

ORIGINAL REPORT

## PROBLEMS COMPLETING QUESTIONNAIRES ON HEALTH STATUS IN MEDICAL REHABILITATION PATIENTS

Thorsten Meyer PhD, Ruth Deck PhD and Heiner Raspe MD, PhD

*From the University of Lübeck, Institute for Social Medicine, Lübeck, Germany*

**Objective:** The validity of health status questionnaires in patients attending medical rehabilitation services has been questioned. The objectives of this study were to identify problems that patients have in completing different health status questionnaires, and thus identify possible major pitfalls in interpretation of the scores.

**Methods:** The study comprised a consecutive sample of 105 patients scheduled for inpatient rehabilitation who had completed a health status questionnaire prior to admission. They underwent a cognitive interview at admission (response rate 95.5%).

**Results:** Patients were motivated to provide the clinic with a clear-cut picture of their illness and life situation. However, the content and response formats of the questionnaire were not specifically tailored to meet their motivation. For example, time-references predefined in the instructions were not meaningful to patients with variable symptoms. Patients' understanding of response categories was found to be ambiguous. In cases of uncertainty, patients were likely to select the "normal" middle category of response.

**Discussion:** It is important to be aware of the problems that rehabilitation patients have in providing answers to health-related questions, because these problems are likely go unnoticed, since patients tend to provide answers even in cases of uncertainty. Instruments need to be tailored towards the motivational states, needs, cognitive capacities and subjective meanings of the respondents.

**Key words:** validity, self-report questionnaire, cognitive interview.

J Rehabil Med 2007; 39: 633–639

*Correspondence address:* Thorsten Meyer, Universität zu Lübeck, Institut für Sozialmedizin, Beckergrube 43–47, DE-23552 Lübeck, Germany. E-mail: thorsten.meyer@uk-sh.de

Submitted May 22, 2006; accepted March 23, 2007

### INTRODUCTION

Self-report questionnaires of patients have become an integral part of healthcare (1), and of rehabilitation and rehabilitation research (2). However, acceptance of self-reported health status measures in rehabilitation cannot be taken for granted. Questionnaires may score high on psychometric criteria, such as objectivity and reliability, while at the same time allowing for over- or under-reporting of symptoms (3) and may place high demands on the cognitive abilities and motivational states

of the respondents (4). Despite the proliferation of health status questionnaires in rehabilitation, we were confronted with clinicians who were reluctant to utilize them and who questioned their validity, as has been reported in other clinical domains (5).

Validity has several important aspects (6). Different approaches have been used to appraise content and substantive validity aspects of an item that are constitutive to validity judgements. In particular, cognitive interview techniques (7) have been applied, including concurrent or retrospective think-aloud, and cognitive probes, such as special comprehension, information retrieval and general probing (8). Different cognitive models of the response process have been developed (e.g. 4, 9–11).

Four steps are regarded as necessary in order to elicit a reasonable answer to a question (10): (i) the respondent has to interpret the question and understand its intended meaning; (ii) he or she has to search his or her memory for relevant information; (iii) he or she has to integrate all the information retrieved into a single judgement; and (iv) translate this judgement into a response by selecting the most appropriate response category.

The extent to which a person is inclined to answer the questions faithfully and assiduously is called "optimizing" (10). "Satisficing" denotes superficial responses. A person may spend less energy on answering the questions or may mix up the information retrieval and judgement processes. In "satisficing" the selection of responses may be oriented toward external stimuli. It is facilitated by: (i) difficult items, (ii) low skills in answering the items; and (iii) low motivation to respond assiduously or strong concurrent motives (10). As in everyday conversation, respondents interpret the meaning of a question both in terms of its literal and pragmatic meaning (11). These are open to different interpretations, especially if the context of the questions is ambiguous. Therefore, the layout, response format (12), or the communicated intention of the author might affect content or substantive validity aspects.

