A search strategy to identify studies on the prognosis of work disability: a diagnostic test framework

Rob Kok,1,2 Jos A H M Verbeek,3 Babs Faber,1,2 Frank J H van Dijk,1,2 Jan L Hoving1,2

ABSTRACT

Objective: Searching the medical literature for evidence on prognosis is an important aspect of evidence-based disability evaluation. To facilitate this, we aimed to develop and evaluate a comprehensive and efficient search strategy in PubMed, to be used by either researchers or practitioners and that will identify articles on the prognosis of work disability.

Methods: We used a diagnostic test analytic framework. First, we created a reference set of 225 articles on the prognosis of work disability by screening a total of 65,692 titles and abstracts from 10 journals in the period 2000–2009. Included studies had a minimum follow-up of 6 months, participants in the age of 18–64 with a minimum sick leave of 4 weeks or longer or having serious activity limitations in 50% of the cases and outcome measures that reflect impairments, activity limitations or participation restrictions. Using text mining methods, we extracted search terms from the reference set and, according to sensitivity and relative frequency, we combined these into search strings.

Results: Both the research and the practice search filter outperformed existing filters in occupational health, all combined with the Yale-prognostic filter. The Work Disability Prognosis filter for Research showed a comprehensiveness of 90% (95% CI 86 to 94) and efficiency expressed more user-friendly as Number Needed to Read=20 (95% CI 17 to 34).

Conclusions: The Work Disability Prognosis filter will help practitioners and researchers who want to find prognostic evidence in the area of work disability evaluation. However, further refining of this filter is possible and needed, especially for the practitioner for whom efficiency is especially important.

INTRODUCTION

In many countries in the world, disability evaluation, involves an examination by physicians to evaluate the nature and extent of the disability.1,2 An important objective of this examination is to evaluate the prognosis of the disease, the impairments and activity limitations and especially the participation restrictions of the patient.3–6 Consequently, a substantial number of health-related uncertainties and related search questions of insurance physicians are of a prognostic nature (39%).7

Making a prognosis of the chances of still being disabled after a period of time implies rather complex predictions that can generate complex search questions related to the natural course of a disease and to a multitude of factors that can have an impact on the future course of that disease. These factors include not only common determinants such as the severity of the disease, physical and mental condition, age, gender, level of education and having manual/non-manual work, but also the impact of a new therapy, rehabilitation efforts or other medical interventions, the willingness of an employer to support the impaired worker in return to work or social support from relatives.

Up till now, considerable effort has been put into developing search strategies, including the development of new filters for EBM questions in different medical domains such as diagnosis, prognosis and therapy. These methodological strategies are now integrated into PubMed to facilitate searching in MEDLINE. These search strategies include also the topic of prognosis or prediction.8
In their study on prognostic factors for work ability in sick-listed employees with chronic diseases, Slebus et al. used the Yale University’s methodological research filter for prognosis and natural history available in PubMed but they did not have a clearly outlined search strategy for disability evaluation to combine with the prognostic filter. In occupational health, although several search filters have been developed to locate studies on return to work, occupational health interventions and work participation in workers with a chronic disease, these are not specifically targeted toward the topic of disability evaluation.

As none of these occupational health filters or existing available prognostic filters suffice to identify studies on prognosis in the setting of disability evaluation, we set out to develop an adequate strategy to identify these studies. As a point of departure, knowing that sufficient efforts have been put into developing a search string for prognosis, we decided to use the Yale methodological research filter for prognosis and natural history to identify prognostic studies. Since there is no search strategy available for disability evaluation, we decided to develop a new search string to use in combination with the Yale prognostic filter. Subsequently, we planned to evaluate how the new combined search filter performs in terms of comprehensiveness, finding all relevant studies and efficiency, in terms of optimising the ratio of relevant to non-relevant studies in the yield. Comprehensiveness and efficiency will be judged for two different purposes. First, as a search strategy for the researcher, who does not want to miss too many relevant hits and cares less about finding many non-relevant articles. Second, as a search strategy for the practitioner, who has limited time and therefore does not want to find too many non-relevant articles. The search strategies should enable the identification of articles about the prognosis of not only disease-related impairments, but also of activity limitations and participation restrictions. We use the word ‘search strategy’ to refer to the comprehensive process of deciding on the resources or databases needed for the search and on the search terms and filters. The word ‘search filter’ is used for a concrete string of search terms used to identify studies in a database. Filters often consist of terms relating to study type and/or terms associated with the subject of the study. Examples are the ‘clinical queries’ filters in MEDLINE, but also the search strings developed by Gehanno et al. and Verbeek et al.

