



BMJ Open German primary care data collection projects: a scoping review

Konstantin Moser ,^{1,2} Janka Massag,² Thomas Frese,¹ Rafael Mikolajczyk,² Jan Christoph,³ Joshi Pushpa,¹ Johanna Straube,¹ Susanne Unverzagt ¹

To cite: Moser K, Massag J, Frese T, *et al.* German primary care data collection projects: a scoping review. *BMJ Open* 2024;**14**:e074566. doi:10.1136/bmjopen-2023-074566

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2023-074566>).

Received 10 April 2023

Accepted 31 January 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Medical Faculty of the Martin Luther University Halle-Wittenberg, Institute of General Practice and Family Medicine, Halle, Germany

²Medical Faculty of the Martin Luther University Halle-Wittenberg, Institute of Medical Epidemiology, Biometrics, and Informatics, Halle, Germany

³Medical Faculty of the Martin Luther University Halle-Wittenberg, Junior Research Group (Bio-)Medical Data Science, Halle, Germany

Correspondence to

Konstantin Moser;
konstantin_moser@web.de

ABSTRACT

Background The widespread use of electronic health records (EHRs) has led to a growing number of large routine primary care data collection projects globally, making these records a valuable resource for health services and epidemiological and clinical research. This scoping review aims to comprehensively assess and compare strengths and limitations of all German primary care data collection projects and relevant research publications that extract data directly from practice management systems (PMS).

Methods A literature search was conducted in the electronic databases in May 2021 and in June 2022. The search string included terms related to general practice, routine data, and Germany. The retrieved studies were classified as applied studies and methodological studies, and categorised by type of research, subject area, sample of publications, disease category, or main medication analysed.

Results A total of 962 references were identified, with 241 studies included from six German projects in which databases are populated by EHRs from PMS. The projects exhibited significant heterogeneity in terms of size, data collection methods, and variables collected. The majority of the applied studies (n = 205, 85%) originated from one database with a primary focus on pharmacoepidemiological topics (n = 127, 52%) including prescription patterns (n = 68, 28%) and studies about treatment outcomes, compliance, and treatment effectiveness (n = 34, 14%). Epidemiological studies (n = 77, 32%) mainly focused on incidence and prevalence studies (n = 41, 17%) and risk and comorbidity analysis studies (n = 31, 12%). Only 10% (n = 23) of studies were in the field of health services research, such as hospitalisation.

Conclusion The development and durability of primary care data collection projects in Germany is hindered by insufficient public funding, technical issues of data extraction, and strict data protection regulations. There is a need for further research and collaboration to improve the usability of EHRs for health services and research.

INTRODUCTION

Electronic health records (EHRs) serve as a comprehensive record of a patient's health information, capturing crucial details from each medical visit.¹ While originally created for clinical purposes, EHRs are now widely utilised in epidemiological and clinical

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This scoping review is the first in the literature to conduct a comprehensive literature search in electronic databases, spanning two time points (May 2021 and June 2022). It ensures a thorough overview of primary care data collection projects and research publications in Germany dedicated to extracting data from practice management systems.
- ⇒ The inclusion of 241 studies from six German projects enabled a detailed analysis, revealing significant heterogeneity in terms of project size, data collection methods, and variables collected. This provided valuable insights into the diversity of approaches.
- ⇒ The study effectively identifies and discusses key challenges in primary care data collection projects in Germany, such as the extraction of data from diverse practice management systems, the lack of standardised interfaces, and issues related to data quality.
- ⇒ A limitation of the study is the development of an independent classification system due to the absence of a common method in the literature. This poses a challenge as some publications may have been excluded or misclassified, impacting the accuracy of the analysis.

research, as well as for improving health-care services.^{2 3} Currently, about 36 large routine primary care data collection projects exist globally, in which EHRs are directly collected from practice management systems (PMS). These projects, which allow millions of patients to anonymously contribute data for health sciences, are mainly carried out in English-speaking (UK, USA, and Canada) and European countries. The success and longevity of these projects is influenced by factors such as strong academic and governmental support as well as the use of comprehensive technical facilities for data extraction and analysis.⁴

In Germany, the analysis of EHRs in primary care is largely based on health insurance data rather than primary care data collection projects.⁵ However, health insurance data are primarily recorded for accounting purposes

and lack valuable information such as clinical input data, reasons for encounters, or diagnostic procedures.⁶ Additionally, privately insured patients, which account for approximately 13% of the German population, are often not included in such health insurance databases, potentially leading to selection bias.⁷

Primary care in Germany is predominantly delivered by general practitioners (GPs), but may also encompass any outpatient physician accessible without a referral, irrespective of their specialty.⁸ Between 2002 and 2010, the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung (BMBF)) recognised the importance of family medicine in the improvement of healthcare services and research.⁹ During this time, the ministry also funded two primary care data collection projects, MedVip (Medizinische Versorgung in Praxen) and CONTENT (CONTinuous morbidity registration Epidemiologic NeTwork).¹⁰ However, these projects ended due to limited funding and technical challenges, and a standardised interface for extracting EHRs is still lacking, even though there are over 132 different PMS available on the German market.^{11–13} Despite these challenges, the use of EHRs in outpatient care continues to grow due to the vast amount of data available. In 2020, for example, approximately 688 million outpatient cases were treated by 161 400 outpatient physicians in Germany, representing a ‘real-world data treasure’.¹⁴

