Searching bibliographic databases for literature on chronic disease and work participation

Joke Haafkens¹, Clara Moerman², Merel Schuring¹ and Frank van Dijk¹

Background The work participation of people with chronic diseases is a growing concern within the field of occupational medicine. Information on this topic is dispersed across a variety of data sources, making it difficult for health professionals to find relevant studies for literature reviews and guidelines.

Aim The goal of this project was to identify bibliographic databases and search terms that could be most useful for retrieving relevant studies on this topic.

Methods Five broad questions regarding work participation and chronic disease were formulated, focusing on angina pectoris, depression, diabetes mellitus, hearing impairment and rheumatoid arthritis. A search strategy for retrieving information on these questions was developed and run in five bibliographic databases: Medline, EMBASE, PsycINFO, Cinahl and OSHROM. Relevant publications were selected from the search results. The utility of the selected databases and search terms was evaluated by analysing the number of relevant publications that were retrieved.

Results The number of relevant publications retrieved from each database varied. Most (84%) of the relevant publications that were retrieved from each database were unique to that source. For each database, specific search terms for the concept of ‘work’ were useful for retrieving relevant publications.

Conclusion Medline, EMBASE and PsycINFO are useful databases for quick searches. Useful search terms for the concept of ‘work’ are work capacity, work disability, vocational rehabilitation, occupational health, sick leave, absenteeism, return to work, retirement, employment status and work status. For comprehensive searches, we recommend additional searches in Cinahl and OSHROM, adapting the search terms to specific databases.

Key words Bibliographic databases; chronic disease; occupational health; search strategies; search terms; work disability.

Introduction Advances in medical technology have decreased the rates of mortality and incapacitating morbidity from chronic diseases. Although one consequence of this development is that more individuals with chronic diseases are able to work, the labour participation of this group still lags behind that of the general population [1,2]. In recent years, a number of Western countries have adopted policies for improving employment opportunities for people with chronic diseases [3–5]. Such policies require the active support of both employers and health care providers [6,7]. Occupational health care for patients with chronic diseases is therefore a growing concern in medicine.

In the past, occupational medicine focused primarily on occupational diseases, clinical medicine focused on the clinical aspects of chronic diseases and rehabilitation medicine concentrated on enabling patients to be independent at home rather than at work [8]. Occupational physicians and other health care professionals have only recently begun to develop specific programmes and practice guidelines for managing the work-related problems of patients with chronic diseases [9,10]. Systematic reviews of the professional literature are necessary to support their efforts.

Although this literature can now be retrieved electronically from bibliographic databases, searching these databases is not always easy. Search strategies for retrieving information on ‘clinical’ questions in biomedical databases are already widely available [11–13]. Questions on ‘occupational health’ issues, however, often have a wider
scope and are of an interdisciplinary nature. Retrieving information on such questions from on line bibliographic databases has proved particularly complex and time consuming, even for experienced researchers [14,15].

The aim of this study was to identify bibliographic databases and search terms that could be particularly useful for locating literature on questions that are related to chronic disease and work participation. We hope that the results will facilitate the work of professionals who seek information on occupational health care for people with chronic diseases, including occupational health professionals who are involved in the development of programmes and practice guidelines, occupational health practitioners and the librarians or information specialists who support them.

Materials and methods

The study was conducted between June 2001 and August 2002 by M.S., C.M., J.H. and F.vD., all of whom are experienced IT users. Our goal was to develop a search strategy that would facilitate the retrieval of good-quality general background information on chronic disease and work participation. To this end, we formulated five broad questions regarding the relationship between these two concepts that might be of interest to health care providers:

- How does chronic disease affect work ability?
- How does working affect chronic disease?
- Which diagnostic and prognostic instruments have been developed and evaluated to assess the work ability of individuals with chronic diseases?
- How do clinical interventions affect the work ability of individuals with chronic diseases?
- How do work-related interventions affect the work ability of individuals with chronic diseases?

