Development and evaluation of a quality assessment instrument for occupational physicians


Published in: 
Occupational and Environmental Medicine

DOI: 
10.1136/oem.55.6.375

Citation for published version (APA): 
Development and evaluation of a quality assessment instrument for occupational physicians

WE van der Weide, JH Verbeek, FJ van Dijk and CT Hulshof


Updated information and services can be found at:
http://oem.bmj.com/cgi/content/abstract/55/6/375

These include:

**References**

5 online articles that cite this article can be accessed at:
http://oem.bmj.com/cgi/content/abstract/55/6/375#otherarticles

**Rapid responses**

You can respond to this article at:
http://oem.bmj.com/cgi/eletter-submit/55/6/375

**Email alerting service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
http://www.bmjjournals.com/cgi/reprintform

To subscribe to *Occupational and Environmental Medicine* go to:
http://www.bmjjournals.com/subscriptions/
Development and evaluation of a quality assessment instrument for occupational physicians

Willeke E van der Weide, Jos H A M Verbeek, Frank J H van Dijk, Carel T J Hulshof

Abstract

Objectives—To develop and apply a method for assessing the quality of the process of occupational health care for individual patients.

Methods—The scientific literature was studied to develop a method to assess the quality of the process of occupational rehabilitation for workers with low back pain. The method was applied to health care and university workers with low back pain who were rehabilitated by their occupational physicians.

Results—Assessment of quality of care is regarded as a four step approach. Firstly, guidelines should be developed and implemented. Secondly, indicators for quality and criteria to demarcate good and deviant quality were derived from the guidelines. Thirdly, a method for data collection was chosen. Finally, quality was scored. For occupational rehabilitation, there was some deviance from the guidelines for most cases, especially in continuity of care with a deviant rate of 47%. Other indicators deviated from 1.4%–17.4%. Occupational physicians agreed on the relevance of the indicators and criteria, but for three indicators they evaluated the criteria as too rigid. They did not agree with their own performance scores in 66% of the deviant cases.

Conclusion—Assessing the quality of the process of occupational health care with this method is an asset to present methods, but more specific criteria are needed for a more sensitive assessment.

Keywords: quality assessment method; guidelines; evaluation

Quality of care attracts increasing interest in different fields of health care—such as hospitals, ambulatory settings, and general practice. 1-3 Compared with these fields quality assessment in occupational health care is largely undeveloped. Occupational health consultation and occupational rehabilitation are good examples. Occupational rehabilitation is the most time consuming task of occupational physicians in The Netherlands, 4-5 but we hardly know what “good quality practice” is and how it should be assessed. At the same time there is an increasing demand for accountability of occupational health services due to commercialisation and competition. To be able to account for the use of economic resources the development of a valid method for quality assessment is of paramount importance.

Unlike most other fields of medicine, in occupational rehabilitation many parties are necessary to yield good results. Apart from patients and occupational physicians, in most cases employers, department staff and colleagues, general practitioners, medical specialists, and insurance companies are involved. Thus, return to work as an outcome variable for occupational rehabilitation is influenced by many variables. Even if the outcome is negative the occupational physicians may have performed quite well. So, it is sensible to look firstly at the process of care. Furthermore, process data are usually more sensitive and more informative measures of quality than outcome data. 6-7 Variables to measure the quality of occupational health care are still not readily available. Some studies, especially from the United Kingdom, Finland, and the United States, refer to this topic. 8-13 However, a clear basis for the definition of good quality practice is often lacking. 8-10 or the emphasis is on procedures instead of the content of care. 11-14 Therefore, we decided to develop and evaluate a method for defining and measuring the quality of the occupational physician’s contribution to occupational rehabilitation.

In this paper we present the development of a method for quality assessment in occupational health care based on publications from other fields of medicine. This method was developed for quality assessment of occupational rehabilitation of workers with low back pain in a research setting, but the situation is comparable with the assessment of quality in everyday practice. In the results section we describe the published findings as well as the choices made for occupational rehabilitation. These two parts are not separated to facilitate understanding. We also report on evaluative aspects of the method.