The objective of this study is therefore to develop and evaluate a comprehensive and efficient search strategy, including search filters in PubMed, to be used by either researchers or practitioners, which identifies articles about the prognosis of work disability.

**METHODS**

For our study we used a diagnostic test analytic framework procedure with the following three steps to develop and test the search strategy in PubMed. See also flow chart in figure 1 for an overview of how we created the search filters:

**Construction of the reference set**

First we developed a reference set of highly relevant prognostic articles in the field of disability evaluation, based on clear inclusion criteria, that we would like our search to find. Journals that publish both studies in the field of disability evaluation and studies on 10 prevalent chronic diseases, frequently subject to disability evaluation, were used for the reference set. First we selected three general medical journals publishing on a wide spectrum of diseases: BMJ, PLoS ONE and the Journal of the American Medical Association (JAMA). Next we selected three journals often publishing in the field of disability evaluation: Journal of Occupational Rehabilitation (JOR), Occupational and Environmental Medicine (OEM) and Scandinavian Journal of Work Environment and Health (SJWEH). Finally we included four high-impact disease-specific journals: Spine, Journal of Anxiety Disorders, Stroke and Cancer. We decided to select the 10-year period 2000–2009 to ascertain that all articles included would have been indexed for MEDLINE at the time of searching (2013), and screened all articles in this time period.

To develop inclusion criteria for the selection of articles most relevant for the prognosis of work disability, we considered the following PICO, adapted from Cornelius:

1. **Type of studies**: we included prospective and retrospective follow-up studies with a minimum follow-up period of 6 months. So studies with an inception cohort studying either prognostic factors or the so-called ‘natural’ course of a disease (with or without taking therapy into account) were included. Furthermore, we included studies reporting an RCT presenting data on a control group (usual care or ‘without treatment’) that enabled evaluation of the course of a disease or disease-related functioning. Reviews were excluded as we wanted to develop a search strategy for original studies. Survival analysis studies were also excluded because filters identifying these studies have already been developed.

2. **Type of participants**: we wanted the patients included in the studies to be more or less similar to the practice of work disability evaluation: workers with any disease, chronic or otherwise, claiming a work disability pension. Therefore, we included studies with...
participants aged 18–64 years that were fully or partially work disabled at the start of the study. For studies that reported only a mean age, we took a maximum of 60 years. Where no information about work disability was provided, we included studies for which we judged the consequences of disease to be so severe that this would nearly always lead to serious problems with work ability, such as for late stages of cancer. At baseline the participants had to be on sick leave for at least 4 weeks or had to present serious activity limitations in 50% of the cases in the population.

3. Type of outcome measures: to support a professional quality disability evaluation, we are especially interested in studies in which the outcomes are measured in line with concepts presented in the International Classification of Functioning, Disability and Health (ICF) model of WHO. These concepts and related terms are frequently used in studies in the field of rehabilitation, occupational medicine, insurance medicine and vocational training. The following outcomes were considered relevant for the prognosis of work disability: level of functioning at work, level of disability, level of work disability, level of work participation such as return to work rates. We also included the level of recovery or deterioration of symptoms and signs where these symptoms or signs are more or less equivalent to the level of functioning such as in patients with a major depressive disorder. These patients have, by definition, substantial mental limitations and, typically, work functioning problems. To be included, a study had to measure either an outcome according to the ICF or a symptom or sign that could be considered equivalent. See box 1 for an overview of the inclusion criteria (box 1).