We limited the study to five chronic conditions: angina pectoris, depression, diabetes mellitus, hearing impairment and rheumatoid arthritis. These conditions have relatively high prevalence and are likely to affect the work ability of patients [16]. To select bibliographic databases for our search, we scrutinized the contents of a list of 15 potentially relevant databases on occupational health and occupational injuries, as recommended by Beahler et al. [15]. From this list, we selected the biomedical databases Medline, EMBASE and Cinahl, the occupational health database OSHROM and the social science database PsycINFO. These databases offer more general information on chronic disease and work participation than do other databases that are more specialized, and they are more easily accessible to the general user (Table 1).

The selected databases can be accessed through various search engines. Our department uses Ovid as a search engine for accessing Medline, EMBASE and Cinahl, and it uses WebSPIRS (Silverplatter) for accessing PsycINFO and OSHROM. We therefore used Ovid and WebSPIRS to develop our search strategies.

All five databases allow the use of ‘free-text words’ to trace articles. Combined with the ability to search in ‘all fields’, free-text words can be used to locate any word or combination of words used in the title, abstract, author’s name, research institution or keyword lists. To facilitate searching, articles in Medline, EMBASE, Cinahl and PsycINFO are indexed in a thesaurus of terms or subject headings (in Medline, they are known as ‘Medical Subject Headings’ or MeSH terms). With the exception of PsycINFO, the thesauri of these databases have a hierarchical tree structure, ordered from general to specific. All articles are indexed using the most specific subject heading available. In addition, the thesauri of Medline, EMBASE and Cinahl contain subheadings that classify articles under particular subject headings (e.g. diabetes mellitus) into more specific medical categories (e.g. aetiology, diagnosis).

Search terms for the ‘five chronic diseases’, were derived from the subject headings in the thesauri of Medline, EMBASE and Cinahl. Because these databases are of biomedical origin, we assumed that chronic diseases would be appropriately defined by those subject headings.

Search terms related to the concept of work participation were selected by exploring subject headings from thesauri and keywords found in the abstracts of relevant publications, as well as through consultation with experts in the field. For a number of reasons, we did not limit the selection to subject headings that are indexed in the thesauri of the databases. First, the only specialized database on occupational health issues, OSHROM, does not select: ‘Occupational health database’: Transportation Research Information Services (TRIS). ‘Business database’: ABI Inform. ‘Criminal justice and social science databases’: Criminal Justice Periodicals Index (CJPI), Sociofile, Educational Resources Information Centre (ERIC). ‘Agricultural database’: Agris International, Agrolca. ‘US Government databases’: National Technical Information Services (NTIS), Public Affairs Information Service (PAIS). ‘General databases’: Books in Print, Dissertation Abstracts, Expanded Academic Index.

Table 1. Bibliographic databases selected for literature searches on chronic disease illness and work participation*

<table>
<thead>
<tr>
<th>Number</th>
<th>Database</th>
<th>Description</th>
</tr>
</thead>
</table>

not provide a thesaurus. Second, we discovered that subject headings related to the concept of ‘work’ are not always consistently defined and classified in the thesauri of the other databases. For example, the Medline thesaurus defines the subject heading ‘work’ as ‘purposeful activity (differentiate from employment for pay)’ and it has one lower order subject heading: ‘work schedule tolerance’. In contrast, the EMBASE thesaurus defines the subject heading ‘work’ as ‘job; job description; protected work; work physiology’, and it has nineteen lower order subject headings, including ‘absenteeism’, ‘job performance’, ‘job satisfaction’ or ‘work schedule’.

Our initial selection of search terms for work participation consisted of 49 items (on request available from the authors). To reduce this list, M.S. conducted test searches in Medline (1966 to May 2001), EMBASE (1988 to April 2001), Cinahl (1982 to April 2001), PsycINFO (1966 to July 2001) and OSHROM (Rilosh, Hseline and Cisdoc:1985–2001, NIOSHTIC:1985–1998), using a combination of the search terms for ‘work’ and the search terms for the five chronic conditions. In these searches, the search terms were used as both free-text words (using all fields) and subject headings. To increase the sensitivity of the searches (the probability of retrieving all relevant publications), the facility to ‘explode’ subject headings was used for a number of search terms related to work participation. To increase the specificity of the searches (the probability of retrieving only information on specific topics), search terms for chronic diseases were restricted to a selection of subheadings (e.g. diagnosis or therapy) [17].

