BMJ Open

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 Hoving[1,2]

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.

[1]Research Center for Insurance Medicine AMC-UMCG-UWV-VUMC, Amsterdam, The Netherlands
[2]Coronel Institute of Occupational Health, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[3]Surveillance and Reviews Team, Finnish Institute of Occupational Health, Kuopio, Finland

Correspondence to
Dr Rob Kok;
r.kok07@gmail.com

Strengths and limitations of this study

- Searching the medical literature for evidence on prognosis is an important aspect of evidence-based disability evaluation.
- This is the first study to describe the development and evaluation of a search strategy and search filters for the prognosis of work disability.
- The Work Disability Prognosis filters will help practitioners and researchers who want to find prognostic evidence in the area of work disability evaluation.
- Further refining of these filters is needed, especially for the practitioner for whom efficiency is especially important.

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]

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

In their study on prognostic factors for work ability in sick-listed employees with chronic diseases, Slebus et al[9] 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,[10] occupational health interventions[11] and work participation in workers with a chronic disease,[12] these are not specifically targeted toward the topic of disability evaluation.

As none of these occupational health filters or existing available prognostic filters[13 14] 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.[14] 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.[15] 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.[10 11 16] 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,[13] but also the search strings developed by Gehanno et al[10] and Verbeek et al.[11]

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[10 17 18] with the following three steps to
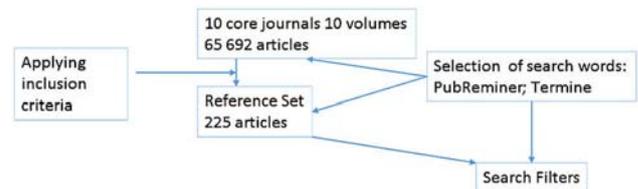
Figure 1 flow chart of articles and search terms.

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:[19]

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.[4] [20] [21] 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.[14] 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.[18] First, we used the program *GoPubMed/PubReminer*[22] to identify the most frequently occurring single-text words and MeSH terms both in the relevant and all non-relevant sets of articles.[18] Second, we used the program *Termine*[23] 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.[16] 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.[10] [15]

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

---

Box 1   Criteria for inclusion of articles in reference set.

types of studies
- ▶ follow-up studies AND
- ▶ with a minimum follow-up period of 6 months AND
- ▶ original studies, no reviews

types of participants
- ▶ age between 18–64 years, mean age max 60 years AND
- ▶ humans AND
- ▶ ill AND
- ▶ consequently they must be on sick leave for a longer period (4 weeks) OR have serious activity limitations in 50% of the cases

types of outcome measures
- ▶ improvement of functioning OR reduction of disability OR increase of work participation OR return to work OR symptom/sign recovery (when equivalent with functioning like recovery of cognitive limitations in major depression).

Table 1 Formulae for the calculation of operating characteristics for a search filter for locating studies in MEDLINE

	Relevant articles	Non-relevant articles	Total
Search term +	a (true positive)	b (false positive)	a+b
Search term −	c (false negative)	d (true negative)	c+d
Total	a+c	b+d	a+b+c+d

Comprehensiveness=Sensitivity=Recall=a/(a+c).
Efficiency (=Precision=a/(a+b)), expressed more user-friendly as Number Needed to Read (=1/precision).
Specificity=d/(b+d).
Accuracy=(a+d)/(a+b+c+d).

characteristics (table 1). Please note that various terms are used for comprehensiveness such as sensitivity and recall. Although efficiency and precision are used as equivalent terms, we choose to measure efficiency as the Number Needed to Read (NNR=1/precision) to identify one relevant article.

We prefer the terms comprehensiveness and efficiency because they are intuitively easier to understand. For the search string to use in practice, we decided that the efficiency, expressed more user-friendly as NNR, should be 10 at most, in combination with a comprehensiveness of ≥65%.[16] For the search string to use in research, we decided on an efficiency (NNR) of 60 at most in combination with a comprehensiveness of ≥90%.