EHRs have evolved from their initial purpose of billing to becoming a valuable tool for epidemiological and clinical research.^{2,3} The increasing functionality and quality of EHRs have made them an affordable and accessible data source.¹⁵ In clinical research, for example, EHRs can facilitate patient identification and recruitment, assess study feasibility, and streamline data collection at baseline and follow-up.^{15–17}

The aim of this scoping review is to identify and describe all primary care data collection projects and research publications in Germany dedicated to extracting data from PMS. This might facilitate further research by describing the methodological problems, amplifying possible solutions, and proposing the potential of the projects to inform health policy and practice. To this end, we chose to conduct a scoping review, since our goal is to identify and map study characteristics and not to answer a clinically meaningful question.¹⁸

METHODS

Search strategy

This scoping review follows the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.¹⁹ In order to identify studies relevant for our research question, we explored two electronic databases, Medline (via OVID) and LIVIVO, the latter of which is a German database for life sciences. The search was conducted in May 2021 and updated in June 2022, searching for all records until this time point without any time restrictions. The

search string combined the terms ‘general practice’ with synonyms like ‘family physician’ as well as ‘routine data’. Other terms such as ‘electronic health record’ or ‘Germany’ were included to cover all relevant aspects of our research questions. For each keyword, relevant Medical Subject Headings terms were identified for the Medline exploration. The LIVIVO search was conducted in German with the equivalent terms. When relevant projects were identified, the project names were added to the search string to find further publications. In addition, we searched the project websites and contacted the project’s principal investigators (PIs) using a comprehensive checklist that included a list of publications retrieved by the search to identify any missing project information that was not publicly available. With encouragement from the PI of the IQVIA disease analyser (DA), we also conducted a search on PubMed (National Library of Medicine) using the keywords ‘DA’ and ‘Germany’ to gather all relevant publications from this database, since a considerable number of publications were identified through the PubMed search which were not previously found through the Ovid Medline search. The complete search strategy can be found in the online supplemental table S1.

Inclusion/exclusion criteria

Abstract, title, and subsequently full-texts were reviewed independently by three researchers (KM, JM, and JS) and checked for eligibility. All disagreements were resolved through consensus. If no consensus was reached, a fourth researcher was consulted (SU). We used two online tools for the screening process. Rayyan (<https://www.rayyan.ai/>) was used for title and abstract screening and Covidence (<https://www.covidence.org/>) was used for full-text screening. Both tools allow for each reviewer to decide if the text should be included, excluded, or if it is undecided and to add a reason for this decision. Decisions were blinded until both reviewers were done with the screening. After both reviewers were able to see if they agreed or disagreed on the inclusion of a text.

Studies were eligible if they met the following inclusion criteria: (1) the study population consisted of patients who received treatment from primary care physicians but could also include patients who received care from other specialists who were not considered primary care physicians; (2) use of EHR data that were initially entered into the PMS independently of primary or secondary purpose; (3) EHR data were extracted from PMS and transferred to a database; (4) studies utilising data collected as part of routine clinical practice; and (5) full-text publications in English or German language. The following were excluded: (1) health research studies using primary data, health insurance data, and data from disease registries; (2) conference contributions and publications in languages other than English or German; and (3) studies collecting supplementary data beyond usual care.

Data management

The identified references were downloaded into the reference manager EndNote V.X7.8 where potential duplicates were identified with the respective tool. Duplicates that were not identified by the automated tool due to different spelling were removed manually during the review process.

Data extraction

Information from the retrieved publications was extracted by KM, JM, and JS. JM and JS each reviewed the included publications using a standardised data extraction template created with Microsoft Word. The data were double checked by KM and entered in online supplemental table S2. We extracted information on the following: German primary care data collection projects including general information, data collection methods, data evaluation, and recruitment strategies, and classified

studies as applied studies and methodological studies and categorised type of research into subject area, sample of publications, disease category, or main medication analysed.

Patient and public involvement

None.

RESULTS

We identified 962 references, screened a 291 of those as potentially eligible studies, and included 241 studies conducted with data from six German projects in which databases are filled with EHR from PMS (see [figure 1](#)).

Database characteristics

Four out of six primary healthcare data collection projects are currently active and two have been completed