Methods

We performed a literature survey in the medline database from 1981 to 1995. We selected non-specific reviews—that is, not about a specific disease—on practice guidelines, process assessment, peer review, medical audit, health services research, physician practice patterns, clinical protocols, consensus development conferences, diffusion of innovation, and feedback. References in pertinent publications were checked for relevant studies, and we used citations provided by colleagues.

For our quality assessment method we decided firstly on scientific evidence for the efficacy of interventions and of implementation
methods, secondly on methodological criteria such as assumed validity and reliability, and thirdly on acceptability in practice. However, our process of development was restrained by time and budget, both not unusual in practice.

For a first evaluation of the quality assessment method we used the setting of a clinical trial about the efficacy of occupational rehabilitation for healthcare workers and university workers (1995–7). Nineteen occupational physicians from eight occupational health services in The Netherlands participated in this trial. One hundred and twenty consecutive workers who were on sick leave with low back pain for at least 10 days were randomly assigned to an intervention group (an occupational rehabilitation programme (n=61)) or to a control group (support by their superior only (n=59)). Two patients in the intervention group did not visit the occupational physician although 14 patients from the control group did. So, 73 workers received occupational rehabilitation by an occupational physician. We report on 69 of these 73 workers because of missing data of four patients.

At the end of the trial the occupational physicians received a questionnaire for a limited evaluation of the quality assessment method. One occupational physician did not answer the questionnaire. Technical (validity, reliability), process (acceptability), and strategic (utility) criteria were differentiated in the evaluation.15

Results
DEVELOPMENT
The development of a method for assessing quality was divided into four steps.16–23 The first step consisted of development, dissemination, and implementation of guidelines taken together. Grimshaw and Russell distinguish these elements. They defined dissemination as interventions aimed at influencing clinicians’ attitudes, awareness, knowledge, and understanding of guidelines, and implementation as interventions aimed at improving clinicians’ compliance with guideline recommendations.24 However, in practice all three phases were so closely connected, that we combined them in this first step. In the second step indicators for quality were derived from these guidelines. For every indicator criteria were formulated which demarcated appropriate and deviant care. These indicators and criteria enabled the next step of data collection in the field. The fourth and final step was to set up a scoring method, which led to a final assessment of good or deviant quality.

Step 1: development and implementation of practice guidelines
Guidelines provide systematically developed statements to assist practitioner decisions about appropriate health care for specific clinical circumstances.25 It is widely accepted that scientific evidence should form the basis of these guidelines.26–29 To fill in existing gaps in scientific knowledge “experts” should be consulted to reach consensus, which is supposed to be better than random variation between doctors.30 Different structured methods for consultation can be used—such as the nominal group technique, a highly structured meeting with experts, or the Delphi method in which questionnaires are used for consulting experts.31 Both methods try to reach a degree of consensus about statements on a certain topic—for example, indications for a specific intervention.

Factors that influence the applicability of guidelines are involvement of practitioners in its construction, and a preliminary trial. The acceptability is affected by the maintenance of professional autonomy, adjustment to local circumstances, and individual instruction.25–30 Furthermore, clinical behaviour of physicians is influenced by financial incentives or disincentives and by patient characteristics—such as expectations of treatment and level of knowledge.31–34 Reminders at the time of consultation and patient specific feedback were confirmed to be effective in sustaining practice changes.35–38

Our draft guidelines for occupational rehabilitation for this trial were mainly based on guidelines for general practitioners in different countries39–44 and on a review of the literature about effective interventions for patients with low back pain.38

The guidelines can be divided into two parts: diagnostics and interventions. Three diagnostic categories for low back pain were distinguished, based on medical history and a sensitive test for nerve root compression, the straight leg raising test.39–40 Prognostic factors that have been shown to be related to the duration of low back pain should be assessed—namely, impairments,41–42 psychosocial problems,43–46 heavy work,47–51 organisational problems,48–53 and inadequate treatment.49 Experts in occupational medicine and neurology were consulted for comment.

