Received: 7 March 2024 | Revised: 14 August 2024 | Accepted: 16 September 2024

DOI: 10.1002/jrsm.1763

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Development and validation of a geographic search filter for MEDLINE (PubMed) to identify studies about Germany

Alexander Pachanov[1,2] | **Catharina Münte**[1,2] | **Julian Hirt**[3,4,5] | **Dawid Pieper**[1,2]

[1]Faculty of Health Sciences Brandenburg, Brandenburg Medical School (Theodor Fontane), Institute for Health Services and Health System Research, Rüdersdorf, Germany

[2]Center for Health Services Research, Brandenburg Medical School (Theodor Fontane), Rüdersdorf, Germany

[3]Pragmatic Evidence Lab, Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel and University of Basel, Basel, Switzerland

[4]Department of Health, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland

[5]Institute of Health and Nursing Science, Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

Correspondence
Alexander Pachanov, Institute for Health Services and Health System Research, Brandenburg Medical School (Theodor Fontane), Seebad 82/83, 15562 Rüdersdorf bei Berlin, Germany.
Email: alexander.pachanov@mhb-fontane.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 471387773

Abstract

While geographic search filters exist, few of them are validated and there are currently none that focus on Germany. We aimed to develop and validate a highly sensitive geographic search filter for MEDLINE (PubMed) that identifies studies about Germany. First, using the relative recall method, we created a gold standard set of studies about Germany, dividing it into 'development' and 'testing' sets. Next, candidate search terms were identified using (i) term frequency analyses in the 'development set' and a random set of MEDLINE records; and (ii) a list of German geographic locations, compiled by our team. Then, we iteratively created the filter, evaluating it against the 'development' and 'testing' sets. To validate the filter, we conducted a number of case studies (CSs) and a simulation study. For this validation we used systematic reviews (SRs) that had included studies about Germany but did not restrict their search strategy geographically. When applying the filter to the original search strategies of the 17 SRs eligible for CSs, the median precision was 2.64% (interquartile range [IQR]: 1.34%–6.88%) versus 0.16% (IQR: 0.10%–0.49%) without the filter. The median number-needed-to-read (NNR) decreased from 625 (IQR: 211–1042) to 38 (IQR: 15–76). The filter achieved 100% sensitivity in 13 CSs, 85.71% in 2 CSs and 87.50% and 80% in the remaining 2 CSs. In a simulation study, the filter demonstrated an overall sensitivity of 97.19% and NNR of 42. The filter reliably identifies studies about Germany, enhancing screening efficiency and can be applied in evidence syntheses focusing on Germany.

KEYWORDS
bibliographic databases, geographic search filters, literature searching, MEDLINE, PubMed

Highlights

What is already known?

- While validated geographic search filters exist for certain countries like the UK or Spain, none has been developed for Germany until now.

What is new?

- A validated, highly sensitive geographic search filter for MEDLINE (PubMed) designed to retrieve records related to studies about Germany is now available.
- This search filter has the potential to significantly reduce the time and resources required for the screening process.

Potential impact for Research Synthesis Methods readers

- Our search filter can assist evidence synthesis researchers and other users seeking comprehensive retrieval of studies about Germany.

1 | INTRODUCTION

Decision-making in healthcare often relies on the results of systematic reviews (SRs).[1] A cornerstone of a reliable SR is a carefully developed search strategy, which encompasses the core components of a research question, databases to be searched, as well as inclusion and exclusion criteria for primary studies.

Integrating search filters into a search strategy can assist systematic reviewers in achieving maximum efficiency in retrieving relevant literature and during the screening process. Search filters are sets of search terms (searched single words or group of words), combined using appropriate syntax elements (e.g. Boolean operators, field tags, etc.) that aim to retrieve publications with a common feature from bibliographic databases.[2,3] They may focus on achieving high sensitivity (aiming to identify all relevant records), high precision (aiming to minimize retrieval of irrelevant records) or finding a balance between both.[4] Jenkins[5] defines three types (or generations) of search filters that differ in their development and validation methodology. First generation search filters contain search terms that were identified based on the authors' expertise on a topic and their knowledge of a database thesaurus (i.e. subjectively developed search filters). Such search filters do not undergo evaluation of performance measures (e.g. sensitivity, precision). Second generation search filters are filters that were subjectively developed, but their performance was validated using at least one set of relevant records (i.e. a gold standard set; GSS). Third generation search filters are filters whose development was informed using objective methods (e.g. by identifying search terms through a frequency analysis within a GSS) and that have undergone validation using an independent GSS that was not used for filter development. In addition to classification based on the development methods, search filters can also be distinguished by the main objective of their creation, that is, methodological search filters that are designed to identify records with a particular research design (e.g. randomized controlled trials or SRs), and topical search filters that aim to retrieve records focused on a particular subject (e.g. addressing age-related, disease-specific or geographical aspects).[6]

Geographical differences in health outcomes, as well as in the use or provision of health services, are often observed between and within countries.[7–10] Identifying the local context factors that lead to these variations and understanding why it happens can help precisely address various challenges related to health and healthcare. Topic search filters developed to retrieve research with a common geographic feature (hereafter referred to as geographic search filters),[11] are particularly useful in such case. To the best of our knowledge, the current availability of geographic search filters is somewhat limited, and very few of them have been validated. Among the validated geographic search filters are those for Spain,[12] Africa,[13] the United Kingdom,[3,11] German-speaking countries (specific for high-impact factor nursing journals),[14] the group of 37 countries of the Organization for Economic Co-operation and Development (OECD),[2,15] as well as the United States.[16]

The shortcomings resulting from the absence of geographic search filters were discussed in a methodologic review of search strategies applied in SRs to identify health-related studies about Germany.[17] The search strategies were mostly not well elaborated, used inappropriate syntax and missed relevant search terms. Alongside highlighting these limitations, the authors emphasized the need for a geographic search filter that can reliably retrieve research about Germany.

We, therefore, aimed to develop and validate a highly sensitive geographic search filter for MEDLINE (PubMed) that identifies primary studies about Germany. We chose to create a highly sensitive search filter because it reduces the risk of overlooking relevant records, making it suitable for comprehensive evidence syntheses. Furthermore, approaches to yield high precision are quite intuitive and already used in practice.[17]

2 | METHODS

We applied a methodology based on Jenkins' recommendations for the objective development and validation of 3rd generation search filters, combining it with the subjective approach.[5] As illustrated in Figure 1, during the development we generated sets of references to identify relevant search terms; iteratively created the search filter versions and evaluated them. In the validation stage, we selected the final version of the filter based on its performance measures.