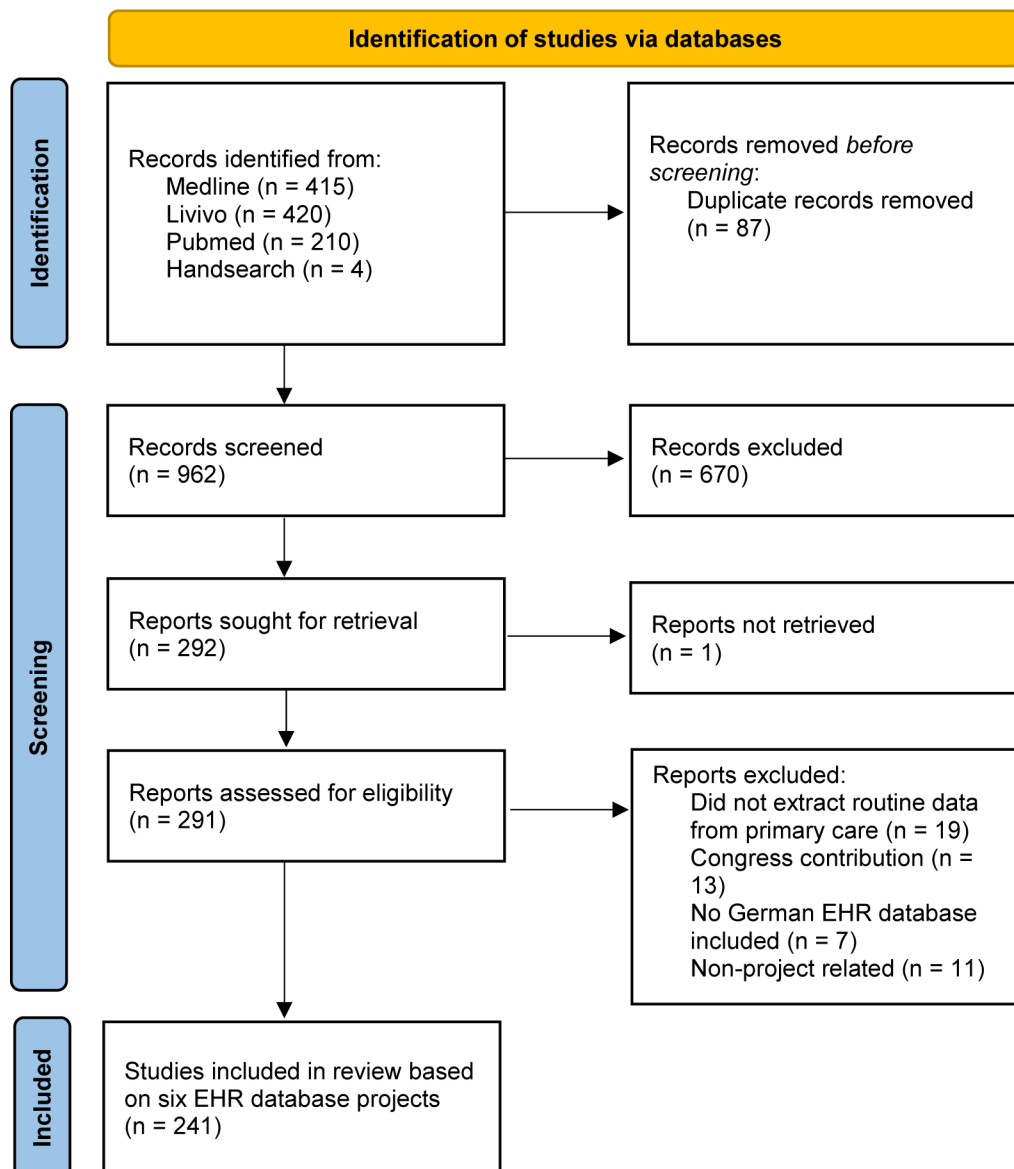


Figure 1 PRISMA 2020 flow diagram for new systematic reviews which included searches of databases only. PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analyses.

(table 1). This overview is sorted by the year in which data collection began.

Of the six, the IQVIA DA is the only German project out of the six identified by this review that is exclusively funded by the pharmaceutical sector. It is specialised in pharmacoepidemiological research and is used as an information system for federal health monitoring.²⁰ Currently, it includes patient records from around 2815 practices, mostly general practices but also including other specialties like cardiology, dermatology, and paediatrics, which are not linked across practices.²¹ With approximately 34 million cases included, it is the largest German primary data collection database and considered to be nationally representative.²²

The other five primary care data collection databases are publicly funded and organised by local academic research groups. Main financiers are the BMBF and the German Research Foundation. The MedVip project aimed to realise first solutions for the use of routine data documentation in the general practice setting. At its peak, a total of 165 practices with approximately 153 000 patient data sets were extracted from 21 different PMS providers. The CONTENT project was based on the International Classification of Primary Care (ICPC) of episodes of care as the primary classification system.^{23 24} Up to 23 practices provided data including approximately 200 000 cases. The project ended because of very high costs and organisational demand. BeoNet (Beobachtungspraxen-Netzwerk)-Hannover was integrated within the German Centre for Lung Research with an initial focus on lung diseases and collects data from approximately 16 practices. Currently, the database includes 343 796 cases.²⁵ RADARplus (Routine Anonymised Data for Advanced Health Services Research plus) aims to develop the infrastructure and technologies, including electronic consent management due to the German data protection regulations, and collects data from seven practices including 100 pseudonymised cases.²¹ BeoNet-Halle is the most recent database and includes anonymised as well as linked pseudonymised data sets from general practices and other types of practices in Germany.²⁶ The database includes 71 911 anonymised and 471 pseudonymised data sets from five practices in Saxony-Anhalt region.

The frequency of data collection by the projects ranges from weekly (BeoNet-Hannover), monthly (DA and BeoNet-Halle), and quarterly (CONTENT), to time points without a fixed interval (MedVip and RADARplus). It is crucial to note that in principle the data export interval can be configured to any desired value, including very short intervals.

Data collection methods

Anonymised data are exclusively collected by the DA and BeoNet-Halle, whereas all other projects except for the DA obtain pseudonymised data. In order to collect pseudonymised data, BeoNet-Hannover, RADARplus, and BeoNet-Halle have instituted informed consent procedures (table 2). RADARplus and BeoNet-Halle employ

an adapted version of the modular Broad Consent, as per the template provided by the Medical Informatics Initiative (MII), allowing for the transfer of identifiable data in compliance with data protection regulations.²⁷ Using Broad Consent, patients have the option to provide consent for various modules, encompassing data collection, processing, scientific utilisation of their patient data, as well as the transfer and scientific use of their health insurance data, along with the possibility for further contact. BeoNet-Hannover has introduced a study-specific consent procedure. The projects exhibit significant heterogeneity in their workflows related to data collection, transfer, and storage, including the integration of trust offices in the cases of RADARplus and BeoNet-Halle.