In the second part of the guidelines effective interventions were described in these areas. Encouragement of physical activity and avoiding bed rest were interventions proved to be effective in shortening time until return to work.49 Decisions about work related interventions, communication with the treating physicians, and time until follow up consultation had to be based on consensus due to lack of evidence.

In a pilot study, the guidelines were used by a group of eight occupational physicians for 44 patients. In two meetings the guidelines were discussed and adjusted based on experiences with real patients and also on hypothetical cases. When we introduced the guidelines among the physicians for the trial with healthcare workers and university workers, they followed the same process to enhance the acceptability. The flexibility of the guidelines was guaranteed as much as possible by leaving latitude for individual clinical judgement and patient preferences. The occupational physicians had to complete a specific registration form summarising the guidelines, which was not only used for data collection but also for reviewing the guidelines. The medical records of relevant employees were tagged so that the occupational physicians were reminded of the
Step 2: selection of indicators and criteria

An indicator for the quality of care was defined as an essential element of care which gives a valid and representative picture of this quality. Indicators can be selected for structure, process, or outcome aspects depending on the element of interest.59 Indicators should be made operational by criteria. A criterion answers the question of “what represents appropriate or desirable care?” and demarcates good from deviant care for a specific indicator. However, the content of care for a patient partly depends on specific patient characteristics. Patients that are more severely ill require different or more intense care than others. Therefore, algorithmic criteria should be used which are adaptable to individual patient characteristics. In this “criteria mapping”56 57 or “branching logic” method,16 process criteria are formulated in “if-then” statements—namely, if a patient has a certain characteristic then the physician should perform a specific action. This method better predicts a favourable outcome of care than applying the same criteria to every patient.15 16 Furthermore, these explicit criteria are objective, can be used by external assessors, and have a high reliability.16 99

Process indicators can be grouped into categories—for instance, into a diagnostic and therapeutic category, or into categories related to specific problems. Finally, a single visit or a whole care episode can be represented in the indicators.10

Because of our interest in the process of care as far as the occupational physician is involved, we focused on the performance of this physician. Therefore, process indicators were called performance indicators. From the guidelines we derived seven performance indicators with explicit criteria in the form of a criteria map (table 1). We used the algorithm that if a problem exists that delays return to work the occupational physician should try to solve that problem. Because of the branching logic, for each indicator several criteria can be met, not

### Table 1 Performance indicators and criteria for rehabilitation of workers with low back pain

<table>
<thead>
<tr>
<th>Number</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Encouragement of activity (applicable for non-specific low back pain only):</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>Psychosocial problems:</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>Inadequate treatment:</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>Occupational disabilities:</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>Organisational impediments:</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>Advice for return to work:</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>Continuity of care (not applicable if return to work full time within 2 weeks was advised):</td>
<td>A</td>
</tr>
</tbody>
</table>

guidelines at the time of consultation. During the research period every second or third month a 90 minute meeting was organised for peer review about actual care of specific patients. Deviations from the guidelines were discussed. Occupational physicians were trained, with help from an actor, in their communication with employees with back pain. One of the researchers visited physicians that enrolled in the study, discussed the guidelines with them, and trained them. The physicians were informed about recent developments in the field.

Because of our interest in the process of care as far as the occupational physician is involved, we focused on the performance of this physician. Therefore, process indicators were called performance indicators. From the guidelines we derived seven performance indicators with explicit criteria in the form of a criteria map (table 1). We used the algorithm that if a problem exists that delays return to work the occupational physician should try to solve that problem. Because of the branching logic, for each indicator several criteria can be met, not
be met (deviant), or be inapplicable. The example (figure) shows the criteria for inadequate treatment. For every indicator criteria were formulated for a whole episode, from the first consultation until return to work. The criteria for each indicator were thoroughly discussed between the two assessors (Wvd-W,Jv) on the basis of real cases after which the definition of diagnostic categories, organizational impediments, and occupational disabilities was refined.