To select the articles for the reference set, we applied a two-step procedure. In the first step, two of the authors (RK and BF) separately analysed all the titles and abstracts of the years 2008 and 2009 of the 10 journals and excluded all articles that obviously did not fulfil the inclusion criteria, based on a judgment of title and abstract. Subsequently, both authors judged the remaining articles after reading the full-text article (about 20–30 articles per year). Where the authors’ opinions differed, both (RK and BF) discussed the deviances until consensus was reached.

In the second step, we were able to improve the efficiency of the search process. First we checked whether all articles identified by screening the volumes 2008 and 2009 (a total of 40 articles) were included when only using the Yale methodological research filter for prognosis and natural history. When this was the case, we used the Yale filter for the remaining eight volumes to preselect the references for further screening. After this preselection, both authors continued the procedure described in step 1. The results of both steps produced the final reference set.

Creating search strategies and search filters

Next, we collected potential search terms both from these articles and based on our own expertise and tested how well these search terms and combinations of terms were able to identify the articles of the reference set developed in step 1.

In the final reference set of articles, discriminating text words, phrases and MeSH terms in titles and abstracts were identified and evaluated by an independent experienced information specialist, using two different approaches. First, we used the program GoPubMed/PubReminer to identify the most frequently occurring single-text words and MeSH terms both in the relevant and all non-relevant sets of articles. Second, we used the program Termine from the National Centre for Text Mining (NaCTem) to identify the frequency of phrases (2–5 terms) in titles and abstracts in the same sets of articles.

We considered a search term, respectively, a phrase, as discriminating when it fulfilled both the following selection criteria: (1) it occurred in at least 5% of the articles in the reference set; and (2) it occurred five times more often in relevant articles as in non-relevant articles.

To determine the ranking order of the selected search terms, we used a cross-product of both selection criteria. Finally, this method resulted in ranking lists of discriminating text words, MeSH terms and phrases.

The terms in these three lists were, in ranking order, subsequently combined with the Boolean operator OR in order to create search filters with a high comprehensiveness and efficiency for the next phase.

Performance criteria for search filters

Based on the best-performing individual search terms, we developed various strings of search terms, termed search filters, which we used in combination with the Yale prognostic filter. We calculated the comprehensiveness, efficiency, specificity and accuracy. See table 1 for an overview of formulae for calculating these operating...
Table 2 Total number of articles and articles included in the reference set from 10 selected journals in the publication years 2000–2009. In parentheses the percentages

<table>
<thead>
<tr>
<th>Name journal</th>
<th>Total articles (%)</th>
<th>Relevant articles (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJWEH</td>
<td>679 (1)</td>
<td>6 (2.7)</td>
</tr>
<tr>
<td>JOR</td>
<td>328 (0.5)</td>
<td>10 (4.4)</td>
</tr>
<tr>
<td>OEM</td>
<td>1654 (2.5)</td>
<td>4 (1.8)</td>
</tr>
<tr>
<td>Spine</td>
<td>6092 (9.3)</td>
<td>167 (74.2)</td>
</tr>
<tr>
<td>Stroke</td>
<td>6200 (9.4)</td>
<td>12 (5.3)</td>
</tr>
<tr>
<td>Cancer</td>
<td>7804 (11.9)</td>
<td>13 (5.8)</td>
</tr>
<tr>
<td>J Anx Dis</td>
<td>742 (1.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>JAMA</td>
<td>11 541 (17.6)</td>
<td>8 (3.6)</td>
</tr>
<tr>
<td>PloS ONE</td>
<td>8489 (12.9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>BMJ</td>
<td>22 163 (33.7)</td>
<td>5 (2.2)</td>
</tr>
<tr>
<td>Total</td>
<td>65 692 (100)</td>
<td>225 (100)</td>
</tr>
</tbody>
</table>
Work Disability Prognosis—Practice filter (WDP-P) having as characteristics, respectively, a comprehensiveness of 90% and efficiency with NNR=20 and a comprehensiveness of 68% and efficiency with NNR=10. See also online supplementary appendix 1 in which the two filters are in bold among the others that were tested.