Three projects (MedVip, BeoNet-Hannover, and RADARplus) extract data using a universal interface (Behandlungsdatentransfer (BDT)). BDT was implemented by the central institute for statutory healthcare to support data exchange between different PMS. The MedVip project has shown the feasibility of data extraction using BDT with various implementations by different software providers. However, its use requires partly that PMS providers assist onsite in extracting the requested data. Despite several updates to the BDT interface, it may still cause inadequate data quality when extracting data from different PMS. Since June 2021, an 'archive and exchange interface' is mandatory in PMS which shall replace BDT. It is based on the interoperability standard HL7 FHIR (Health Level Seven International Fast Healthcare Interoperability Resources), which has gained widespread adoption in the healthcare industry and facilitates interoperability.

The other projects (DA, CONTENT, and BeoNet-Halle) developed their own software solutions to extract predefined data sets. The CONTENT project developed a tailored data extraction software and a modular ICPC software. For BeoNet-Halle, specific exporting modules allow anonymised or pseudonymised data extraction depending on a patient's consent status.

Some projects (DA, CONTENT, BeoNet-Hannover, and BeoNet-Halle) provide training on how to use the software and others provide onsite support to extract data (MedVip and RADARplus). For most projects, data can be uploaded manually by the physician or the research team. Some projects (BeoNet-Hannover and BeoNet-Halle) have also implemented automatic upload to a secure network within the database location. Data validation and integrity checks are run in all projects before data is uploaded to the database and subsequently to an analysis server that can be assessed by researchers. This process is generally facilitated by a database administrator.

Anonymisation and pseudonymisation processes

We could not find publications on specific details of the anonymisation process by the DA. In the case of MedVip, a custom Java programme in doctors' offices removes identifiable BDT fields, except for the patient ID, and

Table 1 Overview of German primary care data collection projects

	IQVIA disease analyser (DA)	MedVip (not active)	CONTENT (not active)	BeoNet-Hannover	RADARplus*	BeoNet-Halle
Funding sources	Private			Public		
Homepage	https://www.iqvia.com/	n.a.	http://content-info.org/	https://www.mhh.de/forschung/beonet	https://generalpractice.umg.eu/forschung/projekte/radarplus/	http://www.beonet.org
Research group	IQVIA Commercial GmbH & Co. OHG	University Medical Centre Goettingen	Department of general practice and health services research, Heidelberg University Hospital	Hannover Medical School and German Centre for Lung Research	University Medical Centre Goettingen	Medical Faculty of the Martin Luther University Halle-Wittenberg
Period of data collection	Since 1992	2002–2010	2003–2014	Since 2016	Since 2016	Since 2020
Included region	Whole Germany	Goettingen and Freiburg	Baden-Wuerttemberg, Hessen, Lower Saxony and Rhineland-Palatinate	Whole Germany	Goettingen	Whole Germany
Frequency of transferring data from PMS to central data collection site	Monthly	No fixed interval (after a practice appointment)	Quarterly	Weekly	No fixed interval (after a practice appointment)	Monthly
Total number of practices (physicians) included (n)	2815 (3540) (November 2022)	165 (n.a.) (May 2008)	23 (41) (March 2014)	16 (27) (March 2023)	7 (n.a.) (February 2022)	5 (40) (February 2023)
Total number of patients (n) per data category	34 million	–	–	–	–	71.911
	Pseudonymised data –	153 000	200 000	343 796†	100	471

The data sources include both published and unpublished sources.
 *Data provided refer to the completed project RADAR, as data from the ongoing project RADARplus are not yet available.
 †The table reflects our findings, although we received contradictory information regarding the process and status of pseudonymisation and obtaining the necessary declarations of consent for this project, so the legal status remains unclear.
 CONTENT, Continuous morbidity registration Epidemiologic NeTwork; n, number; n.a., not available; RADARplus, Routine Anonymised Data for Advanced Health Services Research plus.

Table 2 Data collection methods

	IQVIA disease analyser	MedVip (not active)	CONTENT (not active)	BeoNet-Hannover	RADARplus	BeoNet-Halle
Export types	Anonymous ✓ Pseudonymous –	– ✓	– ✓	– ✓*	– ✓	✓ ✓
Export format	n.a.	BDT	XML	BDT	BDT	CSV
Medium used to upload into the central database	n.a.	Floppy disc or CD send via mail or onsite export	CD, Disc, DVD, email, direct website upload, digital data transfer using GUS box	Internet and secure HTTPS protocol	Via USB into custom software	Internet and secure HTTPS protocol
Import to database	n.a.	Manual	Manual	Automatic	Manual	Automatic or manual
Software details	Not based on BDT interface	Interface for BDT-data export	Modular ICPC classification software	Interface for BDT-data export	Interface for BDT-data export	Universal interface to create a copy of the PMS database
Export from different PMSs (n)	2	PMSs with BDT interface	2	2	PMSs with BDT interface	>70
Database Location	Unknown	Medical Centre Goettingen	Heidelberg University Clinic hospital	Hannover Medical School Location	Medical Centre Goettingen	Martin Luther University Halle-Wittenberg
Database Developer	n.a.	MySQL	n.a.	Postgre SQL	MySQL	Postgre SQL
Graphical user interface	n.a.	Perl	n.a.	PrimeFaces	n.a.	–
Operating language	n.a.	Java	n.a.	Java EE6	n.a.	Python
Linkage to other databases or death records	<ul style="list-style-type: none"> ▶ No linkage to other IQVIA databases ▶ Linkage to death records available in a subgroup of patients (~20%) 	–	–	–	–	–

The data sources include both published and unpublished sources.

