No suitable precise or optimized epidemiologic search filters were available for bibliographic databases

Siw Waffenschmidt[a,*], Tatjana Hermanns[a], Andreas Gerber-Grote[b], Sarah Mostardt[c]

[a]*Information Management Unit, Institute for Quality and Efficiency in Health Care (IQWiG), Im Mediapark 8, 50670 Cologne, Germany*
[b]*School of Health Professions, Zurich University of Applied Sciences, Winterthur, Switzerland*
[c]*Division Health Economics, IQWiG, Im Mediapark 8, 50670 Cologne, Germany*

Accepted 18 August 2016; Published online 26 August 2016

Abstract

Objectives: To determine a suitable approach to a systematic search for epidemiologic publications in bibliographic databases. For this purpose, suitable sensitive, precise, and optimized filters were to be selected for MEDLINE searches. In addition, the relevance of bibliographic databases was determined.

Study Design and Setting: Epidemiologic systematic reviews (SRs) retrieved in a systematic search and company dossiers were screened to identify epidemiologic publications (primary studies and SRs) published since 2007. These publications were used to generate a test and validation set. Furthermore, each SR's search strategy was reviewed, and epidemiologic filters were extracted. The search syntaxes were validated using the relative recall method.

Results: The test set comprises 729 relevant epidemiologic publications, of which 566 were MEDLINE-indexed. About 27 epidemiologic filters were extracted. One suitable sensitive filter was identified (Larney et al. 2013: 95.94% sensitivity). Precision was presumably underestimated so that no precise or optimized filters can be recommended. About 77.64% of the publications were found in MEDLINE.

Conclusion: There is currently no suitable approach to conducting efficient systematic searches for epidemiologic publications in bibliographic databases. The filter by Larney et al. (2013) can be used for sensitive MEDLINE searches. No robust conclusions can be drawn on precise or optimized filters. Additional search approaches should be considered. © 2016 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Information storage and retrieval; Evidence-based medicine; Databases, bibliographic; MEDLINE; Review literature as topic; Epidemiology

1. Introduction

Epidemiologic data on diseases (prevalence, incidence, and mortality) are indispensable in health economic evaluations, for example, when determining background mortality or performing budget impact analyses. These data are also indispensable for planning health care interventions or determining the burden of disease in a population and thus for health policy decision making.

In 2004, the German Institute for Quality and Efficiency in Health Care (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen [IQWiG]) was established to inform decision making in the German statutory health insurance system. Initially, the main responsibility of IQWiG was to produce evidence-based benefit assessments (i.e., systematic reviews [SRs]) of drug and nondrug interventions that had largely been available in the market for several years. Since 2011, IQWiG has also been conducting early benefit assessments of newly approved drugs on the basis of dossiers provided by pharmaceutical companies [1]. These so-called dossier assessments have to be completed by IQWiG within 3 months. In such an assessment, IQWiG must also estimate the size of the target population likely to benefit from the new drug. For this purpose, IQWiG uses prevalence and incidence data provided in the dossiers. These epidemiologic data represent published information from various sources such as public registries and medical journals and are relevant as a basis for price negotiations between the statutory health insurance umbrella organization and the pharmaceutical companies.

Different sources can be used to identify descriptive epidemiologic data. Public sources have a special status and should be primarily searched as they are most likely to fulfill the quality requirements for epidemiologic data (representativeness, completeness, etc.) [2]. The sources for German data

What is new?

Key findings

- No suitable approach to conducting efficient systematic searches for epidemiologic publications in bibliographic databases is currently available.

- The filter by Larney et al. (2013) can be used for sensitive searches in MEDLINE.

What this adds to what was known?

- Handsearching additional sources to increase the number of relevant publications should be considered.

What is the implication and what should change now?

- Additional search approaches, such as combining epidemiologic filters with country- or population-specific search terms, should also be considered.

- It is still necessary to further develop search techniques to make systematic searches more efficient.

include the Federal Statistical Office [3], Robert Koch Institute [4], and European Center for Disease Control [5]. If data from these sources cannot be used for the assessment, epidemiologic publications need to be searched for in bibliographic databases [2]. However, to the best of our knowledge, no standard for this type of search has yet been described in the literature. For instance, according to an analysis from 2005 by Royle et al. [6] on the efficiency of searching in support of SRs of the epidemiology of diabetes, no standardized approach exists for identifying epidemiologic publications on diabetes. Furthermore, no validated search filters for identifying epidemiologic publications (epidemiologic search filters) are available on well-known Web sites for filters [7].

The desired performance of a filter varies, depending on the objective of the search (sensitivity-maximizing approach, precision-maximizing approach, or a balance between sensitivity and precision [8]). In the latter text, the respective filters are referred to as sensitive, precise, and optimized.

2. Aim

The aim of the present study was to determine a suitable approach to a systematic search for epidemiologic publications (primary studies and SRs) in bibliographic databases. For this purpose, three different types of suitable filters (sensitive, precise, and optimized) were to be selected for searches in MEDLINE. In addition, the relevance of MEDLINE and further bibliographic databases as a source for epidemiologic publications was to be determined so that

conclusions could be drawn on which databases should routinely be used.

3. Methods

The approach was divided into five steps (Fig. 1) following the standards currently used for filter evaluation (e.g., suitable evaluation methods and performance measures) [9].

3.1. Identification of relevant SRs

In November 2013, a systematic search for epidemiologic SRs was conducted in MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects, the Cochrane Methodology Register, and the Health Technology Assessment Database. The corresponding search strategies are presented in Appendix A (see on the journal's Web site at www.elsevier.com). Because of the large number of hits, the search was restricted to three therapeutic indications. For this purpose, an infectious disease, a metabolic disease, and a neurologic disease (hepatitis C, diabetes, and multiple sclerosis, respectively) were chosen for which new drugs had recently been approved in Germany; these drugs are thus highly relevant for IQWiG.

Eligible SRs were English- or German-language publications investigating an epidemiologic question and containing data on the prevalence, incidence, or mortality of hepatitis C, diabetes, or multiple sclerosis. As epidemiologic data change over time, it is essential that they are up to date; only SRs published since January 1, 2007 were therefore considered. The screening of titles and abstracts as well as full texts subsequently selected was conducted by two reviewers independently of one other. Disagreements were resolved by consensus.

3.2. Documentation of the search syntaxes used

Each SR's search strategy was reviewed, and those filters were selected that contained epidemiologic search terms and were fully documented. In addition, a search for epidemiologic search filters was conducted on the Web sites of the Scottish Intercollegiate Guidelines Network [10], the InterTASC Information Specialists' Sub-Group [7], and of Orphanet, the portal for rare diseases and orphan drugs [11]. Those filters that were fully reproducible were extracted and, if necessary, adapted to the database-specific syntax of MEDLINE (Ovid).

3.3. Generation of the test and validation sets

A test and a validation set were generated as the basis for filter validation and for evaluation of the relevance of selected bibliographic databases. Epidemiologic publications (primary studies and SRs) published since January 1, 2007 were considered. The type of publication was not