Compared to other occupational health search filters, the new WDP-R and WDP-P filters performed considerably better in comprehensiveness (90% and 68%) compared to a maximum of 41% of the other filters (see table 3). The practice filter developed by Gehanno had a relatively high score on efficiency expressed more user-friendly as NNR=5, compared with an NNR=20 and NNR=13 for the WDP-R, respectively, the WDP-P. Unfortunately the comprehensiveness was rather low (17%, with 95% CI 12% to 22%), making this filter inappropriate for our goal. Although using only the Yale filter gave 100% comprehensiveness, by definition, it is far from efficient in that the NNR is as high as 103 (95% CI 91 to 119).

To illustrate the use of our filters for a patient with a specific disease and because search filters behave differently when combined with specific disease terms, we planned to apply both our search filters (WDP-R and WDP-P) in combination with the Yale methodological research filter for prognosis and natural history to identify articles about the prognosis of work disability. For the practitioner, we developed a version with a comprehensiveness of 90%, with 95% CI 86% to 94% and efficiency with NNR=20, with 95% CI 17 to 23 (WDP-R). For the practitioner, we developed a version with a comprehensiveness of 68%, with CI 61% to 74% and efficiency with NNR=13, with CI 11 to 15 (WDP-P). Since the latter missed three out of five articles in our three case studies we would advice practitioners also to use the WDP-R filter instead.

### DISCUSSION

We developed and evaluated a search filter to be used in PubMed in combination with the Yale methodological research filter for prognosis and natural history to identify articles about the prognosis of work disability. For the practitioner, we developed a version with a comprehensiveness of 90%, with 95% CI 86% to 94% and efficiency with NNR=20, with 95% CI 17 to 23 (WDP-R). For the practitioner, we developed a version with a comprehensiveness of 68%, with CI 61% to 74% and efficiency with NNR=13, with CI 11 to 15 (WDP-P). Since the latter missed three out of five articles in our three case studies we would advice practitioners also to use the WDP-R filter instead.

### Strengths and limitations of our study

This is the first study describing the development and evaluation of a search strategy and search filters for the prognosis of work disability. Disability evaluation is a common task for a variety of medical disciplines all over the world, with a high impact for the work participation and financial compensation of working patients involved. Therefore, a better scientific foundation of this task should receive a high priority. A validated search strategy for identifying studies on the prognosis of work disability may contribute to a more evidence-based medical practice.

A strength of our study is that we used a reference set, or gold standard, of articles that could be identified in a following step through a newly developed search strategy and filters. The set was constructed based on inclusion criteria for studies relevant to the topic chosen, selecting all relevant studies from a large set of studies present in MEDLINE, deemed relevant for evidence-based disability evaluation. We used a comprehensive text mining method to find potentially relevant search terms. Next, we decided on clear criteria to determine the minimum performance of our new search filter in two versions for two groups of users with differences in demands and needs: researchers and practitioners. We simulated the use in practice of the new strategy and filters for three disorders with low, medium and high numbers of relevant studies in MEDLINE, and for which physicians frequently perform a disability evaluation. The problem with search filters is that they behave differently when combined with specific disease terms: prognosis in general will yield different results from prognosis in cancer or in rheumatoid arthritis. We believe that, therefore, the case studies are more useful than validation sets that just show that the search strategy can be replicated in general but do not tell how the strategy behaves in reality when combined with specific and relevant

### Table 3 Comparison of search filter performance scores on comprehensiveness and efficiency expressed more user-friendly as the Number Needed to Read (NNR=1/precision), related to the capability to identify relevant articles on prognosis of work disability

<table>
<thead>
<tr>
<th>Comprehensiveness (%) (95% CI)</th>
<th>Efficiency (NNR) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yale filter only</td>
<td>100</td>
</tr>
<tr>
<td>Work disability</td>
<td>90 (86 to 94)</td>
</tr>
<tr>
<td>Prognosis filter (R)</td>
<td>68 (61 to 74)</td>
</tr>
<tr>
<td>Prognosis filter (P)</td>
<td>35 (28 to 41)</td>
</tr>
<tr>
<td>Gehanno filter</td>
<td>41 (35 to 48)</td>
</tr>
<tr>
<td>Gehanno filter</td>
<td>16 (11 to 21)</td>
</tr>
<tr>
<td>Gehanno filter</td>
<td>31 (25 to 37)</td>
</tr>
<tr>
<td>Gehanno filter</td>
<td>17 (12 to 22)</td>
</tr>
</tbody>
</table>