A number of studies have used cognitive interview techniques to gain insight into the thoughts, reasoning and meanings that people use when completing health-related questionnaires (e.g. 13–15). Others have applied in-depth interviews or analysed the comments people make (13, 16, 17). In rehabilitation patients it could be shown that credibility or self-disclosure problems were not prominent (15). The presence of others while completing a questionnaire has

been found to be common in rehabilitation patients (15). However, indications for both positive (correcting improper responses) and negative effects (social desirability) have been reported (15).

In medical patients, symptom variability was found to be a central problem in answering symptom-related questions, e.g. in pain items relating to a specified time period (13, 16). Interviewees tended to provide the interviewer with additional information on standardized responses to put their answer into context, especially when the respondents felt that the questions did not adequately match their situation (13). Sparse or equivocal context information led to different interpretations of the meanings of items (17). However, a tendency to answer questions despite being unsure about their meaning could be found (16). Of special relevance to the rehabilitation field were accounts of how patients dealt with co-morbidity in rating their health status or specifying symptoms (14, 16), how they select an appropriate comparison standard for judging one's health status (13, 14, 16, 17) or how underlying scales or concepts were changed over the course of time ("response shift") (14, 16).

To our knowledge, these problems described for medical patients have not been investigated in rehabilitation patients. Rehabilitation patients are different from acute or general care medical patients, e.g. regarding treatment goals (role functioning vs cure), a preponderance of functional symptoms or disabilities, and a tension between individual and institutional treatment goals.

The aim of the present study was to identify possible major pitfalls in the interpretation of health status questionnaire scores in rehabilitation patients. In particular, the study aimed to identify the problems that the patients have in completing different health status questionnaires that have a strong potential to distort the answers systematically and are related to formal and content-related weaknesses in the questionnaires.

## METHODS

Patients scheduled for inpatient rehabilitation underwent a guided open-ended interview, including cognitive interview techniques. The subjects comprised a consecutive sample of all patients admitted to 2 rehabilitation clinics in the Northern German state of Schleswig-Holstein. Primary indications for admission to the clinics were musculoskeletal, cardiac and pulmonary disorders. Any patient undergoing planned admission to these clinics is provided with a questionnaire battery, including questions about diverse health and functional dimensions. An overview of the questionnaires is given in Table I. The selection of questionnaires was not within the remit of this study, but it was based on the clinics' information needs. It comprised both internationally known and validated instruments, as well as questions set up by the clinics themselves. All patients who had received the questionnaire battery prior to admission were included. Direct admissions from acute clinics or admissions within days after an acute hospital stay ("Anschluss-Rehabilitation" in Germany) were therefore not included in the sample. All patients were contacted within the first 3 days of their stay and were asked to participate in the study with the stated aim of improving the questionnaire the clinic had sent them. All patients were asked to give written informed consent to participate. Due to special organizational requirements in conducting the interview, i.e. a maximum time-slot of 60 min, without the possibility of re-interviewing patients if deemed necessary, and the rural locations of the clinics, we decided to interview approximately 50 patients in each clinic. We expected this strategy to be overly cautious with regard to emerging problems, but it allowed for different problems and topics being reconciled in subsequent interviews. Also, it had the positive effect of a better acceptance of the results among clinicians unfamiliar with qualitative sampling approaches. The patients were made aware that the interviewer was not a member of the clinic's staff, but of an independent university research institute. It was emphasized to all patients that the interview and the results of the study were independent of their individual treatment in the clinic, so as to foster self-disclosure about personal topics in the interview. At the time the interviews were conducted, the interviewer (TM) was new to the field of rehabilitation research. Having participated in other questionnaire development projects in psychiatric research (e.g. 30, 31) the first author had both an appreciation of the potential of questionnaire methodology as well as experience of difficulties that mental patients have in completing symptom or quality of life questionnaires.

Table I. Questionnaires that were part of the questionnaire batteries in clinics 1 and 2.

Domain	Clinic	Questionnaire
General health status	1	SF-36 general health item (18)
	2	Numerical rating scale (NRS)
Pain	1	Pain-subscale of the Nottingham Health Profile (NHP) (19); Pain Related Self Statements Scale (German version) (20)
	2	Pain localization (pain mannequin); pain intensity (NRS); single item questions on pain-duration, -sensation, -presence during the day, onset of pain, changes, successful means of pain reduction
Somatization	1	Somatization subscale of the Symptom Checklist (SCL90-R) (21)
	1	Energy level, sleep (subscales of the NHP) (19)
Depression, anxiety	1	Center for Epidemiologic Studies Depression Scale (CES-D; German version) (22, 23)
	2	Hamilton Anxiety and Depression Scale (HADS-D) (24)
Functional capacity	1 & 2	Hannover Functional Status Questionnaire (Funktionsfragebogen Hannover, FFbH-R) (25); English version (26)
Drug consumption	1	Questions on alcohol and nicotine consumption
	2	Fagerstrom Test for Nicotine Dependence (27)
Other quality of life aspects	2	Questionnaire on general wellbeing (from FEG) (28)
	2	Questions on disabilities due to pulmonary obstructive disorders
Motivation	2	Expectations towards rehabilitation (FREM-17) (29)

Qualitative methodology has been chosen because an approach was needed that was able to identify problems not thought of a priori. As reported, studies with similar goals have successfully applied qualitative approaches. The cognitive interview approach appeared to be suitable to identify both content and substantive aspects of validity (6), at the same time allowing for inclusion of a sufficient number of subjects. The cognitive interview addressed problems that patients explicitly stated they had in handling the questionnaire or problems that might be inferred from patients' responses. It combined techniques of special comprehension probing (persons were asked to elaborate on specific terms or aspects, e.g. "What do you mean by health?"), information retrieval probing (persons were asked to recall the process of informational retrieval, e.g. "How did you come up with this answer?") and general probing (e.g. "Did you run into problems answering this part of the questionnaire?"). The patients' answers were recorded with pencil and paper during the interview. If necessary, additional notes were taken subsequent to the interview.

Analysis of the guided interview was carried out using a qualitative content analysis applying an outline approach with inductive category development according to Mayring (32, 33). Codes were developed for interviewees' recorded statements, comprising the meaning of the statements. The codes were refined with the inclusion of additional interviews to represent interviewees' responses adequately. To declare a problem as summarized in the codes a substantial threat to questionnaire validity, we thought through the possible consequences of the reported problem and reasoned whether they would be likely to result in a biased response. In addition, the problem should not be restricted to the statements of a single patient. Due to our methodological approach (patients did not complete the questionnaire immediately before or during the cognitive interview) our focus was on content-related issues. Formal weaknesses of the questionnaires were reported as they emerged during the interview. However, they were mixed with content-related problems, as the example of problems with response formats will show. Due to limited resources the processes of developing codes and coding were not performed by a qualitative research team, but by the first author in close consultation with the other authors.

## RESULTS

A total of 105 out of 110 eligible patients participated in this study (participation rate 95.5%). The characteristics of the sample are summarized in Table II. The participants were primarily blue-collar workers with a basic level of education.

### General results

The patients were motivated to provide the clinic with a clear-cut picture of their illness and life-situation by means of the questionnaire. This motivational state became apparent in different ways. There was a broad willingness and interest in highlighting problems the patients had in completing the questionnaire. Only 6 patients reported that it took less than 15 minutes to work through the questionnaire, which might be a strong indication of satisficing. The assumption of high motivation among the patients to present their health status provides a plausible explanation of the anger expressed by patients who were unable to understand questions or response options, or who judged them to be inappropriate for their life situation. Anger appeared to be pronounced if the patients realized that the doctors did not take notice of their responses in the questionnaire. Anticipation of possible disinterest by the professionals with regard to the patients' questionnaire responses could result in "satisficing", as the following quote

exemplifies: "...did not think a long time about it, I thought nobody was going to read it anyway."