*Marks a disagreement between our analysis and the projects principle investigator. The table indicates the statement of the principle investigator. BDT, Behandlungsdatentransfer; CONTENT, CONTinuous morbidity registration Epidemiologic Network; n.a., not available; RADARplus, Routine Anonymised Data for Advanced Health Services Research plus.

encrypts BDT files. For CONTENT, the patient's name is replaced with a unique case number before export. BeoNet Hannover generates automatic pseudonyms from patient IDs for studies, and data are pseudonymised again before leaving the practice, with data processing managed by the data manager. RADARplus follows a privacy-by-design approach, manually documenting consented patients and separating identifiable and medical data. Identifiable data are encrypted and replaced by a pseudonym provided by a trusted third party. For anonymised data, BeoNet Halle assigns unique 35-character keys to patients created from the patient ID which changes from export to export. For pseudonymised data, it creates temporary pseudonyms for consenting patients sent to a trusted third party for generating permanent pseudonyms, allowing data linkage across multiple sources.

Collected variables and data quality

Most projects collect data that are part of health insurance records, encompassing basic patient demographics, diagnoses, drug prescriptions, and billing codes (online supplemental table S3).²⁸

Laboratory tests, such as HbA1c, and health utilisation variables like referrals or hospitalisations, are documented by most projects. Additionally, the majority of ongoing projects (DA, MedVip, BeoNet-Hannover, and BeoNet-Halle) capture essential vital signs, including blood pressure, height, weight, and Body Mass Index, as well as lifestyle-related factors such as smoking status and allergies (DA, BeoNet-Hannover, and BeoNet-Halle). Regarding sociodemographic variables (eg, education and income), number of children, or substance abuse, these variables are not systematically recorded in German PMS. These variables may be entered into structured or free-text fields. To fill this information gap, some projects use standardised questionnaires (BeoNet-Hannover, BeoNet-Halle) given out to patients who consented.

As for the extraction of free-text data, limited information is available, except for BeoNet-Halle, which extracts pseudonymised free text. The MedVip project has partially extracted free-text data due to the absence of data protection regulations during that period.

The CONTENT project can be considered as the only project that attempted to improve data quality at the point of data entry. Several quality circles were implemented and proposed solutions were discussed on a regular basis including training on ICPC-2 coding.

Recruitment strategies

Strategies to recruit GPs and other specialists comprise various financial and non-financial incentives (online supplemental table S4). The DA provides financial incentives of an undisclosed amount, supports practices by using the exporting software, and provides quarterly feedback reports. Its popularity further seems to contribute to its recruitment success.

Publicly funded projects use only some of these recruitment strategies along their project trajectories. Snowball

recruitment is usually implemented at the start of the project to get it running. There have been some 'cold' acquisition attempts (MedVip and RADARplus) including the distribution of circulars, but they were associated with low recruitment rates. Some projects use regular or one-time financial incentives (MedVip, BeoNet-Halle, and CONTENT), while others claim to support practices with establishing a research infrastructure (BeoNet-Hannover, BeoNet-Halle, and CONTENT). Regular feedback reports are provided by some projects (DA, MedVip, CONTENT, and BeoNet-Halle). CONTENT particularly targeted practices with long-term commitment and willingness to code with ICPC. It is also the only project that developed a protected access area where the patients' own data could be accessed. BeoNet-Halle and RADARplus favour practices that integrate consent management.

Applications of the databases

A total of 241 publications were identified (online supplemental table S2). Most articles described applied studies (n=230, 95%) and 5% (n=11) of the articles described methods (figure 2). Methodological studies mainly deal with project-specific issues, such as project descriptions or data collection issues; 30% (n=72) of the studies were industry-funded, while only 9% (n=21) of the publications used data from more than one database. The mean time of recruitment varied from study to study. However, the overall mean time of recruitment across all studies was 7 years in the DA, 4.75 years in MedVip, and 3 years in CONTENT.

Of the 241 publications included, 85% (n=205) were contributed by the DA (figure 2 and online supplemental table S2).

In total, 52% (n=127) of the studies deal with pharmacoepidemiological topics including prescription patterns (n=68, 28%) and studies on treatment outcomes, compliance, and treatment effectiveness (n=34, 14%). Epidemiological studies (n=77, 32%) mainly focused on incidence and prevalence (n=41, 17%) along with risk and comorbidity analysis (n=31, 12%). A small proportion included health services research studies (n=23, 10%) with topics such as hospitalisation.

DISCUSSION

The findings presented in the results section shed light on the landscape of primary care data collection projects in Germany, where databases are populated with EHRs from PMS. In this discussion, we delve into the implications of these findings, drawing comparisons with other countries and addressing key challenges and potential avenues for improvement.