A research (R) and practice (P) search filter on prognosis of work disability are compared with five other occupational health search filters developed by Haafkens, Verbeek (work component only of R and P) and Gehanno (R and P). All filters were applied in combination with the Yale methodological research filter for prognosis and natural history.
Table 4  Number of (relevant) hits on prognosis of work disability in PubMed in a restricted time period in three diseases: rheumatoid arthritis, major depression and cystic fibrosis (using MeSH terms)

<table>
<thead>
<tr>
<th>Diseases selected by MeSH terms</th>
<th>Rheumatoid arthritis</th>
<th>Major depression</th>
<th>Cystic fibrosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hits in PubMed (restricted time period)</td>
<td>624 (1 month)</td>
<td>473 (2 months)</td>
<td>682 (7 months)</td>
</tr>
<tr>
<td>Number of relevant hits</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Number of hits by adding Yale filter (total number of articles found)</td>
<td>1 (262)</td>
<td>2 (260)</td>
<td>2 (265)</td>
</tr>
<tr>
<td>Comprehensiveness (CI)</td>
<td>1.0 (NA)</td>
<td>1.0 (NA)</td>
<td>1.0 (NA)</td>
</tr>
<tr>
<td>Efficiency (CI)</td>
<td>262 (5 to 11 111)</td>
<td>130 (36 to 1075)</td>
<td>133 (4 to 1099)</td>
</tr>
<tr>
<td>Number of hits by adding Yale filter in combination with WDP-R filter (total number of articles found)</td>
<td>1 (72)</td>
<td>2 (80)</td>
<td>2 (42)</td>
</tr>
<tr>
<td>Comprehensiveness (CI)</td>
<td>1.0 (NA)</td>
<td>1.0 (NA)</td>
<td>1.0 (NA)</td>
</tr>
<tr>
<td>Efficiency (CI)</td>
<td>72 (1 to 3333)</td>
<td>40 (1 to 133)</td>
<td>21 (6 to 172)</td>
</tr>
<tr>
<td>Number of hits by adding Yale filter in combination with WDP-P filter (total number of articles found)</td>
<td>1 (44)</td>
<td>0 (36)</td>
<td>1 (20)</td>
</tr>
<tr>
<td>Comprehensiveness (CI)</td>
<td>1.0 (NA)</td>
<td>0 (NA)</td>
<td>0.50 (0.01 to 0.99)</td>
</tr>
<tr>
<td>Efficiency (CI)</td>
<td>44 (15 to 47)</td>
<td>NA</td>
<td>20 (7 to 22)</td>
</tr>
</tbody>
</table>

For each search filter or combination of filters comprehensiveness (sensitivity) and efficiency, expressed more user-friendly as the Number Needed to Read (NNR=1/precision), were calculated. In parentheses the 95% CI.


disease terms. This is especially true in this study in which the WDP-P filter missed three out of five relevant articles compared to none missed by the WDP-R filter. Therefore, our advise to practitioners, as well as researchers, would be to use only the research filter.

As with any study on search filters, the construction of a reference set of articles has been based on a deliberately chosen but restricted sample of journals available in MEDLINE. Although the reference set included sufficient numbers of relevant articles, a topic that is not well studied will have had a low chance of being included in this sample. However, we believe that the choices made represent a good compromise. Although the number of 225 relevant articles is large enough to yield credible results, splitting this set into a reference set and a validation set, would have substantially reduced the power of our study, without adding to the validation. For this reason, we did not use a validation set. Moreover the number of journals is sufficiently diverse to represent the most relevant journals in MEDLINE. However, we had not expected the large number of articles in the reference set published in *Spine* with 167 out of the 225 articles (74%). We decided not to change the choices made, realising that musculoskeletal disorders like chronic low back pain, chronic complaints of arm, neck and shoulder and rheumatoid arthritis are frequent causes of work disability and are among the best studied diseases in relation to the prognosis of work disability. Furthermore, there is no reason to believe that the journal *Spine* would be indexed differently than the other included journals leading to the introduction of bias, which otherwise could have been a reason to change our choices. However, to avoid the development of a disease-specific search filter instead of a more generic filter for the prognosis of work disability, we removed disease-specific words during the development of the new search filters. The fact that the filter performed so well, for example, in the use for a worker with cystic fibrosis or with a major depressive disorder, underlines that this was a successful method.

Comparison with other studies

A systematic review of search filters applied in reviews of prognostic studies showed that no prognostic filter was more comprehensive than 95% with an efficiency expressed as NNR around 10, illustrating the less satisfactory operating characteristics of prognostic filters in comparison with therapeutic filters. Against this background, the prognostic search strategy and filter in the review version (WDP-R) with a comprehensiveness of 90% and an efficiency with NNR=20 is not unusual. Both new filters perform better in comprehensiveness than other occupational health search filters, demonstrating the advantage of developing specific filters for prognostic studies related to work disability.

On the other hand, our research filter researchers would still miss 10% of the relevant studies, and practitioners would need to show persistence in going through 120 references to identify four relevant studies. In diseases such as rheumatoid arthritis, with a large number of relevant hits, practitioners will need to select reviews to reduce the workload of going through too many titles after applying our filter.

It could be that a more sophisticated process for search filter development would have produced better results. Garg et al. used an automated process for combining and testing filters by using a computer algorithm. It would be worthwhile to see if such an algorithm could create search filters with higher comprehensiveness (>95%) and an acceptable efficiency with NNR around 10. Or to create a search filter with acceptable comprehensiveness (>65-70%) and very good efficiency with...
which will likely improve the metrics of the reference set by for instance including more journals MEDLINE. Another possibility would be to enlarge the MEDLINE has greatly increased the comprehensiveness better tagging of randomised controlled trials in occupational health version), as the comprehensiveness is superior to existing. Therefore, in more prevalent diseases in which subsequently more research is undertaken practitioners are be advised to filter on review articles after applying the WDP-R filter. The example of cystic fibrosis, in which only 200 titles over a 3-year period have to be screened after the same time period of 3 years, shows that with more rare diseases application of the WDP-R filter is efficient for researchers and practitioners alike. The above examples nicely illustrates that the performance of a search filter is indeed dependent on the total number of publications of the disease involved.

Especially to improve the feasibility for practice, new efforts are needed to enhance search filter performance with the ultimate goal of improving patient care.\textsuperscript{15} Since better tagging of randomised controlled trials in MEDLINE has greatly increased the comprehensiveness of searches for RCTs, one aspect could be to better tag studies on prognosis and work disability evaluation in MEDLINE. Another possibility would be to enlarge the reference set by for instance including more journals which will likely improve the metrics of the filters as Yao\textsuperscript{27} showed. Also specialised bibliographic software that is capable to make every combination of search terms with optimisation for comprehensiveness and effectiveness, like the Hedge team uses, could further improve these filters.\textsuperscript{8}

For researchers and practitioners alike, it is worthwhile to enlist the help of information specialists in developing and running search strategies. We recommend the development of expert centres with a helpdesk to improve knowledge translation in practice.

**CONCLUSION**

The Work Disability Prognosis filter (WDP-R) will help practitioners and researchers who want to find prognostic evidence in the area of work disability evaluation. However, as we discussed above, further refining of these filters is possible and needed, especially for the practitioner for whom efficiency is especially important.

**Acknowledgements**

The authors thank Joost Daams for providing us with the data of the text analysis with the appropriate software programs and for his advice in identifying the correct set of articles in the inclusion and exclusion set.

**Contributors**

RK, JH, JHV were involved in the concept and methodology of the article; gathering of the material and reference; analysis; formulation of the text. BF was involved in the gathering of material and reference; formulation of the text. FJ HvD was involved in the concept and methodology of the article; formulation of the text.

**Funding**

The project was financed by the National Institute for Employee Benefit Schemes (UWV) as part of the Research Center for Insurance Medicine. However, no funding bodies had any role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Competing interests**

None declared.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Data sharing statement**

No additional data are available.

**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-commericially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

**REFERENCES**