Based on the number of publications (hits) found by the searches, as well as on discussions among the members of the project team, we selected 22 search terms related to work participation that were likely to yield good results. Terms that yielded more than 650 hits in more than one database were excluded, as they were too broad and insufficiently precise. The term ‘work’ is one example. Used as both a subject heading and a free-text word, it yielded 9373 hits in Medline, 5896 in EMBASE, 637 in Cinahl, 4405 in PsycINFO and 638 in OSHROM. ‘Employment’, ‘occupation’ and similar terms were excluded for the same reason. We also excluded terms that yielded fewer than 10 hits in four databases (e.g. occupational nurse, occupational questionnaire, safety hazard). Upon closer examination, these terms appeared to be either insufficiently formulated or related to other areas of interest in occupational medicine. Finally, specific terms (e.g. work capacity evaluation) were excluded if more general terms (e.g. work capacity) were included that were likely to yield the same information. The usefulness of the five selected databases and the 22 search terms was evaluated as follows.

First, M.S. searched all five databases using the selected search terms for the five chronic conditions and the 22 selected search terms for work. Searches were conducted in Medline (1985 to May 2001) EMBASE (1988 to May 2001), Cinahl (1985 to May 2001), PsycINFO (1988 to May 2001), OSHROM (1985 to April 2001) and NIOSHTIC (1985 to September 1998). The searches were restricted to human studies. Second, we used the Reference Manager software to create a separate file for each publication that was retrieved. Each file contains the title and abstract of the publication, its keyword lists and the database(s) in which it was found.

The third step in the evaluation involved the selection of relevant publications. Two researchers (M.S. and C.M.) screened the abstracts and titles of the publications that were retrieved and determined independently whether they provided relevant information on any of the five initial research questions. To be considered relevant, a publication had to contain data from one or more original studies and meet the following inclusion criteria:

- Study participants: people between the ages of 18 and 64 who had been diagnosed with any of the selected chronic diseases and who had no other important comorbidity.
- Interventions (if applicable): any intervention that was targeted at the health situations of working people or the employment situations of people with the selected chronic diseases.
- Outcome measures: outcome measures had to include either work-related or disease-related aspects (e.g. work performance, experienced quality of work, sickness absence or change in health status).

Publications were rejected if both reviewers considered that they did not meet the inclusion criteria. If the reviewers did not agree or if one of them was uncertain, a third assessor (J.H. or F.vD.) was consulted. Differences of opinion were settled by consensus.

Fourth, to identify the databases that were the most useful, we assessed the percentage of relevant publications that were retrieved from each database. Because different databases may refer to the same publication, we also assessed the proportion of relevant publications that were unique to individual databases. To this end, we analysed data from the Reference Manager files to determine the number of relevant publications that had been obtained from each database and the number of relevant publications that had been identified in one database, but not in the others.

Fifth, to identify the search terms for ‘work’ that had been most useful for the search, for each database, we computed the number of relevant publications containing one or more of the 22 selected search terms. Because these terms had been combined in our search strategy, we could not know which specific search terms had led to retrieval of a given publication. To solve this problem, C.M. and J.H. attempted to locate them by screening the titles, abstracts or keyword lists of the relevant publications, using information from the Reference Manager.
files. After discussion, the research team decided to regard a search term as particularly useful if it had occurred in at least 5% of the relevant publications that had been retrieved from a particular database.

To avoid bias, the evaluations of databases and search terms in the fourth and fifth steps were based on the relevant publications that had been published within the same time frame: between 1988 and the beginning of 2001.

Results

Searches conducted using the 22 selected search terms for work participation and those for the five chronic conditions yielded 3063 publications from the five databases (Table 2). The most publications for diabetes mellitus and angina pectoris were found in Medline; EMBASE contained the most publications for rheumatoid arthritis and hearing impairment and, not surprisingly, PsycINFO proved to be the best source for publications on depression. In many instances, more than one database referred to the same publication. After controlling for double entries, the searches yielded 2467 unique publications, varying from 1212 for depression to 142 for angina pectoris. After screening, only 436 publications appeared to contain information that was relevant to the research questions.