In Germany, one notable challenge arises from the extraction of data from more than 132 different PMS, which currently hinders the uniform consolidation of data for research purposes.^{13 29} Despite the existence of mandatory exchange interfaces, such as BDT or the 'archive and exchange' interface, no discernible improvements

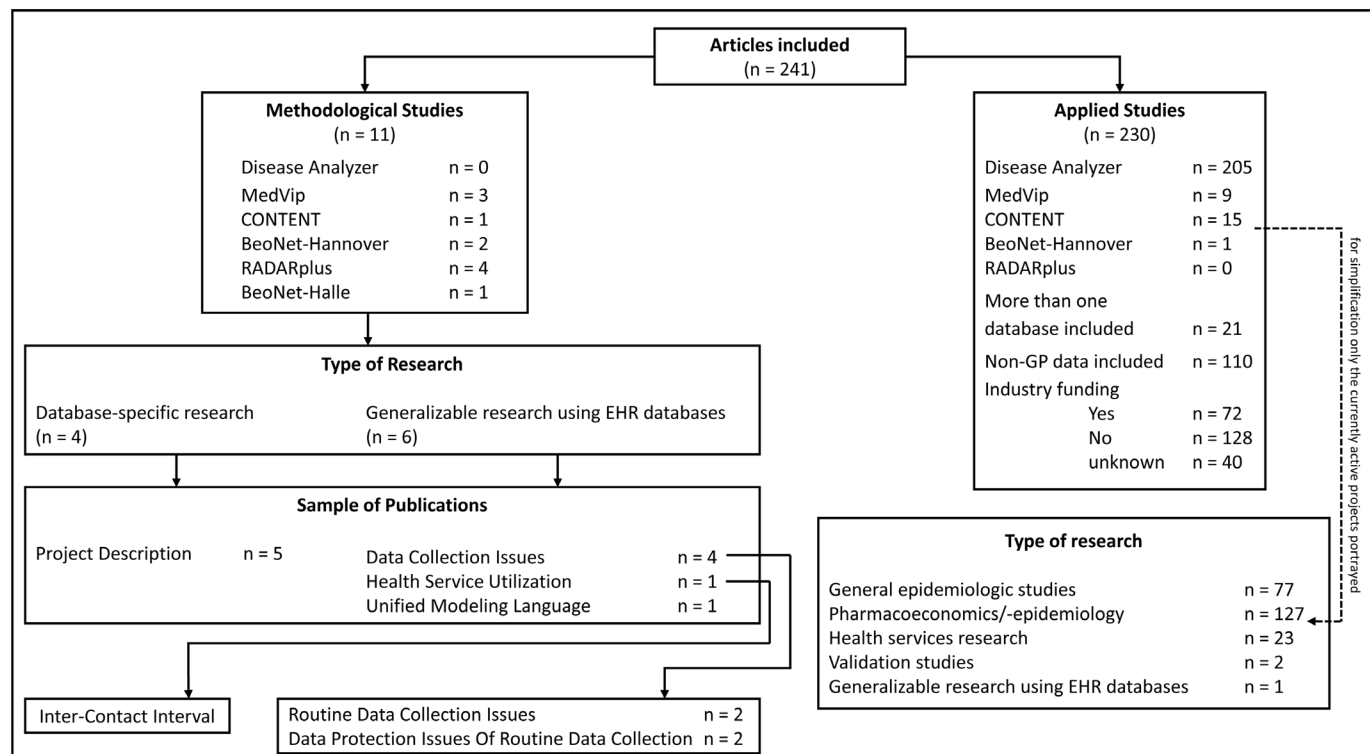


Figure 2 Flow diagram of the extracted articles and their arrangement.

in the ambulatory sector have manifested in this regard. In contrast, the hospital sector boasts well-established standardised interfaces for research.¹¹ The development of standardised interfaces has proven to be a complex and collaborative effort, engaging various stakeholders, including patients, PMS vendors, standards organisations, and academic institutions.^{3,30} Further complicating the situation is the resistance of PMS vendors to external software modifications.³¹

One challenge associated with extracting data from diverse PMS lies in the limited control over the data collection process, thereby compromising the assurance of data quality.³² To illustrate, data may be gathered as part of routine patient care, encompassing information inputted by physicians for primary purposes such as patient care, billing processes, or documentation requirements. Alternatively, data may be collected supplementary to routine care, serving secondary purposes like research, quality improvement, or public health initiatives. The differentiation between these purposes becomes challenging due to the integration of data collected through a complex array of modules and interfaces from various PMS. This complexity is particularly pronounced in cases involving industrial funding, which was evident in a significant proportion of studies (n=72, 30%). It underscores the critical need for transparency and rigour in such studies to maintain scientific integrity, particularly in light of the increasing use of real-world evidence in early benefit assessments of novel therapies.³³

Another challenge in data quality is a predominance of free-text entries in PMS, making complete anonymisation a complex task.³⁴ EHRs encompass structured data,

which is organised, quantifiable, and easily analysable due to its mostly standardised format, and unstructured data, including free-text and images. A comprehensive understanding of a patients' health history necessitates the integration of both types.³ Collaboration with the MII has introduced a Broad Consent concept that allows patients to agree to the scientific use of their data, potentially easing the extraction of free-text information in the future.²⁷ Therefore, informed consent emerges as a vital component for advancing EHR-based research.

The limited progress and short duration of publicly funded projects, as observed in this review, may be attributed to insufficient funding and inadequate government support. Recent projects have received notably meagre funding, especially when compared with government-supported initiatives in other nations.⁴ The initial projects highlighted in this review enjoyed comparatively substantial public funding, indicating the need for sustained investment in healthcare research.⁹ The private funding of the DA by pharmaceutical companies appears to be a contributing factor to its success.

The results indicate that Germany ranks 16th out of 20 analysed countries in terms of EHR implementation. This ranking places Germany behind countries like Sweden, Estonia, and the UK, which have emerged as pioneers in EHR adoption and integration.^{35,36} Therefore, we conclude that the rapid digitalisation of healthcare systems has significantly influenced the development of primary care data collection initiatives.⁴ It is crucial to examine the reasons behind this disparity in EHR adoption and its impact on healthcare research.

Sweden, for example, has efficiently collected and managed patient data through an integrated system including a unique personal identity number, focusing on patient consent and supporting research and quality enhancement.³⁷ Estonia adopted a comprehensive eHealth strategy in 2008, utilising incentives and penalties to establish a cohesive eHealth infrastructure.³⁸ The UK's Clinical Practice Research Datalink stands out as a prominent real-world research service that has contributed data to over 3000 publications, surpassing all German projects combined by more than 12-fold.³⁹ The success of these initiatives can be attributed to factors like opt-out regulations, data quality improvements, and the engagement of healthcare providers.⁴⁰

Our findings, as presented in the results section, also hold implications for the use of databases filled with EHR in healthcare and epidemiological research. The results highlight the versatility of such databases in addressing a wide range of healthcare-related questions, such as evaluating prescription patterns, treatment outcomes, and analysing incidence, prevalence, and comorbidities.

Limitations

One major limitation of this scoping review is incomplete information about some projects. Some information, especially from the DA, is not publicly available due to company confidentiality reasons. A second limitation was mainly identified during the phase of classifying the publications. We developed our own classification system, as we were not able to identify a common classification method in the literature. Some publications listed by the projects' homepages were not included in our final analysis, because we were not able to verify that they included data from PMS. Out of the 241 included publications, we retrieved full-text for 210 papers and extracted information from the abstracts for the remaining 31. Many studies did not describe their study design in detail and might have been classified wrongly. Finally, we only used three literature databases for our investigation, including one database (LIVIVO) that also includes grey literature.

CONCLUSION

The development and sustainability of German primary care data collection projects face several challenges, including limited funding, technical issues related to data extraction, and stringent data protection regulations. Interfaces for data exchange and research remain inadequately implemented. Furthermore, questions regarding data quality and the broad utilisation of ambulatory EHRs for research persist, largely due to the significant amount of information entered in free-text fields. This data can only be partially extracted with patients' informed consent, thereby constraining the range of research publications, primarily focusing on (pharmaco)epidemiological topics derived from a privately funded database. As a result, Germany has yet to fully realise the potential for research made possible by EHRs.

Acknowledgements For proofreading we acknowledge Dawn M Bielawski, PhD.

Contributors KM, JM, and SU developed the methodological concept. KM, JM, and JS screened study titles and abstracts and examined the full texts for inclusion. KM, JM, JS, JC, TF, and JP developed the figures and tables. KM, JM, SU, TF, RM, JP, and JC participated in reading and approving the final manuscript. KM assumes responsibility as the guarantor for overseeing the entirety of the study's content.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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/>.

ORCID iDs

Konstantin Moser <http://orcid.org/0000-0001-8645-4350>

Susanne Unverzagt <http://orcid.org/0000-0002-0108-0415>

REFERENCES

- 1 The Office of the National Coordinator for Health Information Technology (ONC). FAQ: what is an electronic health record. 2023. Available: [https://www.healthit.gov/faq/what-electronic-health-record-ehr#:~:text=An%20electronic%20health%20record%20\(EHR\)%20is%20a%20digital%20version%20of,and%20securely%20to%20authorized%20users](https://www.healthit.gov/faq/what-electronic-health-record-ehr#:~:text=An%20electronic%20health%20record%20(EHR)%20is%20a%20digital%20version%20of,and%20securely%20to%20authorized%20users)
- 2 Modi S, Feldman SS. The value of electronic health records since the health information technology for economic and clinical health act: systematic review. *JMIR Med Inform* 2022;10:e37283.
- 3 Nordo AH, Levaux HP, Becnel LB, et al. Use of EHRs data for clinical research: historical progress and current applications. *Learn Health Syst* 2019;3:e10076.
- 4 Gentil M-L, Cuggia M, Fiquet L, et al. Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. *BMC Med Inform Decis Mak* 2017;17:139.
- 5 Swart E, Gothe H, Hoffmann F, et al. Sonderheft methodische Aspekte der Sekundärdatenanalyse. *Gesundheitswesen* 2020;82:S1-3.
- 6 Schubert I, Köster I, Küpper-Nybelen J, et al. [Health services research based on routine data generated by the SHI. Potential uses of health insurance fund data in health services research]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitschutz* 2008;51:1095-105.
- 7 Akmatov MK, Holstiege J, Steffen A, et al. Utilization of influenza vaccination among chronically ill individuals in Germany: a nationwide claims-based analysis. *Vaccine* 2021;39:952-60.
- 8 Blümel M, Spranger A, Achstetter K, et al. Germany: health system review. *Health Syst Transit* 2020;22:1-272.
- 9 GemeinsamerB. Innovationsfonds. 2022. Available: <https://innovationsfonds.g-ba.de/>
- 10 Bundesministerium für Forschung und Bildung (BMBF). Allgemeinmedizin: Liste der abgeschlossenen Vorhaben. 2022.

- Available: <https://www.gesundheitsforschung-bmbf.de/de/allgemeinmedizin-2624.php>
- 11 Gematik. Informationstechnische Systeme im Krankenhaus. 2022. Available: <https://fachportal.gematik.de/informationen-fuer/isik#c3740>
 - 12 Mach Sv. Praxissoftware: Leichter zum PVS-Wechsel?: hausarzt.digital. 2021. Available: <https://www.hausarzt.digital/praxis/praxisfuehrung/leichter-zum-pvs-wechsel-95031.html>
 - 13 KassenärztlicheB. KBV-Installationsstatistik; 2022.
 - 14 Statista. Anzahl ambulanter ärztlicher Behandlungsfälle und behandelter Personen in Deutschland in den Jahren 2004 bis 2020. 2020. Available: <https://de.statista.com/statistik/daten/studie/75608/umfrage/von-aerzten-behandelte-personen-und-aerztliche->
 - 15 Casey JA, Schwartz BS, Stewart WF, *et al.* Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health* 2016;37:61–81.
 - 16 Kruse CS, Stein A, Thomas H, *et al.* The use of electronic health records to support population health: a systematic review of the literature. *J Med Syst* 2018;42:214.
 - 17 Cowie MR, Blomster JI, Curtis LH, *et al.* Electronic health records to facilitate clinical research. *Clin Res Cardiol* 2017;106:1–9.
 - 18 Munn Z, Peters MDJ, Stern C, *et al.* Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018;18:143.
 - 19 Tricco AC, Lillie E, Zarin W, *et al.* PRISMA extension for scoping reviews (PRISMA-SCR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
 - 20 Gesundheitsberichterstattungdes B. Datenquelle: IQVIATM disease Analyzer. 2023. Available: https://www.gbe-bund.de/gbe/ergebnisse.prc_tab?fid=453&suchstring=&query_id=&sprache=D&fund_typ=DQ&methode=&vt=&verwandte=1&page_ret=0&seite=1&p_sprachkz=D&p_uid=&p_lfd_nr=&p_news=&p_aid=&hlp_nr=&p_janein=
 - 21 Bahls T, Pung J, Heinemann S, *et al.* Designing and piloting a generic research architecture and Workflows to unlock German primary care data for secondary use. *J Transl Med* 2020;18:394.
 - 22 Becher H, Kostev K, Schröder-Bernhardi D. Validity and representativeness of the "disease analyzer" patient database for use in pharmacoepidemiological and pharmaco-economic studies. *Int J Clin Pharmacol Ther* 2009;47:617–26.
 - 23 Laux G, Koerner T, Rosemann T, *et al.* The CONTENT project: a problem-oriented, episode-based electronic patient record in primary care. *Inform Prim Care* 2005;13:249–55.
 - 24 World Organization of Family Doctors' (WONCA). International classification of primary care; 2016. Available: <https://www.globalfamilydoctor.com/site/DefaultSite/filesystem/documents/Groups/WICC/International%20Classification%20of%20Primary%20Care%20Dec16.pdf>
 - 25 Lingner H, Aumann I, Wacker M, *et al.* Health science research with primary care routine data from electronic patient records: the Beonet registry. *Gesundheitswesen* 2018;80:1026–34.
 - 26 Moser K, Mikolajczyk R, Bauer A, *et al.* [Beonet-Halle-development of a Multifunctional database for the automated extraction of healthcare data from general practitioner and specialist practices]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2023;66:569–77.
 - 27 Medizininformatikinitiative. Template text for patient consent forms. 2023. Available: <https://www.medizininformatik-initiative.de/en/template-text-patient-consent-forms>
 - 28 KassenärztlicheB. IT in der Arztpraxis: Datensatzbeschreibung KVDT 2023. 2023. Available: https://update.kbv.de/ita-update/Abrechnung/KBV_ITA_VGEX_Datensatzbeschreibung_KVDT.pdf
 - 29 Sauer J. *Forschung aus der Praxis für die Praxis*. Der Hausarzt, 2023.
 - 30 Justiz Bd. Integration offener und standardisierter Schnittstellen in informationstechnische Systeme. 2022. Available: https://www.gesetze-im-internet.de/sgb_5/_371.html
 - 31 Ayaz M, Pasha MF, Alzahrani MY, *et al.* The fast health interoperability resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. *JMIR Med Inform* 2021;9:e21929.
 - 32 Swart E, Gothe H, Geyer S, *et al.* Good practice of secondary data analysis (GPS): guidelines and recommendations. *Gesundheitswesen* 2015;77:120–6.
 - 33 Schaf F, Thronicke A. Real-world evidence-current developments and perspectives. *Int J Environ Res Public Health* 2022;19:16.
 - 34 Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, *et al.* Secondary use and analysis of big data collected for patient care. *Yearb Med Inform* 2017;26:28–37.
 - 35 Bertram N, Püschner F, Gonçalves ASO, *et al.* Einführung einer elektronischen Patientenakte in Deutschland vor dem Hintergrund der internationalen Erfahrungen. In: Klauber J, Geraedts M, Friedrich J, *et al.*, eds. *Krankenhaus-report 2019: das digitale krankenhaus*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019: 3–16.
 - 36 Amelung VBS, Bertram N, Chase DP, *et al.* *Die elektronische Patientenakte – Fundament einer effektiven und effizienten Gesundheitsversorgung*. Heidelberg: medhochzwei Verlag, 2016.
 - 37 Kajbjerg K, Nordberg R, Klein GO, eds. *Electronic health records in Sweden: from administrative management to clinical decision support. History of nordic computing 3*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
 - 38 World Health Organization. Regional office for Europe Eoohsap. In: Habicht T, Reinap M, Kasekamp K, eds. *Estonia: health system review*. Regional Office for Europe, 2018.
 - 39 Clinical Practice Research Datalink (CPRD). Clinical Practice Research Datalink. 2022. Available: <https://cprd.com/>
 - 40 Clinical Practice Research Datalink (CPRD). Safeguarding patient data. 2023. Available: <https://cprd.com/safeguarding-patient-data>