2.1 | Development of the search filter

2.1.1 | Generating the gold standard set

We generated our GSS using the relative recall method. This approach involves creating a GSS from records included in evidence syntheses relevant to the topic of the search filter being designed.[18] In the context of our filter, such records are records of studies about Germany. We identified them using SRs included in a recently updated methodological review that investigated search strategies applied in SRs with focus on Germany.[19] Updating the original work[17] allowed us to meet the minimal recommended number of 300 relevant records that should comprise a GSS.[20] We followed the recommended minimum number approach, as there are currently no well-established methods specifically designed to determine the optimal sample size for a GSS.

As our goal was to design a search filter for MEDLINE, the initial step was to verify whether the records of reports included in the SRs with focus on Germany[19] were indexed in MEDLINE. This assessment was conducted by AP and CM (each checked 50% of identified records). A record was considered indexed in MEDLINE if the journal in which its report was published was indexed in MEDLINE at the time of the report's publication. To determine this, we searched the NLM (National Library of Medicine) Catalog.

To be included in the GSS, the records had to meet the following criteria:

1. Primary studies with population of people residing in Germany
2. Studies conducted on humans
3. Data were collected in Germany only

By applying these criteria, our aim was to ensure that the GSS contains records related to reports of studies conducted within Germany. Typically, when the first two inclusion criteria are met, it can be assumed that the third criterion also applies. Nevertheless, we added the third criterion to specifically exclude studies involving individuals residing in Germany who may have been elsewhere during data collection (e.g. surveys conducted at vacation destinations).

We excluded records referring to reports of in vitro studies, animal studies, studies that were not exclusively about Germany and evidence syntheses of any kind, as well as comments and editorials.

AP and CM independently evaluated the records (titles and abstracts) first and then assessed the reports (full texts) of potentially relevant records against the inclusion criteria. Any disagreement was resolved by discussion and, if needed, by involving the third team member (DP).

Generating the 'development' and 'testing' sets

Using the RAND function in Excel, 2/3 of the GSS records were randomly assigned to the 'development set' with the remaining 1/3 comprising the 'test set'.[21] Random assignment was based on each record's unique PubMed identifier (PMID).

2.1.2 | Generating the 'population' set

According to the recommendations of Hausner et al.,[21] search term candidates for an objectively developed search strategy should display higher prevalence in the 'development set' compared to the entire database. To enable this comparison, we generated a 'population set' using a random sample of references from

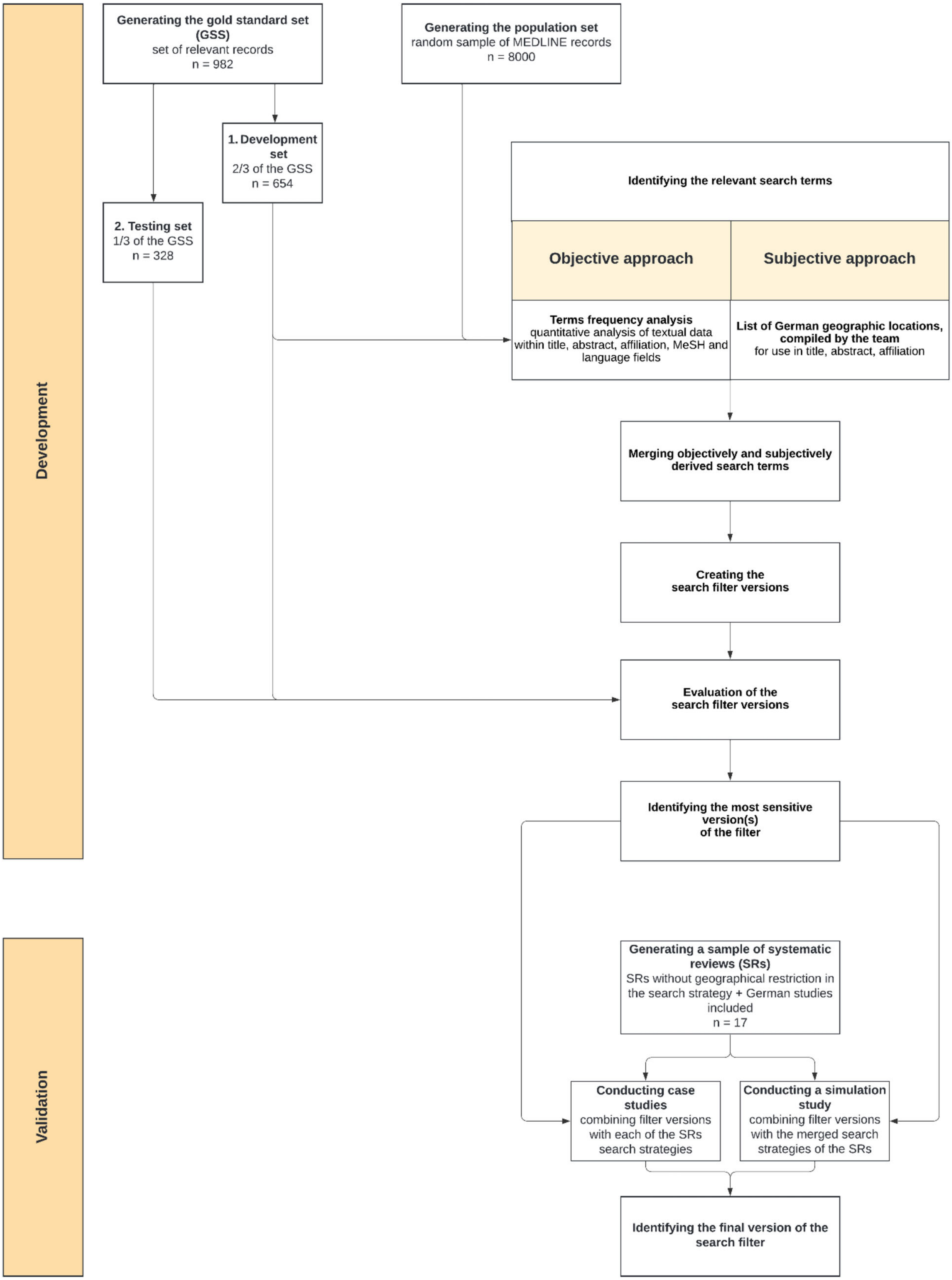

FIGURE 1 Flowchart of the search filter development and validation process.