### Step 3: data collection

Different means of data collection can be used, but it should be noted that the results are often not comparable. Gerbert and Hargreaves concluded from their study that physician and patient interviews, chart audit, video taping and audio taping capture different aspects of care. Agreement between raters was comparable and high for all methods.12 31 61 The advantages of using medical records are low costs and high acceptability for physicians.31 42 However, medical records contain no information about the interpersonal process, and also are often found to be incomplete especially for advisory contents.27 31 61 The disadvantage of observation is the presence of an observer, which can lead to distortions in compliance with the guidelines.

We gathered the information with forms made for this specific purpose because we were not sure that the medical record would contain sufficient information. The forms were explained during the first meeting and tested by the occupational physicians on their patients between the first and second meeting. For each contact a form had to be completed. The forms from every episode of low back pain were scored on the criteria.

### Step 4: assessment of performance

A choice has to be made for a scoring method assessing the performance of the whole process or performance for specific indicators. Performance of occupational health services and of individual physicians might also be of interest. A simple, unweighted summed score of performance has advantages, but assumes that the different indicators throughout the whole process are equally important.64 Another consideration is that performance scores for different indicators give more detailed information about specific aspects of care.16 58 For each case, we scored performance on the specific indicators with a binary outcome (deviant or not) for the whole episode, independent of the number of consultations. Encouragement of activity and continuity of care could also have a third outcome, inapplicable. Performance for an indicator was deviant from the guidelines if one or more of the criteria were not met for that indicator. In case the criteria could not be applied due to forms that were not fully completed, an indicator was also scored as deviant, unless the right answer could be retrieved otherwise.

A performance rate for an indicator was computed as the number of deviant cases for that specific indicator divided by the total number of applicable cases. A case performance score was defined as deviant if one or more indicators for that case were deviant. The performance rates for the specific occupational health services were calculated as well, but we did not find relevant differences. Due to small numbers we did not calculate performance rates for individual physicians.

### Table 2. Performance scores and rates for occupational rehabilitation of 69 workers with low back pain

<table>
<thead>
<tr>
<th>Outcome variables</th>
<th>Performance rate of cases (n (%))</th>
<th>Performance score of cases (n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encouragement of activity*</td>
<td>Not deviant 8 (14.3) deviant 48 (85.7)</td>
<td></td>
</tr>
<tr>
<td>Psychosocial problems</td>
<td>Not deviant 6 (8.7) deviant 63 (91.3)</td>
<td></td>
</tr>
<tr>
<td>Inadequate treatment</td>
<td>Not deviant 12 (17.4) deviant 57 (82.6)</td>
<td></td>
</tr>
<tr>
<td>Occupational disabilities</td>
<td>Not deviant 1 (1.4) deviant 68 (98.6)</td>
<td></td>
</tr>
<tr>
<td>Organisational impediments</td>
<td>Not deviant 3 (4.3) deviant 66 (95.7)</td>
<td></td>
</tr>
<tr>
<td>Advice for return to work</td>
<td>Not deviant 2 (2.9) deviant 67 (97.1)</td>
<td></td>
</tr>
<tr>
<td>Continuity of care**</td>
<td>Not deviant 28 (46.7) deviant 32 (53.3)</td>
<td>Performance score of cases (n (%))</td>
</tr>
<tr>
<td>Technical competence score (1–6)</td>
<td>44 (63.8) 25 (36.2)</td>
<td></td>
</tr>
<tr>
<td>Case performance score (1–7)</td>
<td>26 (37.7) 43 (62.3)</td>
<td></td>
</tr>
</tbody>
</table>

*13 Cases not applicable.  
**9 Cases not applicable.
This is mainly due to continuity of care. Therefore, we also assessed the technical competence separately by summing the other six indicators. Thirty six per cent of the cases were scored as deviant on technical competence. Especially encouragement of activity and inadequate treatment contributed to a deviant score for a specific case.