Finally we compared the performance of our search filters with those of other search filters developed for occupational health[10–12 16] to see if our newly developed filter performed better than existing filters in the broader field of work and health. All in combination with the Yale methodological research filter for prognosis and natural history.[14] In case of the filter from Verbeek et al,[11] developed to identify occupational health interventions, we only used the work component of this filter and not the intervention component, for better comparison.

We calculated 95% CIs for proportions as described by Maceneaney et al.[24]

In addition, we illustrated the use of the new filters for three common diagnoses in the field of disability evaluation to show and explore what their use means in practice of disability evaluation. In this practice search filters are often used for a search when literature information can support the assessment of work disability of a patient with one specific disease. As the total number of relevant articles available for a specific disease, given a certain time period, is different for various diseases, the number of titles that has to be screened after using a filter also depends on the specific disease. So, in order to illustrate the use of the new filters in practice, we applied these on three different diseases, with respectively high, moderate and low numbers of relevant articles in MEDLINE: rheumatoid arthritis, depressive disorders and cystic fibrosis. The time-periods for articles on each disease, (respectively, 1, 2 and 7 months) were chosen because the number of studies on the three topics varied enormously and we wanted the samples to be comparable in size. Again the filters were used in combination with the Yale filter and the same three criteria for relevance as depicted in box 1.

RESULTS

The total number of articles in MEDLINE for the 10 journals in 10 years was 65 692 of which 225 were included in the reference set (table 2).

We identified 16 search terms or combinations of search terms that occurred in 5% or more of the relevant articles and also occurred five times more often in the relevant articles than in the non-relevant ones. These terms were combined one by one, and for each combination, together with the Yale methodological research filter for prognosis and natural history, we calculated the performance characteristics (see online supplementary appendix 1). When a term could be replaced by an overarching term, it was omitted, for example, 'sick leave' was omitted when the term 'sick' was added, since all articles with the term 'sick leave' in title or abstract are also identified by the term 'sick'. Also 'sick leave' as a MeSH term was omitted, because it did not add any relevant or non-relevant article, which reduces the total number of combinations of search terms to 14 (see online supplementary appendix 1).

The best-performing search strings, in combination with the Yale methodological research filter for prognosis and natural history, were named as the Work Disability Prognosis—Research filter (WDP-R) and the

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

Name journal	Total articles (%)	Relevant articles (%)
SJWEH	679 (1)	6 (2.7)
JOR	328 (0.5)	10 (4.4)
OEM	1654 (2.5)	4 (1.8)
Spine	6092 (9.3)	167 (74.2)
Stroke	6200 (9.4)	12 (5.3)
Cancer	7804 (11.9)	13 (5.8)
J Anx Dis	742 (1.1)	0 (0)
JAMA	11 541 (17.6)	8 (3.6)
PloS ONE	8489 (12.9)	0 (0)
BMJ	22 163 (33.7)	5 (2.2)
Total	65 692 (100)	225 (100)

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 gives 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[14] to the three diagnostic groups. The WDP-P filter missed three out of the five relevant articles compared to none missed by the WDP-R filter. For 1-month of publications in PubMed for rheumatoid arthritis, 262 articles including only one relevant article, applying the Yale prognostic filter and our WDP-R filter 72 titles were identified (table 4), in which the relevant article was still present. However, the corresponding number of titles to be screened in a period of 3 years for this disease would be about 2500. For cystic fibrosis, following the same procedure, we found 42 titles in a 7-month period, corresponding to about 200 titles over a 3-year period.