MEDLINE identified via PubMed. The search was conducted on November 10, 2022 using the *easyPubMed* package for *R*.[22] To account for potential changes in indexing that may have occurred in the database over time, we set the range of publication dates to match that of the GSS. Furthermore, we applied the MEDLINE filter provided by PubMed to limit search results to records indexed in MEDLINE. We exported PMIDs of identified records into a Microsoft Excel spreadsheet, and 8000 of them were randomly assigned to the 'population set' using the RAND function. With a sample size of 8000, the prevalence of a search term within a database was calculated with a margin of error of ±1%.[23]

2.1.3 | Identifying the relevant search terms

Objective approach

We used the PubMed2XL web application[24] to extract data found within MEDLINE/PubMed field tags (e.g. title, abstract, etc.) of the records included in the 'development' and 'population' sets. Extracting these data enables their analyses within individual search fields or their combinations. The data were stored in two Microsoft Excel spreadsheets, with one spreadsheet dedicated to each set of records. To identify relevant search terms, we conducted terms frequency analyses in the 'development' and 'population' sets separately using the *quanteda* package[25] for *R* (version 4.2.1). First, were generated five textual datasets (text corpora) per set for the following PubMed fields: title, abstract, affiliation, language and MeSH (Medical Subject Headings) terms. For each corpus we removed stopwords (e.g. in, the, a, etc.), punctuation, numbers and symbols and then converted all words to lower case.

After creating the corpora, we constructed token objects (i.e. lists of terms) for each field and set: a token object for terms consisted of one word, two words (bigrams) and three words (trigrams). We then converted each token object to a document-term matrix (DTM). A DTM represents the frequency with which terms occur within a set of documents (records). To summarize frequencies of terms occurring in the DTMs we used the 'textstat_frequency' function. We merged DTMs for the fields title, abstract and affiliation into a single DTM. This decision is based on the assumption that terms related to a geographical location occurring in one of these fields, are likely to be pertinent in the other two as well. Analyzing a merged DTM can enhance the process of identifying terms that are relevant across all three search fields. Thus, we generated three DTMs for each set of references: one merged DTM for the title, abstract and affiliation fields, one DTM for the MeSH terms, and one DTM for the language field. Finally, we obtained results from the frequency analyses for single-word terms, bigrams and trigrams for each set of records. This included the separate analysis of the language and MeSH terms fields, and the combined analysis of title, abstract and affiliation fields.

Next, we calculated the prevalence of the search terms within the analyzed fields, as following:

$$\text{prevalence of a search term} = \frac{\text{number of records in a set in which a search term appears}}{\text{total number of records comprising a set}} \tag{1}$$

Subsequently, we calculated the prevalence ratios of the search terms:

$$\text{prevalence ratio of a search term} = \frac{\text{prevalence in the 'development set'}}{\text{prevalence in the 'population set'}} \tag{2}$$

Terms with a prevalence ratio >1 were considered as candidate search terms for the application within the respective search fields of the search filter. Subsequently, AP reviewed all single-word terms to identify those truly relevant for the filter. However, due to the vast number of potential bi- and trigram combinations, only those appearing in at least four records of the 'development set' (representing over 0.5% of the total records) were screened for inclusion in the filter. By focusing on more frequent bigrams and trigrams, we ensured a manageable workload for the manual review process while still capturing potentially relevant multi-word terms.

We translated the relevant terms from German into English (e.g. *Köln* into *Cologne*), and vice versa, if versions in both languages could be relevant for use within the search fields.

Subjective approach

Upon reviewing the results of the term frequency analyses, it became evident that the sample of records in the 'development set' was insufficient to identify all relevant search terms. For example, we determined that the term frequency analyses did not identify locations of some German universities (e.g. Bayreuth). Therefore, we decided to employ additionally a subjective approach for the filter development. Using publicly available sources, we generated a list of geographic locations where, in our opinion, studies are likely to be conducted or where researchers' affiliated institutions are located. This subjectively compiled list included names of the German cities with over than 100,000 inhabitants (*Großstädte*),[26] as well as the locations of university hospitals[27] and higher education institutions.[28] All terms included in the list were deemed relevant search terms for use within the title, abstract and affiliation fields. For the language and MeSH fields, there was no need to apply a subjective approach, as all relevant terms were identified through frequency analyses.

Merging objectively and subjectively derived search terms

We merged the list of terms compiled subjectively by the team with the terms related to geographical locations derived objectively through frequency analyses within the title, abstract and affiliation fields. The terms appearing in the merged list were considered as relevant search terms for use within the title/abstract and affiliation fields.

2.1.4 | Creating the search filter versions

We iteratively created search filter versions using the Boolean operator 'OR' to combine candidate terms and search fields. This was then combined with the MEDLINE filter provided by PubMed using the 'AND' operator. We initially anticipated that frequency analysis within the language field would identify German as a relevant search term. However, we recognized that searching for records in German could also retrieve results from other German-speaking countries such as Austria and Switzerland. Therefore, we investigated what impact excluding the language field would have on filter performance. We assumed that this would not affect sensitivity of the filter, given the extensive use of search terms in other fields. It also has the potential to enhance precision by not identifying irrelevant records.

2.1.5 | Evaluating the search filter versions

We used the 'development set' to test how well the search filter versions perform in retrieving records used for their development (a process also known as internal validation).[5] Using the 'testing set' we evaluated the generalizability of the search filter versions (i.e. an assessment of the ability to retrieve records that were not used for the development).[11] Since the 'development' and 'testing' sets exclusively comprised relevant references, we were able to assess only the sensitivity of the search filter. Sensitivity was calculated as the proportion of records identified by a filter version to the total number of references within a set of records (multiplied by 100 to express as percent)[5]:

$$\text{Sensitivity} = \frac{\text{number of relevant records identified by a filter version}}{\text{total number of records included witihin a set}} \times 100 \quad (3)$$

2.2 | Validation of the search filter

The validation of the search filter requires applying it to a set of records that were not part of the initial GSS used during the filter's development.[5] Therefore, the versions of the filter that demonstrated the highest sensitivity in the 'testing set' underwent validation through case studies (CSs). A CS in this context refers to the application of the search filter to the original search strategy of an SR (including the time period of the original search) that did not geographically restrict their search strategy. Comparing the original search results to the results obtained with a search filter allows for the assessment of the reduction in the number of results that would have needed to be viewed if the filter had been used. We conducted CSs by applying the search filter versions to the original search strategies of SRs (labeled so by the authors) that met the following criteria:

1. The original search strategies were not restricted to any geographic location

2. They included at least five reports of studies about Germany (either exclusively about Germany or about Germany and other countries) that were indexed in MEDLINE and were not part of the 'development set'
3. Reproducible search strategy for PubMed is reported
4. Access to full-text

It is worth noting that the criteria and the approach of identifying eligible SRs, differ from those established at the beginning of the project. First, we considered 10 SRs that met the first inclusion criterion among those included in the updated methodological review.[19] Since validating the search filter requires records not used for its development, records related to reports of studies included in these SRs were not a part of the GSS. Second, the criterion number 2 differed from the one presented here. Originally, we set a minimum number of included reports of studies about Germany and indexed in MEDLINE (i.e. relevant studies) to 10. After AP screened the initially considered SRs, none of them was found eligible for CSs, as none met all the inclusion criteria simultaneously. Consequently, we needed an alternative set of SRs eligible for CSs. Therefore, between February and August 2023, AP searched PubMed and LIVIVO to identify such SRs (the search strategy is presented in Appendix S1). Due to challenges in finding SRs meeting the initial target of 10 included relevant studies, we adjusted this number to five, while simultaneously increasing the minimum number of SRs used for conducting CSs to 15. We did not verify geographic settings of relevant studies included in these SRs.

Conducting CSs allows to assess both the search filter sensitivity and precision because the search results contain relevant as well as irrelevant records.[11] Precision of a search filter is defined as the proportion of relevant records identified with the search filter to the total number of records identified with the search filter[5]:

$$\text{Precision} = \frac{\text{No. of relevant records identified with the search filter}}{\text{No. of all identified records with the search filter}} \times 100 \qquad (4)$$

Furthermore, we calculated the number-needed-to-read (NNR) as an additional measure of the search filter efficiency. NNR reflects the number of records needed to be reviewed to identify a relevant one[29] and is calculated as:

$$\text{NNR} = \frac{1}{\text{Precision}} \qquad (5)$$

We additionally present sensitivity, precision and NNR derived from a simulation study, where we developed a comprehensive search strategy to retrieve all records from the CS SRs (simulating a single, larger evidence synthesis). By doing so, we obtained a larger, aggregated set of records (both relevant and irrelevant) for validating our filter, compared to conducting individual CSs with smaller sample sizes. Therefore, the results from the simulation study can provide more precise estimates of the filter's performance measures.

In developing and validating our search filter, we prioritized achieving high sensitivity. However, there is no universally accepted definition of "high" sensitivity, with previous research suggesting a threshold of 90% or above,[20] we adopted this criterion for our study. Consequently, we considered filter versions with a median sensitivity of at least 90% within the CSs and a sensitivity of at least 90% in the simulation study to be highly sensitive.

3 | RESULTS

3.1 | Development of the search filter

3.1.1 | Generating the gold standard set

As depicted in Figure 1, the GSS comprised of 982 records from 57 SRs (Appendix S2), substantially exceeding the minimal required number of 300 records[20] and covering publication dates from May 1, 1992 to September 28, 2021. We assigned 654 records to the 'development set' and allocated the remaining 328 records to the 'testing set'.

3.1.2 | Generating the population set

Using the *easyPubMed* package for *R* to search PubMed, we retrieved 19,545,686 records. Out of these, 8000 randomly selected records formed the 'population set'.

3.1.3 | Identifying the relevant search terms

Objective approach

The term frequency analyses within the title, abstract and affiliations fields resulted in 7518 potentially relevant single-word search terms and 205,447 bi- and trigrams (Appendices S3.1 and S3.2 respectively). In total, all single-word terms and 8409 multi-word terms were reviewed to identify truly relevant terms for the filter. Overall, we classified 221 single-word and 23 bi- and trigrams as truly relevant search terms. These terms encompass full names as well as abbreviations of geographical locations, research institutions, health insurance funds, professional organizations and hospitals.

The terms frequency analyses within the MeSH terms field resulted in 1355 MeSH terms with a prevalence ratio of >1 (Appendix S3.3). After reviewing them, four terms were classified as relevant: *Germany*; *Germany, West*; *Germany, East* and *Berlin*. However, *Berlin* was not included in the filter. The MeSH term *Berlin* is arrayed hierarchically below the MeSH term *Germany*. This means that when searching for the MeSH Term *Germany*, records tagged with both MeSH terms, *Germany* and *Berlin* will be found.

Within the language field, the terms frequency analyses revealed that records written in German were 25 times more prevalent in the 'development' than in the 'population' set (Appendix S3.4). Therefore, for use within the language field the term *German* was considered a relevant search term.

Subjective approach

The subjectively compiled list of German geographical locations comprised 251 search terms related to all higher education institutions, university hospitals, large cities and federal states across Germany.

Merging objectively and subjectively derived search terms

We complemented the list of subjectively derived search terms for geographical locations with the terms identified through the term frequency analyses in the title, abstract and affiliation fields. This resulted in the addition of 38 extra search terms related to geographical locations that were exclusively identified via the term frequency analyses. Next, we decided to remove certain terms from the merged list. For instance, in the case of *Brandenburg* (federal state) and *Brandenburg an der Havel* (city), we kept only the term *Brandenburg*. This decision was made because searching for *Brandenburg* would yield results for both the federal state and the city. In total, the final merged list contained 280 terms related to German geographical locations, which were used in the title/abstract and affiliation fields of the filter.

3.1.4 | Creating the search filter versions

First, we created a comprehensive full version of the filter that included all terms identified in the corresponding search fields through both objective and subjective approaches. The same set of terms applied to two fields: title/abstract and affiliation. However, an exception was introduced for the terms *GDR* and *FRG* (acronyms denoting the German Democratic Republic and the Federal Republic of Germany respectively). Since these countries reunified in 1990, we assumed that thereafter authors' country affiliation would, in the vast majority of cases, be indexed as Germany only. Therefore, we limited the publication date for these terms within the affiliation field to up to 1990. This was done to reduce the number of records that could be retrieved using terms *GDR* and *FRG*, which could potentially denote acronyms not related to Germany. The second version of the filter mirrored the full version, except for the exclusion of the language field. As described in Section 2, this is to reduce retrieval of records about other German speaking nations such as Austria or Switzerland.

Consequently, in the next phase, we evaluated the performance of two versions of the search filter: the full version and its counterpart without the language field. Both versions are shortly described in Table 1 and fully available in Appendix S4.