**EVALUATION**

The validity of indicators and criteria was assessed by asking the opinion of the participating occupational physicians on their relevance and appropriateness. All indicators were valued as relevant by most of the physicians (table 3). Least valued were the relevance of inadequate treatment, involving communication with treating physicians, and continuity of care, a follow up every third week with a few defined exceptions. The formulation of the criteria of these two indicators and the criteria for encouragement of activity, was assessed as too rigid by almost half of the occupational physicians. This was based on practical reasons and reasons of assumed effectiveness. For instance, a follow up every three weeks (continuity of care) was valued as desirable, but not always possible due to holidays, lack of time, a longer duration of a treatment plan of a physiotherapist, or few opportunities at the work site. For communication with treating physicians (inadequate treatment), the occupational physicians wanted a more facultative definition, dependent on assumed effectiveness. The encouragement of activity should, according to the disagreeing physicians, be dependent on the patient’s own initiative. According to most of the physicians, the criteria were correctly formulated for the other four indicators.

For the quality of data collection, the registration forms were accepted by and understandable for the occupational physicians. However, it turned out from the questionnaires that the registration forms of eight cases were not fully completed, meaning that the validity was not optimal. We assessed agreement between raters by comparing the assessment of the seven indicators of 20 cases by two raters. The $\kappa$ statistic (0.88) was satisfactory.

Case performance scores were returned to the occupational physicians as feedback for their own performance. The $\kappa$ statistic for agreement between our and their rating was not high (0.46). They disagreed in 66% of cases with their own deviant performance score.

**Discussion**

We present a method for the evaluation of the activities of occupational physicians. The method is objective with explicit criteria for quality assessment and the agreement between raters was good. This explicit method is an alternative for a medical audit as described by Agius et al.8–10 They presented no guidelines, explicit indicators, or criteria. So, it is difficult to discuss the method they used and to compare the results of their studies with others. A simple method we used ourselves in an audit of occupational health care for employees with low back pain is also not comparable, because of uniform criteria for each patient.65 Only the occupational health care study of Law et al resembles our method.12 They used criteria mapping to assess occupational therapists’ competencies for self care problems of their patients. The basis for their map was formed by assessment of activities of daily living, among which one item referred to the vocational situation. However, they did not specify what kind of interventions could be considered in an intervention plan.

We evaluated different aspects of our quality assessment method, but we did not assess the validity of the guidelines. This validity is best guaranteed if the different elements are evidence-based. However, the evidence about effective interventions was small: only encouragement of activity for patients with non-specific low back pain had been proved to be effective.8–10 66 67 The other interventions were reached by consensus. These interventions could not always be defined as specific as would be desirable to guarantee professional autonomy. On the other hand, this autonomy should enhance the acceptability. Also a consensus process improves the implementation and thereby the acceptability.22 We did evaluate the acceptability implicitly by getting everyone’s agreement after the first two meetings. We tried to improve the change in practice by combining different strategies to maximise compliance. We used almost all methods advocated in the scientific literature.8–10 Therefore, in theory the implementation step was optimal.

We assessed the content validity by inquiring about relevance and formulation as perceived by the occupational physicians. Interestingly,
they objected to the criteria for three performance indicators that they agreed on at the construction of the guidelines. These objections were directed at the indicators with the lowest scores. So, in practice, the physicians acted less strictly than formulated in the guidelines. Sometimes, they considered it to be difficult or even inappropriate to act in accordance with the ideal guidelines. However, we dispute their comments. The criteria for encouragement of activity were formulated to encourage adjusted work at home or at work, so it was less strictly defined as interpreted by the physicians. Furthermore, this is the only evidence-based intervention, so we think that it remains important to stimulate patients, even when they are already active. Good performance on continuity of care can be hampered by internal problems—such as lack of time—or external ones—such as high workload or prolonged treatment plans of the physiotherapist—but these are exactly the factors that can prolong the duration of sick leave. It is important to prevent these internal, organisational problems and to intervene in the case of external problems. The assessment of inadequate treatment as too rigid was based on assumed effectiveness. We think that it is practically impossible to define criteria for exceptional cases, for which communication with treating physicians is useless. Furthermore, it can never be assessed whether communication with treating physicians is really ineffective unless this communication is obligatory when treatments hinder the return of patients to work.