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[14] to identify articles about the prognosis of work disability. For the researcher, 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

	Comprehensiveness (%) (95% CI)	Efficiency (NNR) (95% CI)
Yale filter only[14]	100	103 (91 to 119)
Work disability Prognosis filter (R)	90 (86 to 94)	20 (17 to 23)
Work disability Prognosis filter (P)	68 (61 to 74)	13 (11 to 15)
Filter Haafkens[12]	35 (28 to 41)	19 (16 to 25)
Filter Verbeek[1] (R)	41 (35 to 48)	57 (47 to 73)
Filter Verbeek[1] (P)	16 (11 to 21)	39 (29 to 58)
Filter Gehanno[10] (R)	31 (25 to 37)	11 (9 to 14)
Filter Gehanno[10] (P)	17 (12 to 22)	5 (4 to 8)

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)

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

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 calcualted. In parentheses the 95% CI.
NA, not applicable; WDP-P, Work Disability Prognosis—Practice filter; WDP-R, Work Disability Prognosis—Research filter.

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 diability, we removed disease-specific words during the development

of the new search filters. The fact that the filter peformed 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,[25] 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 research 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, with 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*[26] 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

NNR around 1–2, which would make it more attractive for a practitioner to search for evidence.

Practical implications for science and practice

To researchers starting new studies, and all those experts and practitioners who are preparing systematic reviews or developing knowledge products such as evidence-based guidelines and recommendations, but also to practitioners in need for information in daily practice, we recommend using our strategy and filter (research version), as the comprehensiveness is superior to existing occupational health filters and the NNR is satisfactory.

However, the example of the case study rheumatology, in which applying the WDP-R filter still leaves approximately 2500 articles to be screened in 3 years of PubMed, implies an excessive workload for the practitioner in need for information in daily practice. 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.[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[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.[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.

REFERENCES

1. de Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. *Prat Organ Soins* 2007;38:205–17.
2. SSA. *Social security programs throughout the world: Europe 2012*. Washington: SSA, 2012; Publication No. 13-11801.
3. Gezondheidsraad. *Verzekeringsgeneeskundige protocollen: Algemene inleiding, Overspanning, Depressieve stoornis (Dutch Health Council: Protocols for insurance physicians: general introduction, adjustment disorder, depressive disorder)*. Den Haag: Gezondheidsraad, 2006; Publicatie nr 2006/22.
4. World Health Organization (WHO). *International classification of functioning, disability and health: ICF*. Geneva: World Health Organisation, 2001 http://www.who.int/cassifications/icf/site/beginners/bg.pdf
5. Kok R, Hoving JL, Verbeek J, *et al*. Integrating evidence in disability evaluation by social insurance physicians. *Scand J Work Environ Health* 2011;37:494–501.
6. Anner J, Kunz R, Boer WD. Reporting about disability evaluation in European countries. *Disabil Rehabil* 2014;36:848–54.
7. Kok R, Hoving JL, Verbeek JH, *et al*. Evaluation of a workshop on evidence-based medicine for social insurance physicians. *Occup Med (Lond)* 2008;58:83–7.
8. Wilczynski NL, McKibbon KA, Walter SD, *et al*. MEDLINE clinical queries are robust when searching in recent publishing years. *J Am Med Inform Assoc* 2013;20:363–8.
9. Slebus FG, Kuijer PP, Willems JH, *et al*. Prognostic factors for work ability in sicklisted employees with chronic diseases. *Occup Environ Med* 2007;64:814–19.
10. Gehanno JF, Rollin L, Le Jean T, *et al*. Precision and recall of search strategies for identifying studies on return-to-work in Medline. *J Occup Rehabil* 2009;19:223–30.
11. Verbeek J, Salmi J, Pasternack I, *et al*. A search strategy for occupational health intervention studies. *Occup Environ Med* 2005;62:682–7.
12. Haafkens J, Moerman C, Schuring M, *et al*. Searching bibliographic databases for literature on chronic disease and work participation. *Occ Med* 2006;56:39–45.
13. Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Med* 2004;2:23.
14. Yale Methodological Research filter for Prognosis and Natural History; accessed 6 Mar at http://library.medicine.yale.edu/tutorials/577
15. Shariff SZ, Sontrop JM, Haynes RB, *et al*. Impact of PubMed search filters on the retrieval of evidence by physicians. *CMAJ* 2012;184:E184–90.
16. Schaafsma F, Hulshof C, Verbeek J, *et al*. Developing search strategies in Medline on the occupational origin of diseases. *Am J Ind Med* 2006;49:127–37.

17. Iansavichus AV, Haynes RB, Shariff SZ, *et al*. Optimal search filters for renal information in EMBASE. *Am J Kidney Dis* 2010;56:14–22.
18. van de Glind EM, van Munster BC, Spijker R, *et al*. Search filters to identify geriatric medicine in Medline. *J Am Med Inform Assoc* 2012;19:468–72.
19. Cornelius LR, van der Klink JJ, Groothoff JW, *et al*. Prognostic factors of long term disability due to mental disorders: a systematic review. *J Occup Rehabil* 2011;21:259–74.
20. Brage S, Donceel P, Falez F. Disability of ICF core set for disability evaluation in social security. *Disabil Rehabil* 2008;30:1392–6.
21. Anner J, Schwegler U, Kunz R, *et al*. Evaluation of work disability and the international classification of functioning, disability and health: what to expect and what not. *BMC Public Health* 2012;12:470.
22. Koster JB. Pubmed Pubreminer. Detailed analysis of PubMed Search results. Webbased application. http://bioinfo.amc.uva.nl/human-genetics/pubreminer/ (accessed Aug 2013).
23. Frantzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms. *Int J Digit Libr* 2000;3:117–32.
24. Maceneaney PM, Malone DE. The meaning of diagnostic test results: a spreadsheet for swift data analysis. *Clin Radiol* 2000;55:227–35.
25. Chatterley T, Dennett L Utilisation of search filters in systematic reviews of prognosis questions. *Health Info Libr J* 2012;29:309–22.
26. Garg AX, Iansavichus AV, Wilczynski NL, *et al*. Filtering Medline for a clinical discipline: diagnostic test assessment framework. *BMJ* 2009;339:b3435.
27. Yao X, Wilczynski NL, Walter SD, *et al*. Sample size determination for bibliographic retrieval studies. *BMC Med Inform Decis Mak* 2008;8:43.

Appendix 1: Comparison of search terms, which were combined one by one (in ranking order of sensitivity). For each combination of search terms, together with the Yale methodological research filter for prognosis and natural history, we calculated the comprehensiveness (%) and efficiency User friendly expressed as the Number Needed to Read (NNR). In bold the WDP-P and WDP-R. In parentheses the 95% Confidence Interval (CI).

Combination of search terms	Comprehensiveness (95% CI)	Efficiency (95% CI)
VAS	11 (7-15)	5 (3-7)
VAS OR "disability index"	31 (25-37)	4 (3-5)
VAS OR "disability index" OR disability evaluation[mh]	52 (46-59)	6 (5-8)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog"	55 (48-61)	7 (6-8)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement (WDP-P)	**68 (61-74)**	**13 (11-15)**
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh]	71 (65-77)	14 (12-16)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "sick leave"	76 (70-81)	14 (12-16)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "sick leave" OR "functional outcome"	77 (77-83)	14 (12-16)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "sick leave" OR "functional outcome" OR sick	78 (72-83)	14 (13-17)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function"	78 (73-84)	14 (13-17)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function" OR "quality of life"	86 (82-91)	16 (14-19)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function" OR "quality of life" OR sick leave[mh]	86 (82-91)	16 (14-19)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function" OR "quality of life" OR "functional status"	87 (82-91)	16 (14-19)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function" OR "quality of life" OR "functional status" OR "health status"	87 (82-91)	16 (14-19)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function" OR "quality of life" OR "functional status" OR "health status" OR functioning	88 (84-93)	18 (16-21)
VAS OR "disability index" OR disability evaluation[mh] OR "visual analog" OR improvement OR recovery of function[mh] OR "functional outcome" OR sick OR "physical function" OR "quality of life" OR "functional status" OR "health status" OR functioning OR activities (WDP-R)	**90 (86-94)**	**20 (17-34)**