specificity of 99.4%. The negative predictive value was 99.6% and the positive predictive value 94.3%. The date of death agreed in 96.3% between both data sources.

**Conclusions** The vital status recorded in GePaRD was of high accuracy and discrepancies between dates of death in GePaRD and official dates were rare. This underlines the potential of the database for conducting large cohort studies with mortality as the endpoint.

## INTRODUCTION

Mortality is a major endpoint in epidemiological studies, including those analysing the risks and benefits of drug treatment and other interventions. During the past decade, routinely recorded claims data have become a powerful data source for such epidemiological studies. Yet, as these data are collected for non-scientific purposes, validation of the available information with other data

## Strengths and limitations of this study

- We validated the claims data-based mortality information in German Pharmacoepidemiological Research Database (GePaRD) which is a claims database that includes 17% of the annual German population.
- The validation sample included 256 111 women whose insurance coverage ended in the period 2006–2013 either due to termination of membership at the health insurance provider or due to death.
- We calculated sensitivity, specificity and predictive values for the GePaRD-based vital status compared with official mortality information.
- We analysed discrepancies between GePaRD based and official date of death.
- The validation sample was restricted to women aged 25–80 years.

sources is necessary. This particularly applies to mortality information in German claims data as there is no established legal framework for the collection of these data by the statutory health insurance providers (SHIs) and no unequivocal code indicating death in claims data either. The necessary validation approach requires a linkage of different data sources. Studies conducting such a linkage despite the challenges resulting from strict data protection regulations are still lacking.

The German Pharmacoepidemiological Research Database (GePaRD) contains claims data from four SHIs with information on about 17% of the German population. An earlier study<sup>1</sup> compared mortality in GePaRD only on the population level with official rates and only for the year 2006. Recently, mortality information from a small proportion of the persons included in GePaRD was validated on the individual level by a probabilistic record linkage with the local Bremen

Mortality Index.<sup>2</sup> A total of 83.7% of cases from both data sources were successfully linked, and the date of death was found to be accurate in more than 97% of all linked cases. While these findings and the developed underlying linking procedures were novel, the results were also limited for a number of reasons. First, the linkage was of limited representativity as it was conducted only with one local, relatively small SHI with about 232 000 insured persons. Second, over-reporting of deaths could not be ruled out. Third, due to logistic reasons, the observed time period was rather short and could cover two data years only. Fourth, due to the inherent methodology of the selected approach, important validity measures like sensitivity, specificity and negative predictive value for the status of death information could not be determined. Finally, the quality of the data flow could not be assessed due to strict data protection regulations.

In a project addressing the feasibility of GePaRD for the evaluation of the German Mammography Screening Program (MSP), we embedded a linkage study in which we aimed to validate the GePaRD-based vital status and the date of death in more detail and to overcome the limitations of the first validation approach. In the age range 50–80 years, annually, about 10 600 women died due to BC (about 7.2% of all cases of death) in Germany in the period 2006–2013. The goal of the German MSP is to reduce the number of cases of breast cancer death by offering biennially mammography examination to women aged 50–69 years. As evaluation of the MSP is needed and no primary data for mortality-related analyses have been collected collaterally to the MSP, it is intended to use secondary databases like GePaRD instead. However, these data need to be validated. In cooperation with two nationwide-acting SHIs which contribute 95% of the GePaRD data and based on an individual probabilistic record linkage between GePaRD and the epidemiological cancer registry (ECR) of North Rhine Westphalia (NRW), we used a much larger population for the validation of mortality information in GePaRD including deceased and non-deceased women from 8 data years.

## MATERIALS AND METHODS

### Data sources

For the presented analyses, two data sources were used: (1) GePaRD, which has been described elsewhere in more detail,<sup>2–4</sup> is based on claims data from four SHIs in Germany and currently includes information on more than 20 million persons who have been insured with one of the participating providers since 2004 or later. In addition to demographic data, GePaRD contains information on drug dispensations, outpatient and inpatient services and diagnoses. Per data year, there is information on approximately 17% of the general population and all geographical regions of Germany are represented. The core data include start and end date of insurance period as well as the reasons for the end of insurance coverage (eg, death). Inpatient data comprise, among others,

information on the date and reason for admission and discharge. All data contained in GePaRD are pseudonymised and no further person identifiers are included. In this study, only the data of the two large nationwide-operating SHIs were considered for the linkage sample as the two smaller and predominantly locally operating SHIs included virtually no NRW inhabitants. (2) The epidemiological cancer registry of NRW (ECR-NRW), the most populous federal state in Germany with about 17.9 million residents, was considered gold standard information to validate the GePaRD-based vital status and date of death. The ECR-NRW includes pseudonymised data on the official date and cause of death of all deceased inhabitants of NRW (irrespective of cause of death or previous registration as a case of cancer disease) on the individual level. Thus, the ECR-NRW could be characterised, unlike to other German cancer registries, also as a mortality registry.

### Patient and public involvement

Patients were not involved in this secondary data study. It was based on claims data provided by health insurance funds (see also the 'Ethical standard statement' in the section 'Declarations').

### Study population

As the underlying project aimed at evaluating the breast cancer-related mortality in the German MSP, the present feasibility study was restricted to women. Thus, the selected study population comprised all women in GePaRD whose insurance membership ended between 2006 and 2013, who were aged 25 to 80 years at the end of their membership, resided in NRW and were insured in one of the two large SHIs providing data for GePaRD. Based on GePaRD information, the study population was divided into the categories 'deceased' and 'alive'. Death can be coded in GePaRD either as the reason for the end of insurance coverage or for the hospital discharge; all women with one or both of these codes were classified as deceased. In case of discrepancies between the dates of these events, the earlier one was chosen as the date of death. All other women were classified as alive on the date their insurance coverage ended.

Women with an ongoing insurance period beyond 2013 were considered alive at the end of 2013 and were not included in the study population of this validation study.

### Data flow

Data processing and the record linkage procedure have been described in detail elsewhere.<sup>5</sup> In brief, GePaRD records had to be de-pseudonymised, a procedure which involved a trusted third party. The SHIs had to be involved for the addition of person identifiers to the study sample. This step was necessary for the individual record linkage at the ECR-NRW which included names, date of birth and detailed residence information. The SHIs verified whether the place of residence was actually located in NRW. Reidentified inhabitants of NRW formed the

sample to be linked to the ECR-NRW (linkage sample). Using a software tool provided by the ECR-NRW, the person identifiers of this sample were encrypted at the SHIs and sent to the ECR-NRW where a probabilistic record linkage was conducted. The linkage procedure followed methods described by Felligi and Sunter<sup>6</sup> which resulted in the classification as either 'successful', 'incorrect' or 'probable' matches. After this algorithm-based classification, probable matches could be differentiated manually into successful or incorrect matches based on additional clear text characteristics. Finally, the ECR-NRW added mortality information to successful matches and deleted all identifiers supplied by the SHIs. The resulting data set was then merged with the original data set. Cases included in the linkage sample with a successful match in the ECR-NRW records were considered verified cases of death.<sup>5</sup>

### Statistical analyses

The linkage sample included cases of both types of the GePaRD-based vital status: those classified as alive at the end of their insurance period and those classified as deceased at that time. For both groups, we calculated proportions of successful matches in the record linkage with the ECR-NRW data overall and for each SHI. Further, we calculated rates of successful matches among those classified as deceased stratified by the place of death (inside or outside hospital).

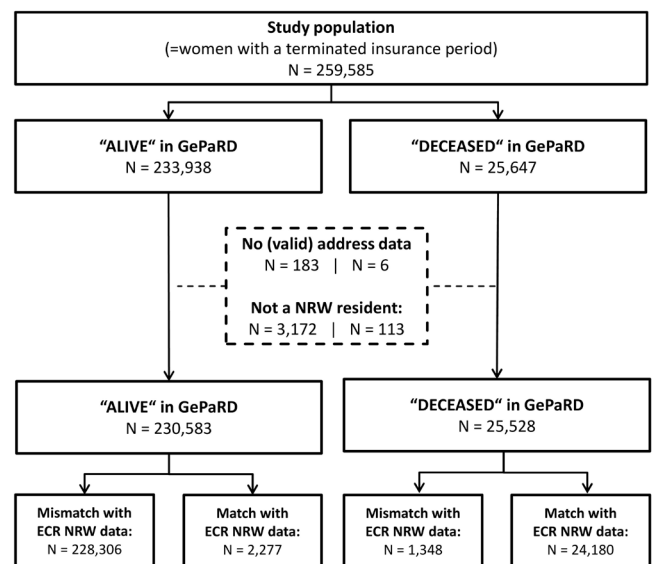
For the subpopulation of cases with a successful match, we calculated the difference between the official date of death and the end date of the insurance period.

The calculation of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for the GePaRD-based vital status 'deceased at the end of the insurance period: yes/no' was compared with a corresponding vital status based on mortality records linked at the ECR-NRW as the gold standard. Since the insurance period in GePaRD data could have ended before the official date of death, the ECR-NRW-based vital status had to be adapted in these cases. Therefore, for individuals who had left the insurance more than 5 days before their matched official date of death, the ECR-NRW-based vital status was set to non-deceased. For all other cases with a successful match, the ECR-NRW-based vital status was set to deceased and correspondingly to not deceased for subjects without a successful match. The validity measures for the GePaRD-based classification as deceased were given as percentages. Corresponding 95% CIs were calculated using the method recommended by Newcombe and Altman.<sup>7</sup>

All analyses were conducted with SAS V.9.3 (© SAS Institute, USA).

## RESULTS

The study population for the validation of mortality information in GePaRD initially consisted of n=259 585 women including n=25 647 women classified as deceased. After



**Figure 1** Flow chart for the study population and the record linkage showing the ascertainment of the linkage sample and the results for the linkage success. ECR NRW, Epidemiological Cancer Registry of North Rhine Westphalia; GePaRD, German Pharmacoepidemiological Research Database.

excluding subjects without valid address data or without NRW residency (see figure 1; see online supplementary table A1 for more details on reasons for exclusion), the GePaRD-based sample consisted of n=230 583 women who were alive and n=25 528 who were deceased. Of the deceased women, 94.72% were successfully linked with the ECR-NRW data (table 1). There was no match in the ECR-NRW for 99.01% of the women classified as alive.

Among those matched with data entries of the ECR (successful matches), there were 69 cases in which a pair of different GePaRD records was linked with the same ECR record. For 67 of these cases, we assume that the GePaRD records of a pair belonged to one and the same person who had changed from one of the included SHIs to the other. There are several facts supporting this assumption. First, in all of these cases, the vital status was classified as

**Table 1** Matches for the GePaRD linkage sample with records of the ECR-NRW

	Match with official mortality registry records at the ECR-NRW				
	No		Yes		All
	N	%	N	%	
Vital status at end of insurance period according to GePaRD					
Alive	228 306	99.01	2277	0.99	230 583
Deceased	1348	5.28	24 180	94.72	25 528
All	229 654	89.67	26 457	10.33	256 111

ECR-NRW, Epidemiological Cancer Registry of North Rhine Westphalia; GePaRD, German Pharmacoepidemiological Research Database.

**Table 2** Distribution of the difference between GePaRD-based and official date of death

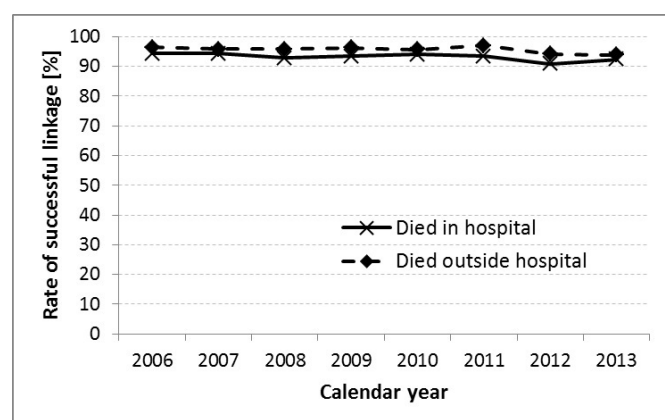
	Vital status at end of insurance period according to GePaRD*			
	Alive		Deceased	
	N	%	N	%
Related to the mortality registry-based date of death, the GePaRD-based date of death was ...				
... more than 1 year earlier	1085	47.62	4	0.017
... more than 1 month to 1 year earlier	153	6.72	16	0.066
... more than 5 days to 1 month earlier	15	0.66	90	0.372
... more than 1 day to 5 days earlier	11	0.48	205	0.848
... 1 day earlier	15	0.66	441	1.824
... the same	616	27.05	23 288	96.311
... more than 1 day to 1 month later	57	2.50	117	0.484
... more than 1 month to 1 year later	319	14.01	17	0.070
... more than 1 year later	6	0.26	2	0.008
All	2277	100.00	24 180	100.00

\*Only cases with a successful match in the linkage at the ECR-NRW.

ECR, epidemiological cancer registry; GePaRD, German Pharmacoepidemiological Research Database; NRW, North Rhine Westphalia; SHI, statutory health insurance provider.

alive in one SHI and deceased in the other one. Second, there were no overlaps in insurance periods, and in most cases, they directly adjoined. Third, in all cases, the insurance period of the record with the vital status classified as deceased followed the one classified as alive.

Matching rates between both data sources were slightly higher for women whose death was coded in hospital data (95.74%) than for those with death coded as the reason for the end of insurance coverage only (93.18%). Matching rates were similar for both SHIs (data not shown) and varied only slightly over the years (figure 2).



**Figure 2** Rates of successful matches with ECR-NRW records for subjects classified as deceased in GePaRD for each calendar year of the study period differentiated by the respective place of death (in or outside hospital). (Abbreviations: GePaRD, German Pharmacoepidemiological Research Database; ECR, epidemiological cancer registry; NRW, North Rhine Westphalia).

Table 2 displays the agreement on date of death between GePaRD and the ECR-NRW data. For those females classified as deceased according to both data sources, the date of death was concordant in 96.31%; the date of death according to GePaRD was earlier in 3.13% and later in 0.49%. For those females classified as alive according to GePaRD but deceased according to the ECR-NRW data (n=2277), the end of insurance coverage was concordant with the official date of death in 27.05%, it occurred earlier in 56.17% (thereof 84.84% more than 1 year earlier) and later in 16.77%, respectively.

When compared with the vital status based on ECR data, the GePaRD-based vital status showed a sensitivity of 95.9% (95% CI 95.7 to 96.2) and a specificity of 99.4% (95% CI 99.3 to 99.4) (table 3). The NPV was 99.6% (95% CI 99.5 to 99.6) and the PPV 94.3% (95% CI 94.0 to 94.6).

**Table 3** Vital status classifications at the end of the insurance period based on GePaRD and based on ECR-NRW records

	Vital status according to ECR-NRW records	Vital status according to GePaRD		
		Alive	Deceased	All
Vital status according to GePaRD	Alive	229 559	1024	230 583
	Deceased	1458	24 070	25 528
All		231 017	25 094	256 111

ECR, epidemiological cancer registry; GePaRD, German Pharmacoepidemiological Research Database; NRW, North Rhine Westphalia.

## DISCUSSION

In the present study, we investigated the validity of mortality-related information in the largest German database with claims data from health insurance providers by conducting a record linkage with a large epidemiological cancer registry in Germany holding comprehensive official mortality data. The data presented here were derived by enhancing a linkage procedure which had been introduced earlier.<sup>2</sup> As was pointed out in the introduction, the original procedure was methodologically limited. The present study aimed at overcoming those limitations.

We found that more than 94% of deceased cases according to the SHI data could be confirmed by official mortality data of the cancer registry and that the date of death according to GePaRD was accurate for 96% of all cases considered. The sensitivity and PPV were 94% and 96%, respectively, while both the specificity and NPV were above 99%. Additionally, we showed that the high accuracy of the vital status varied only slightly over the years and between SHIs.

Many claims databases in other countries are directly linked to official mortality information and thus do not require such a validation approach. Accordingly, the present work provides novel information for settings like Germany where claims data do not include official mortality information. Only few studies on the validation of mortality information in claims data in such settings have been conducted so far. There is a study from Germany comparing claims data-based mortality rates with official national rates on the population level and another study from Japan conducting an internal validation by using different sources of individual information indicating death in the same data source.<sup>1,8</sup> Our previous study<sup>2</sup> which compared mortality information to an external data source on the individual level had some limitations in the temporal and regional coverage as well as in the sample size we were able to overcome in the present analysis. The linkage study as described in this paper was based on two large SHIs, covered a large region of Germany and included 8 data years. Moreover, by including persons who were classified as non-deceased according to GePaRD, we could also determine the sensitivity, specificity and NPV of the mortality information in GePaRD.

Our earlier study<sup>2</sup> resulted in markedly lower successful linkage proportions and thus implied an over-reporting of deaths in GePaRD. At the time of the first study, it was assumed that problems in accurately confirming the place of residence in the federal state of the small sample potentially caused this result. In the current study, 95% of the persons classified as deceased in GePaRD could be linked to the official information on deaths at the ECR-NRW. Given that none of the data sources used in this record linkage could completely rule out data entry errors in their data, a linkage rate of 100% was not expected. About 2%–3% of the records at the ECR-NRW are assumed to contain errors in the identifiers (personal communication with Dr Volker Krieg, ECR-NRW) and it is likely that

a similar error rate is true for the core data from the SHIs. Therefore, the matching rates accomplished in this study seem to be near the practically possible maximum for a probabilistic record linkage based on names, address data and date of birth. Thus, the results of the present study suggest that over-reporting of deaths in GePaRD is unlikely.

In contrast to our previous study,<sup>2</sup> we included all persons with an insurance period ending during the study period who were not classified as deceased. This allowed us to investigate a potential underestimation of deaths in GePaRD as a successful mortality record linkage for those persons would indicate that they were misclassified as alive in GePaRD. Most of the cases classified as alive in GePaRD with a match in record linkage had left the SHI before their official date of death. Therefore, these cases do not represent false negatives. They were correctly classified as alive at the end of the period observable in GePaRD. Overall, misclassification of the vital status in GePaRD was low and resulted from differences between the GePaRD-based and the official date of death. However, for 97% of cases with such a disagreement, the difference was less than 1 month.

As GePaRD comprises only pseudonymised data, the data flow for the actual probabilistic record linkage at the ECR included several merging steps in the form of deterministic record linkages for the de-pseudonymisation and the addition of the necessary identifiers. The procedure was of high quality as only 0.07% of the total sample could not be reidentified by the SHIs. Furthermore, address data were not available at the SHIs for only 0.001% of the sample. Therefore, we can virtually exclude any relevant influence of the deterministic procedures on the matching result.

Our results showed a high accuracy of the vital status in GePaRD and a high precision of the GePaRD-based date of death. Additionally, as records of both involved SHIs showed similarly high matching rates in the record linkage and both SHIs also act nationwide, the accuracy of the vital status and high precision of the date of death in GePaRD can also be assumed for German regions other than NRW. This underlines the potential for a mortality follow-up of large cohorts of persons insured with one of the SHIs contributing data to GePaRD, the largest pharmacoepidemiological research database in Germany.

Some limitations have to be considered when interpreting our work. First of all, due to the research context of this linkage (evaluation of a national mammography screening programme, which is of great public health interest, a precondition for receiving permission from governing authorities to conduct this linkage study), our results are based on women and an age range of 25–80 years. As our study sample did not include men, it cannot be ruled out that unconsidered factors, predominantly occurring in men, might alter the results. However, while we have previously found no sex-specific differences for the matching rates,<sup>2</sup> we find

this rather unlikely. Second, we analysed the validity of the vital status in GePaRD only for persons whose membership in the insurance ended. Of the persons ever recorded in GePaRD between 2006 and 2013, 74.4% had an insurance period that extended beyond 31 December 2013; these persons were not included in this study. However, there are several administrative pathways in Germany that promptly lead to the discontinuation of the payment of the insurance fees in the case of death of an insured person. Even without the information for the reason of the discontinuation, this would result in a termination of the membership of this person. As the SHIs transfer data years to GePaRD with a time lag of about 1 year, membership information for the respective data year is up-to-date at the time of the transfer. Therefore, on the one hand, assigning the vital status of persons with an ongoing insurance period as alive in GePaRD is of high validity which was the rationale for excluding this subpopulation of insured women in this study. On the other hand, our linkage sample was supposed to have included all actual cases of death as we included all women with a terminated membership. With regard to record matching, examinations of the linkage performance gave no indication that a technically correct match between different persons could occur.<sup>5</sup> Third, on switching from one SHI to another, a person will be represented by a new pseudonym in GePaRD and the new data of that person can therefore not be merged with earlier records. Although switching to another SHI is rare, it is possible that our sample included a small number of persons twice for that reason. Actually, there were a few cases where sample individuals from different SHIs were linked to one and the same record at the ECR. In all of these cases, the vital status of these individuals in GePaRD was deceased for one SHI and alive for the other. Further, the corresponding insurance periods did not overlap and the SHI membership of the individual classified as alive was terminated before the official date of death of the linked record of the ECR. Thus, the interpretation of these cases as persons who had switched between the two involved SHIs during the study period appears valid particularly as the homonym error rate—signifying cases where data sets of different persons are incorrectly matched to only one person—for the linkage method used at the ECR was shown to be very low.<sup>9</sup> However, because it is impossible to be insured with more than one SHI at the same time, in these cases, the vital status in GePaRD of the same person would have been classified at different time points and based on information from different SHIs.

## CONCLUSIONS

Overall, our study showed that the vital status recorded in GePaRD was of high accuracy and that both over-reporting and under-reporting of deaths in GePaRD occur only rarely. In addition, regarding the date of

death in GePaRD, in more than 96% of the cases, there was an exact agreement with the official dates. In the rare case of discrepancies, the difference amounted to 5 days or less in 80% of these cases. In summary, information on mortality appears to be of high quality in GePaRD which underlines the potential of the database for conducting large cohort studies with mortality as the endpoint.

**Acknowledgements** We would like to thank Die Techniker (TK) and DAK-Gesundheit, which not only provided data but also collaborated with us in this study. Further, we would like to thank the ECR-NRW for conducting the linkage and for the technical support in preparation of the sample data. The study presented here was conducted as part of a feasibility study for the evaluation of breast cancer-related mortality in the German mammography screening program which was funded by the Federal Office for Radiation Protection (Bundesamt für Strahlenschutz), the Federal Ministry for Environment, Nature Conservation, Building and Nuclear Safety (BMUB), the Federal Ministry of Health (BMG) and the Kooperationsgemeinschaft Mammographie under grant no. UFOPLAN 3610S40002 and 3614S40002. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

**Contributors** IL and CO conceived the study and planned the design. IL and HZ obtained the permission from SHIs and the administrative authority to use the claims data for this study. IL performed the statistical analyses. IL, CO and OR wrote the first draft of the manuscript. UH and HZ provided supervision and critical review of the manuscript. All authors read and approved the final manuscript.

**Funding** The presented study was funded by the Federal Office for Radiation Protection (Bundesamt für Strahlenschutz), the Federal Ministry for Environment, Nature Conservation, Building and Nuclear Safety (BMUB), the Federal Ministry of Health (BMG) and the Kooperationsgemeinschaft Mammographie under grant no. UFOPLAN 3610S40002 and 3614S40002.

**Disclaimer** The contents of this manuscript have never been published and are not under consideration for publication by another journal.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The utilisation of SHI data for scientific research is regulated by the Code of Social Law in Germany (SGB X). The two involved SHIs and their governing authority, the German Federal Insurance Office, approved the use of the data for this study. Informed consent for studies based on GePaRD is not required by law and according to the Ethics Committee of the University of Bremen, these studies are exempt from institutional review board review.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** In accordance with German data protection regulations, access to the data of the German Pharmacoepidemiological Database must not be given to third parties. Furthermore, as we are not the owners of the data, we are not legally entitled to grant access to the data.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Ohlmeier C, Langner I, Hillebrand K, *et al*. Mortality in the German Pharmacoepidemiological Research Database (GePaRD) compared to national data in Germany: results from a validation study. *BMC Public Health* 2015;15:570.
- Ohlmeier C, Langner I, Garbe E, *et al*. Validating mortality in the German Pharmacoepidemiological Research Database (GePaRD) against a mortality registry. *Pharmacoepidemiol Drug Saf* 2016;25:778–84.
- Garbe E, Mikolajczyk RT, Banaschewski T, *et al*. Drug treatment patterns of attention-deficit/hyperactivity disorder in children and

- adolescents in Germany: results from a large population-based cohort study. *J Child Adolesc Psychopharmacol* 2012;22:452–8.
4. Pigeot I, Ahrens W. Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. *Pharmacoepidemiol Drug Saf* 2008;17:215–23.
  5. Langner I, Krieg V, Heidinger O, *et al.* [Enrichment of Claims Data with Official Causes of Death Using a Record Linkage with the Epidemiological Cancer Registry of North Rhine-Westphalia: Feasibility Study and Comparison of Procedures]. *Gesundheitswesen* 2018.
  6. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc* 1969;64:1183–210.
  7. Newcombe RG, Altman DG. *et al* Proportions and their differences. In: Altman DG, Machin D, Bryant TN, Gardner MJ, . eds. *Statistics with confidence*. 2nd edn. Bristol: JW Arrowsmith Ltd, 2001:45–56.
  8. Ooba N, Setoguchi S, Ando T, *et al.* Claims-based definition of death in Japanese claims database: validity and implications. *PLoS One* 2013;8:e66116.
  9. Schmidtman I, Sariyar M, Borg A, *et al.* Quality of record linkage in a highly automated cancer registry that relies on encrypted identity data. *GMS Med Inform Biom Epidemiol* 2016;12.