These difficulties with action related to curative treatment correspond to a recent publication about problems in the cooperation between the two parties.68 In that study it was shown that the main bottleneck for cooperation was the unclear position of the occupational physician. However, the necessity for improvements was recognised by both parties and therefore it is important to show the effectiveness of this cooperation.

Briefly, in practice it will be both difficult and important to strive for optimal care as defined by the criteria. Although occupational physicians considered that the criteria were too rigid, we tend to think that the list of interventions that were defined as appropriate was too long and performance deviated only slowly. More specific guidelines and criteria are needed for a more sensitive quality assessment, but this is very complicated. A lot of unknown factors influence the process, and the specific criteria should incorporate those factors. A step forward in the process of developing more specific guidelines would be an appropriate assessment of psychosocial and disability problems. In a new version of the guidelines, we included questionnaires for both assessments.47 69–71 The usefulness and effectiveness of these questionnaires should be evaluated in further research.

The comments of the occupational physicians on their own performance scores were mainly related to the difference between general criteria and specific individual circumstances which can explain deviations. They did not agree with the deviant performance scores in two thirds of the cases. It is difficult to judge whether they were right in their own cases. It might be related to the actor-observer phenomenon: actors tend to ascribe their behaviour more to circumstances than observers do.72 To unravel this problem, the performance scores can be related to other indicators of the quality of occupational rehabilitation. Other indicators are, for example, realisation of planned adjustments at the worksite and satisfaction of the employee. This relation with other process and outcome measurements refers to the strategic or useful quality of our method. Until now, too little attention has been paid to this quality aspect. In further analyses this question about relation with other quality indicators will be considered.

The performance indicators assess the technical quality and continuity of care, which cover only a small part of quality. Communication and interpersonal manner were not assessed. This restriction was practically based, as observational techniques were not applicable for our trial, but it is also difficult to define criteria for appropriate communication. It should be investigated whether satisfaction questionnaires for employees can validly assess these aspects.

We did not use medical records because we expected incomplete recording.65 Instead, we chose specific registration forms. These forms were at the same time a manual of the guidelines. It seemed that the most objective criterion, the follow up, was the most deviant one. Possibly the performance scores we assessed were too optimistic. On the other hand, the physicians did not know how to fill in the forms for a good performance score, because they were not informed about the evaluation method. It also seemed that in a few cases the registration forms were not adequately completed. It would have been more appropriate to use a method of quality assessment with a broader base, and time could have been saved by the use of computerised registrations and evaluations.

Conclusion

We used this method to assess compliance of physicians in a randomised clinical trial, but it is also applicable for evaluation purposes in such situations as medical audit. By contrast with our method, some quality systems—such as vocational training and certification of occupational health services or professionals—cannot detect gaps in performance. We think that these gaps should be the base for further discussion on separating real deviant performance from those cases with acceptable reasons for its deviance. Our method can be used both for a single evaluation and for a continuing process of improving occupational health care. We conclude that assessing the quality of the work of occupational physicians is evolving and therefore is not without problems. A challenge for the future remains. If a method of quality assessment were to be applied in practice, it would be more appropriate and time saving to use computerised registrations and evaluations.
Vancouver style

All manuscripts submitted to *Occup Environ Med* should conform to the uniform requirements for manuscripts submitted to biomedical journals (known as the Vancouver style.)

*Occup Environ Med*, together with many other international biomedical journals, has agreed to accept articles prepared in accordance with the Vancouver style. The style (described in full in the *JAMA* [1]) is intended to standardise requirements for authors, and is the same as in this issue.

References should be numbered consecutively in the order in which they are first mentioned in the text by Arabic numerals on the line in square brackets on each occasion the reference is cited (Manson[1] confirmed other reports[2][3][4][5]). In future references to papers submitted to *Occup Environ Med* should include: the names of all authors if there are three or less or, if there are more, the first three followed by et al; the title of journal articles or book chapters; the titles of journals abbreviated according to the style of *Index Medicus*; and the first and final page numbers of the article or chapter. Titles not in *Index Medicus* should be given in full.

Examples of common forms of references are: