Accepted Manuscript

Submission Date: 2024-04-12
Accepted Date: 2025-01-17
Accepted Manuscript online: 03.02.2025

Das Gesundheitswesen

Enrichment of health insurance claims data with official death certificate information from three German cancer registries: Proportions of successful linkages and differences by region, year, and age

Franziska Heinze, Ingo Langner, Sebastian Bartholomäus, Martin Meyer, Joachim Kieschke, Kerstin Maaser, Jonas Czwikla.

Affiliations below.

DOI: 10.1055/a-2531-6220

Please cite this article as: Heinze F, Langner I, Bartholomäus S et al. Enrichment of health insurance claims data with official death certificate information from three German cancer registries: Proportions of successful linkages and differences by region, year, and age. Gesundheitswesen Das 2025. doi: 10.1055/a-2531-6220

Conflict of Interest: The authors declare that they have no conflict of interest.

This study was supported by Bundesamt für Strahlenschutz mit Mitteln des Bundesministeriums für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV), des Bundesministeriums für Gesundheit (BMG) und der Kooperationsgemeinschaft Mammographie (KoopG; getragen durch den Spitzenverband der gesetzlichen Krankenkassen und die Kassenärztliche Bundesvereinigung), 3617S42402,3621S42410.

Abstract:
Evaluating breast cancer mortality in the German mammography screening program with health insurance claims data requires the enrichment of claims data with information on causes of death. This work aimed to determine the proportions of successful cause-of-death linkages between the second-largest German statutory health insurance fund and three federal cancer registries and to investigate whether linked proportions differ by region, year, and age. Women aged 40–90 years whose insurance was terminated between 2006 and 2018 were included. Proportions successfully linked to the official death certificate databases of all individuals (available in one registry) and of registered cancer cases (available in three registries) were calculated. Of 150,369 women whose insurance was terminated due to death, 90.0% were linked to the database including all deceased women. Regarding the databases including only registered cancer cases, 35.9% of 150,369, 38.6% of 47,472, and 20.1% of 65,893 deceased women were linked. Linked proportions increased from 2006 to 2018 and peaked in age group 60-69 years. The data will be used for the evaluation of the German Mammography screening program. Since causes of death were not linked for all deceased women and the proportions of linkages differed by region, year, and age, claims-based algorithms will also be considered to enrich claims data with causes of death.

Damit Routinedaten der Gesetzlichen Krankenversicherung für die Evaluation der Brustkrebsmortalität im deutschen Mammographie-Screening-Programm verwendet werden können, müssen diese mit Informationen zur Todesursache angereichert werden. Das Ziel dieser Arbeit war es, die Trefferquoten von Todesursachenabgleichen zwischen der zweitgrößten deutschen gesetzlichen Krankenkasse und drei regionalen Krebsregistern zu bestimmen und zu untersuchen, ob die Trefferquoten nach Region, Jahr und Alter variieren. In die Abgleiche eingeschlossen wurden Frauen im Alter von 40-90 Jahren, deren Versicherungsverhältnis zwischen 2006 und 2018 endete. Trefferquoten wurden in Bezug auf Datenbanken ermittelt, die offizielle Todesursacheninformationen für alle Verstorbenen (verfügbar in einem Krebsregister) bzw. nur für registrierte Krebsfälle (verfügbar in allen drei Krebsregistern) enthalten. Von 150.369 Frauen, deren Versicherungsverhältnis aufgrund von Tod endete, wurden 90,0% mit der Datenbank verlinkt, die Todesursacheninformationen für alle verstorbenen Frauen enthält. Mit den Datenbanken, die Todesursacheninformationen nur für registrierte Krebsfälle enthalten, wurden 35,9% von 150.369, 38,6% von

47.472 und 20,1% von 65.893 verstorbenen Frauen verlinkt. Die Trefferquoten der Abgleiche nahmen von 2006 bis 2018 zu und erreichten ihr Maximum in der Altersgruppe 60-69 Jahre. Die Daten werden für die Evaluation des deutschen Mammographie-Screening-Programms genutzt. Da Todesursacheninformation jedoch nicht für alle verstorbenen Frauen hinzugefügt werden konnten und die Trefferquoten der Abgleiche nach Region, Jahr und Alter variieren, werden zusätzlich Routinedaten-basierte Algorithmen zur Todesursachenanreicherung der Routinedaten eingesetzt.

Corresponding Author:
Dr. Jonas Czwikla, Universität Bremen, SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik, Mary-Somerville-Straße 5, 28359 Bremen, Germany, czwikla@uni-bremen.de

Affiliations:
Franziska Heinze, Universität Bremen, SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik, Bremen, Germany
Franziska Heinze, Universität Bremen, Wissenschaftsschwerpunkt Gesundheitswissenschaften, Bremen, Germany
Ingo Langner, Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Klinische Epidemiologie, Bremen, Germany
[...]
Jonas Czwikla, Universität Bremen, Wissenschaftsschwerpunkt Gesundheitswissenschaften, Bremen, Germany

Introduction

With nearly 10 million deaths in 2020, cancer is a leading cause of death worldwide [1]. Organized screening programs such as mammography screening for the early detection of breast cancer aim to reduce cancer mortality [2, 3, 4]. For the mortality evaluation of screening programs, information on causes of death is needed. Longitudinal health insurance claims data, including individual information on screening participation and cancer diagnoses, represent a potential data source for mortality evaluations but contain no information on causes of death. Therefore, this information needs to be linked from other data sources such as official databases or cancer registries [5]. Alternatively, claims-based algorithms validated against official death certificate information might also be used [6].

In Germany, a mammography screening program (MSP) has been implemented nationwide since 2009 [7]. Whether the program reduces breast cancer mortality is currently evaluated using health insurance claims data, cancer registry data, and screening unit data [7, 8, 9, 10]. The claims data are available from two separate data sources, the German Pharmacoepidemiological Research Database (GePaRD), including 25 million individuals from four health insurance funds, and the BARMER Data Warehouse, including 12.6 million individuals from another insurance fund [11]. Because both data sources include no information on causes of death, this information needs to be linked from cancer registries, the only institutions conducting such linkages for the German MSP evaluation. So far, the proportion of successful cause-of-death linkages has only been determined for a linkage between two insurance funds included in GePaRD and one cancer registry [12]. The proportions of successful linkages are still unclear for other insurance funds and cancer registries.

The purposes of this work were i) to determine the proportions of successful cause-of-death linkages between the BARMER health insurance fund and three federal cancer registries and ii) to investigate whether linked proportions differ by region, year, and age.

Methods

Data sources

BARMER claims data

About 90% of the German population is statutorily (i. e. non-privately) health insured. Of these, almost 9 million people from all over Germany are members of Germany's second-largest statutory health insurance fund, BARMER. For this study, we used BARMER claims data that cover about 10% of the German population [13, 14]. The considered claims data were from

2006 to 2018 and contain information on gender, year of birth, federal state of residence, dates of insurance entry and termination, and the reason for insurance termination (termination due to death or other reasons, such as a switch to another insurance fund or private health insurance). Further data include information on in- and outpatient diagnoses as well as medical procedures and services. The data contain no information on cause of death.

Cancer registry data

The history and legal basis of cancer registries in Germany differ among the 16 German federal states. Beginning in 1926 with the first attempts of cancer registration in the city-state of Hamburg, all federal states had committed to the Federal Cancer Registry Data Act (Bundeskrebsregisterdatengesetz, BKRG, [15]) by 2009. They established individual federal state laws, including quality and data management guidelines, resulting in nationwide coverage with 15 epidemiological cancer registries (Berlin and Brandenburg share their registries) focusing on population-based outcomes and clinical cancer registries focusing on clinical outcomes [16, 17]. All German cancer registries have official death certificate information for registered cancer cases, irrespective of whether they died of cancer or not. The cancer registry in North Rhine-Westphalia (NRW) has additionally full access to population data, including causes of death for the whole deceased population of NRW (i. e. official death certificate information is available for all deceased individuals and not only for registered cancer cases). The cancer registries may use additional death information from periodic linkages with population registers (only date of death) and from hospitals.

For the mortality evaluation of the German MSP, the epidemiological cancer registries of NRW, Lower Saxony, and Bavaria, agreed to cooperate and provide official death certificate information for the linkage with health insurance claims data.

The reporting of this paper followed the RECORD statement [18].

Study Population and Linkage Procedure

In **step 1**, based on pseudonymized BARMER data, women aged 40 to 90 years were selected whose insurance was terminated between 2006 and 2018 and whose last known place of residence was NRW, Lower Saxony, or Bavaria. Besides women whose insurance was terminated due to death (TD), women whose insurance was terminated due to other reasons (TOR) were also considered to evaluate the differentiation between both reasons of insurance termination in the BARMER data. We assigned each selected woman a communication ID for the linkage procedure. Potential cancer diagnoses were not considered in the selection process, and information on whether women were registered in a cancer registry was not available at this stage. In **step 2**, the BARMER – the only party with access to unencrypted

identifying information of insured individuals – re-identified the women selected in step 1 and added encrypted person identifiers according to cancer registry standards. The encrypted information included, i. a., information on name, address, and date of birth. If more than one address was available per woman, the BARMER generated multiple records to potentially increase linkage success. In **step 3**, cancer registries linked the encrypted data to their databases and deleted the encrypted identifiers. In **step 4,** we linked the death information to anonymized claims data via the communication ID. For further information on the approval, data access, and linkage process, see Langner et al. [5].

Statistical analysis

For the analyses, the linkage results were reduced to one record per ID (i. e. duplicates resulting from the linkage procedure were deleted). We calculated for both reasons of insurance termination (i. e. TD or TOR) i) the proportion of women linked to the official death certificate database in NRW, including all deceased individuals, among all women whose insurance was terminated, and ii) the proportions of women linked to the official death certificate databases of registered cancer cases in NRW, Lower Saxony, and Bavaria among all women whose insurance was terminated. These proportions were compared by region (NRW, Lower Saxony, and Bavaria), year (2006 to 2018), and age (40-49, 50-59, 60-69, 70-79, and 80-90 years). In a sensitivity analysis, we restricted the linkage sample to women with a cancer diagnosis (ICD-10 C00-C97) in the claims data up to three years before their insurance termination due to death between 2008 and 2018. Then we calculated the proportions of women linked to the databases of registered cancer cases within this subgroup. For this analysis, hospital discharge and ancillary hospital diagnoses as well as outpatient diagnoses coded as “certain” or “status post” were considered. This analysis was conducted because it can be assumed that with respect to the linkages with the databases including only registered cancer cases, the proportions of successful linkages also depend on the number of cancer cases in the claims data.

Additionally, the agreement between the insurance termination date and the linked official death date was calculated. As Lower Saxony and Bavaria reported only the month and year of death, official death dates in all three federal states were set to the 15th of the reported month. The agreement was measured in two categories: i) difference of <=31 days, and ii) official death date >31 days before insurance termination. Official death date >31 days after insurance termination were ignored/counted as no-matches because we expect that these cases deceased after insurance termination.

Results

For the linkage procedure, 150,369 (TD) and 96,007 (TOR) women in NRW, 47,472 (TD) and 42,007 (TOR) in Lower Saxony, and 65,893 (TD) and 49,566 (TOR) in Bavaria were selected (**Figure 1**). The median age of the women ranged from 81 to 82 years (TD) and 49 to 51 years (TOR) in the three federal states, and hardly varied between 2006 and 2018. The proportions of women with a cancer diagnosis in up to three years before their insurance termination ranged from 45.24% to 45.92% (TD) and 8.39% to 8.93% (TOR), and increased by less than five percentage points in all three federal states between 2008 and 2018. Regarding women with TD, the proportions with cancer increased from the first (40-49 years) to the third (60-69 years) age group, and then decreased to the last age group (80-90 years), whereas regarding women with TOR, the proportions with cancer increased from the first to the last age group in all federal states. The BARMER added encrypted person identifiers for 150,379 (TD) and 96,034 (TOR) records in NRW, 47,491 (TD) and 42,071 (TOR) in Lower Saxony, and 66,490 (TD) and 49,962 (TOR) in Bavaria. In NRW, official death certificate information was provided for 135,291 (TD) and 303 (TOR) records from the database including all deceased women, and 53,958 (TD) and 118 (TOR) records from the database including only registered cancer cases. The latter numbers were 18,337 (TD) and 60 (TOR) in Lower Saxony, and 13,330 (TD) and 41 (TOR) in Bavaria. After 16 (TD) and 0 (TOR) duplicates were deleted, official death certificate information was available for 135,275 (TD) and 303 (TOR) women in NRW among all deceased women. Concerning registered cancer cases, 0 (TD) and 0 (TOR) duplicates in NRW, 11 (TD) and 0 (TOR) in Lower Saxony, and 89 (TD) and 0 (TOR) duplicates in Bavaria were deleted. This resulted in 53,958 (TD) and 118 (TOR) women in NRW, 18,326 (TD) and 60 (TOR) women in Lower Saxony, and 13,241 (TD) and 41 (TOR) women in Bavaria with official death certificate information obtained from the databases including only registered cancer cases.

Proportions of successful linkages by region

In NRW, the proportion of women with TD linked to the database including all deceased individuals, was 89.96% (**Figure 1**), of which 99.92% had a difference of <= 31 days between the insurance termination date and the linked official death date. Regarding the databases including only registered cancer cases, linked proportions were 35.88% in NRW (99.93% with <= 31 days difference), 38.60% in Lower Saxony (99.97% with <= 31 days difference), and 20.09% in Bavaria (99.80% with <=31 days difference). In the sensitivity analysis, linked proportions were 75.04%, 78.71%, and 42.27% in the three federal states.

The proportion of women with TOR linked to the database including all deceased women, was 0.32% in NRW (94.72% with <= 31 days difference). With respect to the databases including only registered cancer cases, the proportions were 0.12% in NRW (93.22% with <= 31 days difference), 0.14% in Lower Saxony (86.67% with <= 31 days difference), and 0.08% in Bavaria (78.05% with <= 31 days difference).

Proportions of successful linkages by year

In NRW, the proportion of women with TD linked to the database including all deceased individuals, increased from 83.14% in 2006 to 93.70% in 2018 (**Figure 2**). Regarding the databases including only registered cancer cases, linked proportions increased from 28.19% to 42.29% in NRW, 34.43% to 44.32% in Lower Saxony, and 18.65% to 20.16% in Bavaria. In the sensitivity analysis, linked proportions increased from 66.38% in 2008 to 80.74% in 2018, 75.51% to 84.71%, and 39.29% to 40.95%, respectively.

The proportion of women with TOR linked to the database including all deceased women decreased from 0.45% to 0.26% in NRW. With respect to the databases including only registered cancer cases, the proportions decreased from 0.19% to 0.12% in NRW and from 0.55% to 0.01% in Lower Saxony, and 0.25% to 0.04% in Bavaria.

Proportions of successful linkages by age

In NRW, the proportion of women with TD linked to the database including all deceased individuals decreased from 95.39% (40-49 years) to 87.70% (80-90 years), whereby the decrease was particularly observed among women >=70 years (**Figure 3**). Regarding the databases including only registered cancer cases, linked proportions increased from the first (40-49 years) to the third (60-69 years) age group, where they were 54.72% (NRW), 56.10% (Lower Saxony), and 32,20% (Bavaria). Subsequently, the proportions decreased to 26.04%, 29.73%, and 14.35% in the last age group (80-90 years). In the sensitivity analysis, linked proportions were 87.43%, 88.74%, and 51.81% in the third age group, and 64.07%, 70.14%, and 35.08% in the last age group.

The proportion of women with TOR linked to the database including all deceased women increased from 0.08% (40-49 years) to 6.91% (80-90 years) in NRW. With respect to registered cancer cases, these proportions increased from 0.02% to 2.34% in NRW, 0.04% to 1.44% in Lower Saxony, and 0.01% to 1.28% in Bavaria.

Discussion

This work determined the proportions of successful cause-of-death linkages between the second-largest German statutory health insurance fund and three cancer registries. We found that 90.0% of women aged 40-90 years whose insurance was terminated due to death were linked to the database of all deceased individuals in NRW. Regarding the databases including only registered cancer cases, linked proportions were 35.9% in NRW, 38.6% in Lower Saxony, and 20.1% in Bavaria. Linked proportions increased from 2006 to 2018 and were highest in age group 60-69 years.

Concerning the linkage with the database including all deceased individuals in NRW, our linked proportion of 90.0% is slightly lower compared to the linked proportion of 94.7% reached in a previous work [12]. This might be explained by the fact that the proportion of linkages is lower among older compared to younger age groups, and our work included more older women than the existing study (40-90 years vs. 25-80 years) [12]. Furthermore, it should be noted that the proportion of linkages in our work increased over time due to data quality improvements and was 93.7% in 2018. The remaining proportion of missing linkages might be due to deviations in the person identifiers available at the different data sources [19]. Furthermore, insurance funds have only the mailing address of insured individuals, which is not necessarily the official registered address documented in the official death certificate. In some cases, it might also be possible that cancer registries have no information on deaths. With respect to the agreement between the insurance termination date and the linked official death date, we observed a difference of <=31 days in 99.9% of the linked cases, which is nearly the same as the 99.8% reported in the previous work [12].

Regarding the linkages with the databases including only registered cancer cases, we observed higher proportions of linked individuals in NRW (35.9%) and Lower Saxony (38.6%) compared to Bavaria (20.1%). This might partly be explained by a lower cancer incidence and cancer mortality among women in Bavaria compared to those in NRW and Lower Saxony (raw incidence and mortality per 100,000 women according to official data [14, 20] in 2013: 527.0 and 221.6 in Bavaria, 622.9 and 267.6 in NRW, and 611.7 and 249.1 in Lower Saxony). However, it should also be considered that in our study population the proportions of women with a cancer diagnosis in up to three years before their insurance termination were almost identical in the three federal states. Furthermore, our sensitivity analysis restricted to women with a cancer diagnosis before their insurance termination also showed higher proportions of linked individuals in NRW (75.0%) and Lower Saxony (78.7%) compared to Bavaria (42.3%), which suggests that the observed differences must also be due to other aspects. A further explanation are differences between cancer registries concerning inclusion of cases with non-melanoma skin cancer (ICD-10 C44). While these cases are registered systematically in NRW

and Lower Saxony, registration in Bavaria is only partial, resulting in a smaller database. Another reason might be different degrees of data quality improvements in the three federal states between 2006 and 2018: While the linked proportion increased by 14.1 (sensitivity analysis 14.4) percentage points in NRW and 9.9 (9.2) percentage points in Lower Saxony, it increased by only 1.5 (1.6) percentage points in Bavaria. Furthermore, due to legal restrictions the cancer registry in Bavaria was unable to perform a linkage between cancer cases and the population register between 2009 and 2016, which resulted in less accurate data on changes of names, places of residence, and deaths. Regarding our main analysis, it must generally be considered that we calculated the proportion of successful linkages among the total population of women with insurance termination due to death because separation of individuals registered in a cancer registry is not possible in claims data. Given that i) according to data of the cancer registry of NRW only 53% of all deaths between 2006 and 2018 among women aged 40 to 90 years were registered cancer cases and ii) deviations in person identifiers and address data, as well as missing death information among registered cancer cases exist, the observed proportions of successful linkages in the linkage with the databases including registered cancer cases were below 50% in all three federal states as expected. However, the proportions of successful linkages decreased from the third to the last age group in all three federal states in both the main and sensitivity analysis, which indicates that deviations in address data might be particularly common among women aged >=70 years. Especially in cases of end-of-life care in institutional settings, deviations between the mailing address of insured individuals and the official registered address documented in the official death certificate are to be expected.

With respect to women whose insurance was terminated due to reasons other than death, the proportions of linked individuals were below 0.4% regarding the database of all deceased individuals in NRW and below 0.2% with respect to the databases including only registered cancer cases in the three federal states. Because the expected proportion was nearly 0%, this indicates a valid claims-based differentiation between insurance terminations due to death and other reasons in the BARMER data. Nevertheless, it must be considered that in claims data in some cases information on deaths is missing in the core data or is only indicated in hospital data as reason for discharge [21]. In our study, the latter was the case for 33.3% of the 303 linked individuals from NRW whose insurance was terminated due to reasons other than death.

Overall, our results show that claims data can be enriched with official death certificate information from cancer registries. The data can be used for studies addressing cancer mortality, such as the evaluation of the German MSP. When conducting such studies, it should be borne in mind that causes of death cannot be linked for all deceased women, and the proportions of linkages differ by region, year, and age. Therefore, claims-based algorithms should also be considered to enrich claims data with causes of death [6, 22].

Strengths and limitations

A major strength of this work is that it provides detailed information on the proportions of successful cause-of-death linkages between the second-largest German health insurance fund and three federal cancer registries for a period covering 12 years. A further strength is that more than 260,000 women whose insurance was terminated due to death as well as nearly 188,000 women whose insurance was terminated due to other reasons were included in the linkage procedure.

There are, however, some important limitations. First, information on all deceased women was available in only one federal state (NRW). This state has, however, the largest population of all German federal states [14]. Second, we used the month and year of death to calculate death dates but were unable to consider the day of death. However, Langner et al. showed that in 99.8% of linked cases, the date of insurance termination and official death was identical (96.5%) or differed by <=31 days (3.5%) [12]. Third, because claims data contain no information on cancer registration, we calculated the proportion of successful linkages among the total population of women with insurance termination due to death. Therefore, the expected proportions of successful linkages with the databases including registered cancer cases were below 50%. Fourth, our study includes only one of the currently 96 German statutory health insurance funds [23] and ignores privately insured individuals, who represent 10% of the German population [24]. Structural differences between insurance funds and statutorily and privately insured individuals may limit the generalizability of our results [25, 26]. The considered health insurance fund is, however, large and insures a sizeable proportion of women. Fifth, we conducted linkages with only three of the 15 German cancer registries. The included registries are, however, from comparatively large federal states representing 46.7% of all women aged 40-90 years. Finally, we conducted probabilistic linkages based on pseudonymized identification variables (control numbers) because a unique person identifier was not available in all data sources. In the future, however, cause-of-death linkages between German insurance funds and cancer registries may also be possible based on health insurance numbers, which should result in higher proportions of successful linkages.

Conclusion

This work showed that claims data of the second-largest German statutory health insurance fund can be enriched with official death certificate information from cancer registries. The linked data will be used for the mortality evaluation of the German MSP. Because causes of death could not be linked for all deceased women and the proportions of linkages differed by region,

increased over time, and was highest in age group 60-69 years, claims-based algorithms will also be used to identify causes of death.

Literature

1. International Agency for Research on Cancer. Global Cancer Observatory: Cancer Today. https://gco.iarc.fr/today; Stand: 10.12.2023
2. Lauby-Secretan B, Scoccianti C, Loomis D et al. Breast-cancer screening–viewpoint of the IARC working group. N Engl J Med 2015; 372: 2353–2358. DOI: 10.1056/NEJMsr1504363
3. International Agency for Research on Cancer. Breast Cancer Screening. IARC Handbooks of Cancer Prevention. Vol. 15. Lyon: International Agency for Research on Cancer; 2016
4. Biesheuvel C, Weigel S, Heindel W. Mammography Screening: Evidence, History and Current Practice in Germany and Other European Countries. Breast Care (Basel) 2011; 6 (2): 104-109. DOI: 10.1159/000327493
5. Langner I, Riedel O, Czwikla J et al. Linkage of Routine Data to Other Data Sources in Germany: A Practical Example Illustrating Challenges and Solutions. Gesundheitswesen 2020; 82 (S 02): S117-S121. DOI: 10.1055/a-0999-5509
6. Langner I, Ohlmeier C, Haug U et al. Implementation of an algorithm for the identification of breast cancer deaths in German health insurance claims data: a validation study based on a record linkage with administrative mortality data. BMJ Open 2019; 9 (7): e026834. DOI: 10.1136/bmjopen-2018-026834
7. Malek D, Kaab-Sanyal V. Implementation of the German Mammography Screening Program (German MSP) and First Results for Initial Examinations, 2005-2009. Breast Care (Basel) 2016; 11 (3): 183-187. DOI: 10.1159/000446359
8. Heinze F, Czwikla J, Heinig M et al. German mammography screening program: program sensitivity between 2010 and 2016 estimated based on German health claims data. BMC Cancer 2023; 23 (1): 852. DOI: 10.1186/s12885-023-11378-0
9. Heinig M, Schafer W, Langner I et al. German mammography screening program: adherence, characteristics of (non-)participants and utilization of non-screening mammography-a longitudinal analysis. BMC Public Health 2023; 23 (1): 1678. DOI: 10.1186/s12889-023-16589-5
10. Fuhs A, Bartholomaus S, Heidinger O, Hense HW. [Evaluation of the impact of the mammography screening program on breast cancer mortality: feasibility study on linking several data sources in North Rhine-Westphalia]. Bundesgesundheitsblatt

Gesundheitsforschung Gesundheitsschutz 2014; 57 (1): 60-67. DOI: 10.1007/s00103-013-1870-7

11. Braitmaier M, Kollhorst B, Heinig M et al. Effectiveness of Mammography Screening on Breast Cancer Mortality - A Study Protocol for Emulation of Target Trials Using German Health Claims Data. Clin Epidemiol 2022; 14: 1293-1303. DOI: 10.2147/CLEP.S376107
12. 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 (7): e028223. DOI: 10.1136/bmjopen-2018-028223
13. BARMER. Geschäftsbericht 2022. https://www.barmer.de/resource/blob/1233990/a6ed08471a1f8ace402b3555c0db905f/barmer-geschaeftsbericht-2022-barrierearm-6202n-data.pdf; Stand: 07.11.2023
14. Gesundheitsberichterstattung des Bundes. Bevölkerung am Jahresende ab 2011 (Grundlage Zensus 2011), (Primärquelle: Statistisches Bundesamt, Fortschreibung des Bevölkerungsstandes). https://www.gbe-bund.de; Stand: 07.11.2023
15. Bundeskrebsregisterdatengesetz vom 10. August 2009 (BGBl. I S. 2702, 2707), das zuletzt durch Artikel 2 des Gesetzes vom 18. August 2021 (BGBl. I S. 3890) geändert worden ist. https://www.gesetze-im-internet.de/bkrg/BKRG.pdf; Stand: 22.02.2024
16. Hundsdorfer G. [Epidemiological cancer registries in Germany: history from a legal point of view]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014; 57 (1): 7-12. DOI: 10.1007/s00103-013-1874-3
17. Katalinic A, Halber M, Meyer M et al. Population-Based Clinical Cancer Registration in Germany. Cancers (Basel) 2023; 15 (15): 3934. DOI: 10.3390/cancers15153934
18. Benchimol EI, Smeeth L, Guttmann A et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015; 12 (10): e1001885. DOI: 10.1371/journal.pmed.1001885
19. Bartholomaus S, Siegert Y, Hense HW, Heidinger O. Secure Linking of Data from Population-Based Cancer Registries with Healthcare Data to Evaluate Screening Programs. Gesundheitswesen 2020; 82 (S 02): S131-S138. DOI: 10.1055/a-1031-9526
20. Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. Atlas der Krebsinzidenz und –mortalität in Deutschland (GEKID-Atlas). Datenstand: Oktober 2023, Lübeck, 2023. www.gekid.de; Stand: 22.02.2024
21. Ohlmeier C, Langner I, Garbe E, Riedel O. Validating mortality in the German Pharmacoepidemiological Research Database (GePaRD) against a mortality registry. Pharmacoepidemiol Drug Saf 2016; 25 (7): 778-784. DOI: 10.1002/pds.4005

22. Gagnon B, Mayo NE, Laurin C et al. Identification in administrative databases of women dying of breast cancer. J Clin Oncol 2006; 24 (6): 856-862. DOI: 10.1200/JCO.2005.02.1790
23. GKV Spitzenverband. Solidarisch. Die gesetzliche Kranken- und die soziale Pflegeversicherung Geschäftsbericht 2022. https://www.gkv-spitzenverband.de/media/dokumente/service_1/publikationen/geschaeftsberichte/GKV_GB2022_barrierefrei_final.pdf; Stand: 17.03.2022
24. Gesundheitsberichterstattung des Bundes. Bevölkerung nach Art des Versicherungsverhältnisses in der Krankenversicherung in 1000. Gliederungsmerkmale: Jahre, Deutschland, Geschlecht, Art des Versicherungsverhältnisses, (Primärquelle: Mikrozensus - Angaben zur Krankenversicherung, Statistisches Bundesamt, Zweigstelle Bonn). https://www.gbe-bund.de; Stand: 17.03.2022
25. Hoffmann F, Koller D. [Different Regions, Differently Insured Populations? Socio-demographic and Health-related Differences Between Insurance Funds]. Gesundheitswesen 2017; 79 (1): e1-e9. DOI: 10.1055/s-0035-1564074
26. Hoffmann F, Icks A. [Structural differences between health insurance funds and their impact on health services research: results from the Bertelsmann Health-Care Monitor]. Gesundheitswesen 2012; 74 (5): 291-297. DOI: 10.1055/s-0031-1275711

List of Figures

Figure 1: Flow chart with proportions of matches for three federal states and by reason for insurance termination. [1]Denominator for displayed percentages in that column; [2]Duplicates are due to different addresses in the BARMER database; TD: Termination due to Death; TOR: Termination due to other reasons

Figure 2: Proportions of matches for the linkage between BARMER health insurance claims data and official death certificate information by year of insurance termination. RCC: Registered Cancer Cases; NRW: North Rhine-Westphalia; BA: Bavaria; LS: Lower Saxony, TD: Termination due to Death

Figure 3: Proportions of matches for the linkage between BARMER health insurance claims data and official death certificate information by age at insurance termination. RCC: Registered Cancer Cases; NRW: North Rhine-Westphalia; BA: Bavaria; LS: Lower Saxony, TD: Termination due to Death

	Federal state of North Rhine-Westphalia				Federal state of Lower Saxony		Federal state of Bavaria	
	Linkage with the database including all deceased women		Linkage with databases including registered cancer cases					
	insurance TD	insurance TOR	insurance TD	insurance TOR	insurance TD	insurance TOR	insurance TD	insurance TOR
1 selected cohort, females aged 40-90	n = 150,369[1]	n = 96,007[1]	n = 150,369[1]	n = 96,007[1]	n = 47,472[1]	n = 42,007[1]	n = 65,893[1]	n = 49,566[1]
2 records of females re-identified by BARMER[2]	n = 150,379	n = 96,034	n = 150,379	n = 96,034	n = 47,491	n = 42,071	n = 66,490	n = 49,962
3 records linked in cancer registries	n = 135,291	n = 303	n = 53,958	n = 118	n = 18,337	n = 60	n = 13,330	n = 41
duplicates	n = 16	n = 0	n = 0	n = 0	n = 11	n = 0	n = 89	n = 0
4 matches	n = 135,275 89.96%	n = 303 0.32%	n = 53,958 35.88%	n = 118 0.12%	n = 18,326 38.60%	n = 60 0.14%	n = 13,241 20.09%	n = 41 0.08%
with a maximum difference of <=31 days	n = 135,167 99.92%	n = 287 94.72%	n = 53,921 99.93%	n = 110 93.22%	n = 18,320 99.97%	n = 52 86.67%	n = 13,214 99.80%	n = 32 78.05%
death date >31 days before termination	n = 108 0.08%	n = 16 5.28%	n = 37 0.07%	n = 8 6.78%	n = 6 0.03%	n = 8 13.33%	n = 27 0.20%	n = 9 21.95%
no-matches or death >31 days after termination	n = 15,094 10.04%	n = 95,704 99.68%	n= 96,411 64.12%	n = 95,889 99.88%	n = 29,146 61.40%	n = 41,947 99.86%	n = 52,652 79.91%	n = 49,525 99.92%