Table 3 provides information on the relevant publications that had been published between 1988 and 2001 (Table 3). Column 1 shows the percentage of relevant publications that would have been retrieved if we had searched 'only one' of the five databases. In this case, Medline, EMBASE or PsycINFO would have yielded 36, 40 or 23%, respectively, of the relevant publications. OSHROM and Cinahl would have provided lower percentages. Column 2 shows the proportion of the relevant publications that were found in ‘one’ particular database, and ‘not’ in others. Surprisingly, most of the relevant publications (84%) were found in ‘only one’ of the databases: 24% in Medline, 22% in EMBASE, 19% in PsycINFO, 13% in OSHROM and 6% in Cinahl. Each database thus offered specific information that was relevant to our research questions and that could not be found in the other databases.

Many of the search terms could be traced back in the titles, abstracts and keyword listings of the relevant publications of the five databases (Table 4). We considered a search term useful for a database if it occurred in at least 5% of the relevant articles that were found in that database. For each database, we found specific search terms that met this criterion: Medline (10 search terms), EMBASE (11), PsycINFO (9), Cinahl (10) and OSHROM (9). A different selection of search terms may therefore be appropriate for specific databases. With few exceptions, however, almost all the terms that yielded at least 5% of the relevant publications in Medline were also relevant for the other databases. We therefore recommend the first 10 terms from Table 4 for a quick search of literature on chronic illness and work participation.

Discussion

Searching literature for literature reviews in the field of chronic illness and work is not an easy task. Potentially relevant literature may be found in a large number of databases, and a great variety of search terms for work...
participation may be used to retrieve publications. In this study, we developed and evaluated a search strategy that could facilitate this task. We found that certain bibliographic databases and certain search terms were more useful than others. A number of caveats are nonetheless in order.

First, we found that a limited number of the 22 search terms for work participation (Table 4) yielded the best results. We excluded such obvious terms as ‘employment’, ‘occupation’ and ‘work’ from this selection, as they were defined differently in the thesauri of the various databases and because they provided too many hits in our test searches. Although their exclusion increased the efficiency of our search strategy, including them may have enabled us to find additional relevant studies. This issue requires further investigation.

Second, we chose to restrict our searches to 5 of the 15 potentially relevant databases for occupational health that were recommended by Beahler et al. [15]: Medline, EMBASE, Cinahl, PsycINFO and OSHROM. Surprisingly, we found little overlap among these databases regarding the retrieval of relevant publications. In each database, we found a considerable number of unique publications on chronic disease and work participation that could not be found in the other databases. Studies on the retrieval of information on other health-care-related topics (e.g. rehabilitation of people with severe mental

### Table 3. Percentage of relevant references, by database

<table>
<thead>
<tr>
<th>Database</th>
<th>% of all relevant references found in this database (N = 390)</th>
<th>% of all relevant references found 'only' in this database (N = 390)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>EMBASE</td>
<td>40</td>
<td>22</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Cinahl</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>OSHROM</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>123</td>
<td>84</td>
</tr>
</tbody>
</table>


*bFor four publications, accurate information about the databases from which they were retrieved was missing.

*cTotal >100%, as one reference may be present in more than one database.

### Table 4. Relevant publications per database, by search terms related to work (presented as percentages of all relevant publications found in that database)

<table>
<thead>
<tr>
<th>Relevant publications (N = 480)</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medline (n = 142)</td>
</tr>
<tr>
<td>Absenteeism</td>
<td>25</td>
</tr>
<tr>
<td>Work disability</td>
<td>22</td>
</tr>
<tr>
<td>Sick leave</td>
<td>17</td>
</tr>
<tr>
<td>Employment status</td>
<td>13</td>
</tr>
<tr>
<td>Work capacity</td>
<td>11</td>
</tr>
<tr>
<td>Rehabilitation, vocational</td>
<td>11</td>
</tr>
<tr>
<td>Occupational health</td>
<td>12</td>
</tr>
<tr>
<td>Return to work</td>
<td>6</td>
</tr>
<tr>
<td>Retirement</td>
<td>10</td>
</tr>
<tr>
<td>Work status</td>
<td>7</td>
</tr>
<tr>
<td>Occupational medicine</td>
<td>4</td>
</tr>
<tr>
<td>Job satisfaction</td>
<td>2</td>
</tr>
<tr>
<td>Work ability</td>
<td>2</td>
</tr>
<tr>
<td>Job performance</td>
<td>1</td>
</tr>
<tr>
<td>Occupational stress</td>
<td>0</td>
</tr>
<tr>
<td>Occupational health service</td>
<td>0</td>
</tr>
</tbody>
</table>


The table mentions only search terms that were present ≥5% of the relevant publications found in at least one database. Search terms that were present in <5% of the relevant publications in each database are as follows: sick absence, vocational guidance, disability pension, occupational rehabilitation, occupational physician and employment record.

Italic numbers refer to recommended search terms for each database.
illness and toxicology) have arrived at similar results [18,19]. It therefore seems constructive to examine whether the databases that were not tested in this study could also yield relevant publications that are unique to those sources. At the same time, it should be acknowledged that we may have found more overlap among the databases had we included such search terms as 'employment', 'occupation' and 'work', which yielded many hits in our exploratory search. This issue also calls for further investigation.

Third, the goal of this study was to identify databases and search terms that could be useful for retrieving a comprehensive set of studies on broad questions concerning chronic disease and work participation. Because the study was an initial attempt at retrieving such information, we feel that any attempt to define the relevant studies we found as the gold standard of evidence in the area to be premature. For that reason, we made no attempt to evaluate the usefulness of the search strategy by calculating the sensitivity and specificity of particular databases and search terms with reference to a gold-standard search. We intend to use this method in future research on optimal search strategies for retrieving information for systematic reviews in the field of occupational health.

Fourth, bibliographic databases are occasionally reconfigured, and subject headings may be added, deleted or defined differently. Conventions in the terminology used by researchers may also change over time. The validity of the relevant search terms we found in our study is therefore debatable. In our opinion, these search terms are quite robust, as many appeared to be relevant for all five databases. Moreover, these search terms could always be used as free-text words, even if the subject headings in the thesauri of the databases were to be reconfigured. Nevertheless, regular re-evaluation of the search terms is advisable.

This study was an initial attempt to facilitate the work of occupational health professionals and others who may seek information on work participation and chronic disease. Despite its limitations and the further work that is needed, the results suggest a number of recommendations:

- For limited searches, at least three databases should be searched: Medline, EMBASE and PsycINFO. Comprehensive searches should make use of OSHROM and Cinahl as well.
- For quick searches covering all databases, the following search terms related to work participation are most useful: work capacity, work disability, vocational rehabilitation, occupational health, sick leave, absenteeism, return to work, retirement, employment status and work status. They can be used as a string of terms, connected with the term OR.
- For searches in specific databases, specific combinations of search terms for work participation may be most useful (Table 4).
- When the results of a search are poor, more general terms (e.g. ‘employment’, ‘work’ or ‘occupation’) may be included.
- The search terms for work participation can be used as free-text words, as subject headings or (preferably) in both ways, according to the characteristics of the database searched.
- In general, subject headings (MeSH terms) that are available in the thesauri of the biomedical databases (Medline and EMBASE) can be recommended as search terms for chronic diseases. These terms may need to be adapted for searches in other databases.

Finally, it should be acknowledged that the use of the recommended databases and search terms does not automatically produce good search results. Each enquirer who carries out a search has specific questions and informational needs. The choice of search terms ultimately depends chiefly on those specific informational needs.

Acknowledgements

This study was part of a larger project on ‘chronic disease and work’, funded by the Dutch Board of Health Insurances. We would like to thank the Board for the financial support, as well as the programme ‘Fatigue and Work’ of the Dutch Organization of Scientific Research. We would also like to acknowledge the librarians of the Library of the Academic Medical Centre in Amsterdam for their support and Jos Verbeek for his useful comments on an earlier version of this paper.

Conflicts of interest

None declared.

References