3.1.5 | Evaluating the search filter versions

Both iterations of the filter demonstrated a sensitivity of 99.85% within the 'development set' and 100% within the 'testing set' (Table 2). The filter versions failed to identify one record from the 'development set'.[30] The only term related to Germany this record had was *German Competence Network Heart Failure* in the corporate author field that was not included in the filter versions.

Considering these findings, no adjustments could be made within the fields included in the filter versions to identify this particular record. Therefore, both versions without amendments were validated in the next phase.

3.2 | Validation of the search filter

We identified 17 SRs eligible for the CSs. Within these SRs, the number of included relevant records (i.e. MEDLINE records related to studies about Germany) ranged from 6 to 35, resulting in a total of 178 records, published between 1981 and 2022.

Table 3 presents results of the validation process of the two versions of the filter by combining them with the original search strategies of the SRs. Both versions demonstrated the median sensitivity of 100% (interquartile range [IQR]: 93.75%–100%), ranging between 80% (1 CS) and 100% (13 CSs). In two CSs, the versions demonstrated the sensitivity of 85.71%, while in the remaining one, a sensitivity of 87.5% was achieved. Both versions failed to identify five relevant records,[31–35] which lacked terms related to Germany in any search field.

Because both versions had equivalent levels of sensitivity, we could not determine the final filter in this phase. As mentioned earlier, we assumed that excluding the language field from the filter would decrease number of irrelevant (false-positive) records. Across the 17 SRs, the median precision of the original search strategies without any version of the filter applied was 0.16% (IQR: 0.10%–0.49%) with a median NNR of 625 (IQR: 211–1042). The full version of the filter achieved median precision of 2.55% (IQR: 1.22%–6.00%) and NNR of 39 (IQR: 17–86). The version without the language field demonstrated a median precision of 2.64% (IQR: 1.34%–6.88%) and an NNR of 38 (IQR: 15–76).

In the simulation study both versions of the filter demonstrated a sensitivity of 97.19% (see Table 4). The same five records as in the CSs were overlooked by both versions. The exhaustive search strategy alone demonstrated precision of 0.16% and NNR of 625. The full version of the filter demonstrated precision of 2.11% and NNR of 47, while the version without the language field achieved precision of 2.36% and NNR of 42. Furthermore, the full version retrieved 850 more records than the version without the language field.

Considering results regarding precision and NNR in the CSs and in the simulation study, the version without the language field was determined as the final version of the search filter (further referred to as the search filter, see Appendix S5).

TABLE 1 Description of the search filter versions.

Search fields	Full version of the filter Search terms	Version of the filter without the language field Search terms
[Title/Abstract]	German* FRG GDR Terms related to the names of: • Geographical locations • Research institutions • Health insurance funds • Professional organizations • Hospitals	German* FRG GDR Terms related to the names of: • Geographical locations • Research institutions • Health insurance funds • Professional organizations • Hospitals
[Affiliation]	German* FRG GDR Terms related to the names of: • Geographical locations • Research institutions • Health insurance funds • Professional organizations • Hospitals	German* FRG GDR Terms related to the names of: • Geographical locations • Research institutions • Health insurance funds • Professional organizations • Hospitals
[MeSH terms]	Germany Germany, West Germany, East	Germany Germany, West Germany, East
[Language]	German	None
[Filter]	MEDLINE	MEDLINE

Abbreviations: FRG, Federal Republic of Germany; GDR, German Democratic Republic.

4 | DISCUSSION

Using objective and subjective approaches, we developed the first validated search filter designed to retrieve primary studies about Germany from MEDLINE (PubMed). The filter reliably identifies studies about Germany. Overall, the filter demonstrated a median sensitivity of 100% within the 17 CSs and a sensitivity of 97.19% in the simulation study.

The search filter failed to identify six of the total 1160 records from the samples that were used for evaluation and validation. These six records lacked adequate geographical details within the search fields incorporated into the filter. For instance, during the evaluation, the filter did not retrieve a record from the 'development set' that had Germany-related term in the corporate author field only.[30] Despite this finding, we decided not to incorporate this field into the filter, as we assumed that it is

TABLE 2 Results of the search filter evaluation.

	Full version of the filter			Version of the filter without the language field		
Set of records	No. of records	No. of records identified	Sensitivity (%)	No. of records	No. of records identified	Sensitivity (%)
'Development set'	654	653	99.85	654	653	99.85
'Testing set'	328	328	100	328	328	100

TABLE 3 Results of the case studies.

First author (year)	No. of relevant records	Original search string			Full version of the filter				Version of the filter without the language field			
		Hits	Precision (%)	NNR	Hits	Sensitivity (%)	Precision (%)	NNR	Hits	Sensitivity (%)	Precision (%)	NNR
Andrejko (2021)[40]	12	4574	0.26	384.62	216	100	5.56	17.99	210	100	5.71	17.51
Lopes (2021)[41]	8	1991	0.40	250.00	103	100	7.77	12.87	85	100	9.41	10.63
Cereda (2016)[42]	9	504	1.79	55.87	29	100	31.05	3.22	29	100	31.03	3.22
Christensen (2023)[43]	13	2166	0.60	166.67	202	100	6.44	15.53	185	100	7.0	14.22
Foley (2022)[44]	7	5623	0.12	833.33	431	85.71	1.39	71.94	405	85.71	1.48	67.57
Gallego (2012)[45]	8	12,179	0.07	1428.57	992	100	0.81	123.46	670	100	1.19	84.03
Gong (2022)[46]	8	1112	0.72	138.89	99	100	8.08	12.38	90	100	8.89	11.25
Hu (2017)[47]	7	4413	0.16	625	328	100	2.13	46.95	291	100	2.41	41.49
Ibrahim (2017)[48]	9	24,601	0.04	2500	1531	100	0.59	169.49	1392	100	0.65	153.85
Kosmopoulos (2023)[49]	8	4105	0.19	526.32	234	87.5	2.99	33.44	104	87.50	6.73	14.86
Vitturi (2022)[50]	5	6726	0.07	1428.57	864	80.00	0.46	217.39	822	80.00	0.49	204.08
Lu (2023)[51]	14	17,372	0.08	1250.00	1138	85.71	1.05	95.24	1097	85.71	1.09	91.74
Meng (2022)[52]	15	7804	0.19	526.32	560	100	2.68	37.31	539	100	2.78	35.97
Raoofi (2023)[53]	6	3464	0.16	625	235	100	2.55	39.22	227	100	2.64	37.88
Singh (2022)[54]	7	5842	0.12	833.33	311	100	2.25	44.44	299	100	2.34	42.74
Ten Cate (2023)[55]	7	4713	0.15	666.67	344	100	2.03	49.26	323	100	2.17	46.08
Ting (2023)[56]	35	6086	0.58	172.41	698	100	5.01	19.96	671	100	5.22	19.16
Median (IQR)		4713 (2815–7265)	0.16 (0.10–0.49)	625 (211–1042)	328 (209–781)	100 (93.75–100)	2.55 (1.22–6.00)	39 (17–86)	299 (145–671)	100 (93.75–100)	2.64 (1.34–6.88)	38 (15–76)

Abbreviations: IQR, interquartile range; NNR, number-needed to read.

TABLE 4 Results of the simulation study.

	Number of hits	Sensitivity (%)	Precision (%)	NNR
Exhaustive search strategy	111,463	100	0.16	625
Full version of the filter	8189	97.19	2.11	47
Version of the filter without the language field	7339	97.19	2.36	42

Abbreviation: NNR, number-needed to read.

uncommon for geographical details to be present there. The other five records[31–35] lacked terms relevant to Germany within any bibliographical fields, making their identification possible only by examining the full texts. An alternative strategy for identifying these records would involve searching another database. For instance, when searching for these six records in Embase, four of them have Germany-related terms within the author address field.[30,31,33,35] This observation, along with existing research on geographic search filters, suggests that records in Embase may be more comprehensively indexed with regard to geographical details.[36] Therefore, we plan to adapt the MEDLINE (PubMed) filter for use in the Embase database. Using both filters together could potentially enhance the retrieval of records related to studies about Germany.

The results of the validation process suggest that application of the filter substantially reduces the workload associated with the screening process. For instance, using the filter in the CSs increased the median precision by a factor of 16.50, reducing the median NNR from 625 to 38. The median number of search results was reduced by 4414 records, corresponding to a 93.66% decrease. Given that 500–1000 abstracts can typically be screened over an 8-hour period,[37] the reduction in search results demonstrated by our filter could potentially save about 35–70 working hours per review member.

It is important to note that the filter application results in an error message because certain search terms, such as "Trossingen", a town in Baden-Württemberg, are not currently indexed in PubMed. However, this does not affect the filter's functionality. We have chosen to keep these search terms, anticipating their potential indexing in PubMed in the future.

4.1 | Strengths and limitations

This study possesses both limitations and strengths that should be taken into account alongside the presented results.

First, the relative recall approach employed to create the GSS may introduce a potential limitation, as it could bias the developed filter towards the search terms identified in the GSS.[38] However, in the case of our filter, this limitation is much reduced, given that our GSS included records related to primary studies covering a diverse range of topics and study types, spanning almost 30 years of publication. This was confirmed by validation of the filter conducted on a set of relevant studies that were published between 1981 and 2022. However, it is important to mention that the oldest publication date (May 1, 1992) in the GSS falls after German reunification. This means that the pre-reunification names of research institutions might not be captured by the GSS and, therefore, are not included in the filter. Another important aspect related to the German reunification is that we limited the dates of publication for FRG and GDR in the affiliation field to up to 1990. However, we acknowledge that some research institutions might have changed the country name in their affiliation (e.g. from GDR to Germany) after reunification. Furthermore, due to the lag between research completion and publication, articles published after 1990 can still have pre-reunification affiliations.

Second, we intended initially to adopt the objective approach alone for the filter development. However, the results of the terms frequency analyses revealed that the size of the 'development set', comprising 654 records, was not sufficient to identify all German geographical locations relevant for conducting and producing research. For instance, some big German cities, locations of higher education institutions and of university hospitals were not identified. Consequently, we decided to complement these terms by creating a subjectively developed list that, in our opinion, covers the vast majority of locations where German research can be carried out. While most terms identified through the terms frequency analyses were also included in the subjectively developed list, 38 terms were exclusively identified via the objective method. Moreover, the objective approach allowed us to identify German terms that are not geographical locations, such as names of professional organizations, forms of health insurances and relevant abbreviations. Therefore, it seems that applying both methods together may yield the best results in terms of the filter sensitivity. However, it is important to note that despite combining objective and subjective approaches, our filter might not contain all relevant search terms. For instance, our subjectively developed list did not contain names of the cities with fewer than 100,000 inhabitants. Consequently, the filter might miss records where the only identifiable terms are related to these locations or other terms not covered by the filter for reasons mentioned above (e.g. the oldest publication date in the GSS and the time limit for the FRG and GDR in the affiliation).

Third, to enhance the sensitivity of the filter, we considered terms resulted from the term frequency analyses with a prevalence ratio >1 as potentially relevant for the filter. This resulted in a large number of candidate search terms, which should be screened manually by the team to identify the truly relevant terms. Such an approach, however, can be seen as a limitation. Manual screening of a large number of candidate search terms could potentially lead to overlooking the truly relevant terms and is time-consuming. Further research on defining the threshold in terms of prevalence ratio or other measures for candidate search terms would be useful to reduce possibilities of human factor mistakes and increase efficiency of the process.

Finally, although our primary objective was to reach the highest degree of sensitivity possible, we employed methods that are likely to positively affect the precision of the filter. For instance, comparing terms in the 'development' and 'population' sets allowed us to select terms both frequently used in the 'development set' and specific to the focus of our filter.[39] Furthermore, we verified that all the records from the GSS are related to studies about Germany. However, we did not verify whether the records included as German studies in the SRs used for the filter validation were indeed about Germany. Despite this, given the methodology applied during the development phase, we believe that latter had no significant impact on the filter's performance.

5 | CONCLUSIONS

A validated, highly sensitive geographic, and easy-to-use search filter to be used in MEDLINE (PubMed) for retrieving primary studies about Germany is now available. The validation process demonstrated that the filter is able to identify nearly all relevant references and has the potential to enhance efficiency of the screening process, saving both time and resources. However, users who are not inclined to conduct thorough searches should be aware that despite the filter's efficiency the screening process may still be time-consuming due to the relatively high values of NNR. Therefore, our filter can be particularly useful for researchers aiming to comprehensively identify primary studies about Germany, facilitating the production of high-quality SRs crucial for evidence informed decision-making.

AUTHOR CONTRIBUTIONS

Alexander Pachanov: Conceptualization; methodology; software; writing – original draft; writing – review and editing; formal analysis; data curation. **Catharina Münte:** Methodology; writing – review and editing; formal analysis. **Julian Hirt:** Conceptualization; methodology; writing – review and editing. **Dawid Pieper:** Conceptualization; methodology; writing – review and editing.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

German Research Foundation (*Deutsche Forschungsgemeinschaft*; project number 471387773).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICAL APPROVAL

Not necessary for this study.

ORCID

Alexander Pachanov https://orcid.org/0000-0003-2579-3142
Catharina Münte https://orcid.org/0000-0002-3640-4010
Julian Hirt https://orcid.org/0000-0001-6589-3936
Dawid Pieper https://orcid.org/0000-0002-0715-5182

REFERENCES

1. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med*. 2016;13(5):e1002028.
2. Ayiku L, Levay P, Hudson T. The NICE OECD countries' geographic search filters: part 1-methodology for developing the draft MEDLINE and Embase (Ovid) filters. *J Med Libr Assoc*. 2021;109(2):258-266.
3. Ayiku L, Levay P, Hudson T, et al. The Embase UK filter: validation of a geographic search filter to retrieve research about the UK from OVID Embase. *Health Info Libr J*. 2019;36(2):121-133.
4. Lefebvre C, Glanville J, Beale S, et al. Assessing the performance of methodological search filters to improve the efficiency of evidence information retrieval: five literature reviews and a qualitative study. *Health Technol Assess*. 2017;21(69):1-148.
5. Jenkins M. Evaluation of methodological search filters – a review. *Health Info Libr J*. 2004;21(3):148-163.

6. Damarell RA, May N, Hammond S, Sladek RM, Tieman JJ. Topic search filters: a systematic scoping review. *Health Info Libr J*. 2019;36(1):4-40.
7. Ho KL, Shiels MS, Ramin C, et al. County-level geographic disparities in cardiovascular disease mortality among US breast cancer survivors, 2000–2018. *JNCI Cancer Spectr*. 2023;7(1):pkac083.
8. Parcha V, Kalra R, Best AF, et al. Geographic inequalities in cardiovascular mortality in the United States: 1999 to 2018. *Mayo Clin Proc*. 2021;96(5):1218-1228.
9. Singh D, Vaccarella S, Gini A, De Paula SN, Steliarova-Foucher E, Bray F. Global patterns of Hodgkin lymphoma incidence and mortality in 2020 and a prediction of the future burden in 2040. *Int J Cancer*. 2022;150(12):1941-1947.
10. Lu VM, Shah AH, Eichberg DG, et al. Geographic disparities in access to glioblastoma treatment based on Hispanic ethnicity in the United States: insights from a national database. *J Neurooncol*. 2020;147(3):711-720.
11. Ayiku L, Levay P, Hudson T, et al. The medline UK filter: development and validation of a geographic search filter to retrieve research about the UK from OVID medline. *Health Info Libr J*. 2017;34(3):200-216.
12. Valderas JM, Mendivil J, Parada A, Losada-Yáñez M, Alonso J. Development of a geographic filter for PubMed to identify studies performed in Spain. *Rev Esp Cardiol*. 2006;59(12):1244-1251.
13. Pienaar E, Grobler L, Busgeeth K, Eisinga A, Siegfried N. Developing a geographic search filter to identify randomised controlled trials in Africa: finding the optimal balance between sensitivity and precision. *Health Info Libr J*. 2011;28(3):210-215.
14. Hirt J, Buhtz C, Meyer G, Balzer K. Publications German-speaking countries in high impact journals: development and validation of a search filter. *Pflege*. 2019;32(2):97-106.
15. Ayiku L, Hudson T, Williams C, Levay P, Jacob C. The NICE OECD countries' geographic search filters: part 2-validation of the MEDLINE and Embase (Ovid) filters. *J Med Libr Assoc*. 2021;109(4):583-589.
16. Cheung A, Popoff E, Szabo SM. Application of text mining to the development and validation of a geographic search filter to facilitate evidence retrieval in Ovid MEDLINE: an example from the United States. *Health Info Libr J*. 2022;40:169-180.
17. Pieper D, Mathes T, Palm R, Hoffmann F. How do authors of systematic reviews restrict their literature searches when only studies from Germany should be included? *Z Evid Fortbild Qual Gesundhwes*. 2016;117:1-8.
18. Sampson M, Zhang L, Morrison A, et al. An alternative to the hand searching gold standard: validating methodological search filters using relative recall. *BMC Med Res Methodol*. 2006;6:33.
19. Pachanov A, Muente C, Hirt J, Hoffmann F, Palm R, Pieper D. How do authors of systematic reviews restrict their literature searches when only studies from Germany should be included? A methodological review update OSF Registries. 2022.
20. Ayiku L, Craven J, Hudson T, Levay P. How to develop a validated geographic search filter: five key steps. *Evid Based Libr Inf Pract*. 2020;15(1):170-178.
21. Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. *Syst Rev*. 2012;1:19.
22. Team RC. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2022.
23. Lunny C, McKenzie JE, McDonald S. Retrieval of overviews of systematic reviews in MEDLINE was improved by the development of an objectively derived and validated search strategy. *J Clin Epidemiol*. 2016;74:107-118.
24. Isaak D. PubMed2XL (version 2.01). *J Med Libr Assoc*. 2016;104(1):92-94.
25. Benoit K, Watanabe K, Wang H, et al. Quanteda: an R package for the quantitative analysis of textual data. *J Open Source Softw*. 2018;3:774.
26. Bundesamt S. *Städte (Alle Gemeinden mit Stadtrecht) nach Fläche, Bevölkerung und Bevölkerungsdichte am 31.12.2022.*2023. Available from: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/05-staedte.html.
27. Universitätsklinika Dd. *Übersicht der Universitätsklinika in Deutschland*. Verband der Universitätsklinika Deutschlands; 2023. Available from: https://www.uniklinika.de/die-deutscheneuniversitaetsklinika/uebersicht-der-universitaetsklinika/.
28. Hochschulkompass. *Download Hochschullisten*. Hochschulrektorenkoferenz; 2023. Available from: https://www.hochschulkompass.de/hochschulen/downloads.html.
29. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc*. 2002;9(6):653-658.
30. Biermann J, Neumann T, Angermann CE, et al. Economic burden of patients with various etiologies of chronic systolic heart failure analyzed by resource use and costs. *Int J Cardiol*. 2012;156(3):323-325.
31. Gianni L, Baselga J, Eiermann W, et al. Feasibility and tolerability of sequential doxorubicin/paclitaxel followed by cyclophosphamide, methotrexate, and fluorouracil and its effects on tumor response as preoperative therapy. *Clin Cancer Res*. 2005;11(24 Pt 1):8715-8721.
32. Ackland SP, Anton A, Breitbach GP, et al. Dose-intensive epirubicin-based chemotherapy is superior to an intensive intravenous cyclophosphamide, methotrexate, and fluorouracil regimen in metastatic breast cancer: a randomized multinational study. *J Clin Oncol*. 2001;19(4):943-953.
33. Schüz B, Wolff JK, Warner LM, Ziegelmann JP, Wurm S. Multiple illness perceptions in older adults: effects on physical functioning and medication adherence. *Psychol Health*. 2014;29(4):442-457.
34. Scheinberg L, Holland N, Larocca N, Laitin P, Bennett A, Hall H. Vocational disability and rehabilitation in multiple sclerosis. *Int J Rehabil Res*. 1981;4(1):61-64.
35. Rienmüller R, Doppman JL, Lissner J, Kemkes BM, Strauer BE. Constrictive pericardial disease: prognostic significance of a nonvisualized left ventricular wall. *Radiology*. 1985;156(3):753-755.
36. Ayiku L, Levay P, Hudson T, Finnegan A. The NICE UK geographic search filters for MEDLINE and Embase (Ovid): post-development study to further evaluate precision and number-needed-to-read when retrieving UK evidence. *Res Synth Methods*. 2020;11(5):669-677.
37. Lefebvre C, Glanville J, Briscoe S, et al. Chapter 4: searching for and selecting studies. *Cochrane Handbook for Systematic Reviews of Interventions [Internet]*. Cochrane; 2023 Available from: https://training.cochrane.org/handbook
38. Hubbard W, Walsh N, Hudson T, Heath A, Dietz J, Rogers G. Development and validation of paired MEDLINE and Embase

search filters for cost-utility studies. *BMC Med Res Methodol*. 2022;22(1):310.
39. Simon M, Hausner E, Klaus SF, Dunton NE. Identifying nurse staffing research in Medline: development and testing of empirically derived search strategies with the PubMed interface. *BMC Med Res Methodol*. 2010;10:76.
40. Andrejko K, Ratnasiri B, Hausdorff WP, Laxminarayan R, Lewnard JA. Antimicrobial resistance in paediatric *Streptococcus pneumoniae* isolates amid global implementation of pneumococcal conjugate vaccines: a systematic review and meta-regression analysis. *Lancet Microbe*. 2021;2(9):e450-e460.
41. Lopes LB, Machado V, Mascarenhas P, Mendes JJ, Botelho J. The prevalence of molar-incisor hypomineralization: a systematic review and meta-analysis. *Sci Rep*. 2021;11(1):22405.
42. Cereda E, Pedrolli C, Klersy C, et al. Nutritional status in older persons according to healthcare setting: a systematic review and meta-analysis of prevalence data using MNA(®). *Clin Nutr*. 2016;35(6):1282-1290.
43. Christensen MO, Barakji YA, Loft N, et al. Prevalence of and association between atopic dermatitis and food sensitivity, food allergy and challenge-proven food allergy: a systematic review and meta-analysis. *J Eur Acad Dermatol Venereol*. 2023;37(5):984-1003.
44. Foley L, Larkin J, Lombard-Vance R, et al. Prevalence and predictors of medication non-adherence among people living with multimorbidity: a systematic review and meta-analysis. *BMJ Open*. 2021;11(9):e044987.
45. Gallego JA, Bonetti J, Zhang J, Kane JM, Correll CU. Prevalence and correlates of antipsychotic polypharmacy: a systematic review and meta-regression of global and regional trends from the 1970s to 2009. *Schizophr Res*. 2012;138(1):18-28.
46. Gong S, Gao Y, Liu J, et al. The prevalence and associated factors of dysphagia in Parkinson's disease: a systematic review and meta-analysis. *Front Neurol*. 2022;13:1000527.
47. Hu Y, Wan JH, Li XY, Zhu Y, Graham DY, Lu NH. Systematic review with meta-analysis: the global recurrence rate of *Helicobacter pylori*. *Aliment Pharmacol Ther*. 2017;46(9):773-779.
48. Ibrahim A, Morais S, Ferro A, Lunet N, Peleteiro B. Sex-differences in the prevalence of *Helicobacter pylori* infection in pediatric and adult populations: systematic review and meta-analysis of 244 studies. *Dig Liver Dis*. 2017;49(7):742-749.
49. Kosmopoulos M, Liatsou E, Theochari C, et al. Updates on the global prevalence and etiology of constrictive pericarditis: a systematic review. *Cardiol Rev*. 2023;32:417-422.
50. Vitturi BK, Rahmani A, Dini G, et al. Spatial and temporal distribution of the prevalence of unemployment and early retirement in people with multiple sclerosis: a systematic review with meta-analysis. *PLoS One*. 2022;17(7):e0272156.
51. Lu Y, Pan W, Deng S, et al. Redefining the incidence and profile of fluoropyrimidine-associated cardiotoxicity in cancer patients: a systematic review and meta-analysis. *Pharmaceuticals (Basel)*. 2023;16(4):510.
52. Meng SQ, Cheng JL, Li YY, et al. Global prevalence of digital addiction in general population: a systematic review and meta-analysis. *Clin Psychol Rev*. 2022;92:102128.
53. Raoofi S, Pashazadeh Kan F, Rafiei S, et al. Global prevalence of nosocomial infection: a systematic review and meta-analysis. *PLoS One*. 2023;18(1):e0274248.
54. Singh J, Stoitsova S, Zakrzewska K, Henszel L, Rosińska M, Duffell E. Healthcare-associated hepatitis B and C transmission to patients in the EU/EEA and UK: a systematic review of reported outbreaks between 2006 and 2021. *BMC Public Health*. 2022;22(1):2260.
55. Ten Cate V, Prochaska JH, Schulz A, et al. Clinical profile and outcome of isolated pulmonary embolism: a systematic review and meta-analysis. *EClinicalMedicine*. 2023;59:101973.
56. Ting RS, Lewis DP, Yang KX, et al. Incidence of multiple organ failure in adult polytrauma patients: a systematic review and meta-analysis. *J Trauma Acute Care Surg*. 2023;94(5):725-734.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pachanov A, Münte C, Hirt J, Pieper D. Development and validation of a geographic search filter for MEDLINE (PubMed) to identify studies about Germany. *Res Syn Meth*. 2024;15(6):1147-1160. doi:10.1002/jrsm.1763