### Time-reference

A substantial problem emerged from the time-frame predefined in the instruction or the items themselves. For example, the somatization subscale of the symptom checklist (SCL90-R) asks for different somatic complaints within the last 7 days. "Today I have a headache. Usually I don't have a headache. Should I report this here?" It is evident that this patient did not just follow the instruction, but took the self-perceived purpose of the questionnaire survey into consideration, i.e. providing a proper picture of his or her disease. Therefore, this patient might not have ticked this item because he did not relate his headache to his illness that made him apply for rehabilitation services. This dilemma becomes even more pronounced in patients who did not experience those symptoms within the last 7 days that at other times were prominent in the course of their disorder: "I had to think a lot about this; I did have this for months but not within the last 7 days". This dilemma – providing an adequate impression of their symptoms and disorder which made them apply for rehabilitation services vs taking the instruction literally – was not resolved in a uniform way. Some ignored the 7 day time-reference: "it could come back any day"; "I did complete it in a more general sense"; "if I had headaches during the last 3 months but not within the last

Table II. Characteristics of the patient sample.

Characteristic	Total n=105	Clinic 1 n = 52	Clinic 2 n = 53
Age, years, mean (SD)	50 (9.7)	51 (10.0)	49 (9.5)
Gender, male %	62	54	70
Family status, %			
Married	68	79	58
Single	13	8	19
Divorced/separated	13	8	19
Widowed	5	6	4
School examinations, %			
Secondary school <sup>a</sup>	58	58	59
Secondary modern school <sup>b</sup>	22	23	21
University entrance degree	11	17	6
Other school examination	4	0	8
No school examination	1	2	0
Educational status, %			
Apprenticeship	62	62	64
Vocational school	9	9	8
University	8	8	8
No respective education/training	19	19	19
Diagnostic groups (multiple diagnosis possible), %			
Musculoskeletal system and connective tissue	71	60	81
Endocrine, nutritional and metabolic diseases	56	64	49
Diseases of the circulatory system	46	52	40
Mental and behavioural disorders	28	15	40
Diseases of the respiratory system	11	0	23

<sup>a</sup>"Hauptschule"; <sup>b</sup>"Realschule".

SD: standard deviation.

7 days I would provide you with a wrong picture of myself”; “it is important to know: do you have this symptom more often?”. Less prevalent, by contrast, were patients who reported that they did not overrule the instruction, i.e. they referred to the 7 day reference period while feeling uncomfortable about not presenting an adequate picture of their usual status.

*Variable symptoms of the disorder*

There is a close correspondence with problems of variable states of disorders, which held especially true for patients with musculoskeletal disorders. For example, this problem emerged in the questionnaire about functional capacity. “Sometimes I can do everything really well ... while at other times I cannot do any of these things”. The majority of the patients did complete the questionnaire as if they suffered from back pain the way they usually did or when it was worst. The problem of variable states was especially important for the pain and somatic complaints items: “It’s difficult to judge because it’s not the same every day; sometimes I do have a headache, sometimes I do not: I did tick ‘a little bit’ in this case, for ‘quite a bit’ I would have to have headaches regularly”; “if I had severe headaches within the last 10 minutes I would have ticked number 4 (extremely), but this would not be true; I would like to refer to a different time-frame; for example, I do have problems with ... frequently”.

Variable pain states made it difficult for patients to specify average values. Six out of 50 interviewees referred to difficulties in averaging pain intensities because of variable pain symptoms: “Since my pain is present at one time, at other times it is not as bad, it is difficult to specify an average value”; “I had to think about this question, to average is difficult, some days 20, at other days I am located at 90” (numbers refer to response categories of a numerical scale ranging from 0 to 100 with steps of 10). Two of the respondents made a virtue of necessity; they did not report a single value, but a range from a minimum to a maximum value.

*Response formats*

Patients preferred a response format in which every response category was labelled with a verbal statement. For example, a 6-point scale response format in a questionnaire about catastrophizing thoughts, where only the extreme values were labelled verbally, provoked uncertainty in the choice of responses. The same was true for a hybrid response format (numerical and visual analogue scale) to judge health in general, ranging from 0 to 100 in steps of 10, with labels of “very bad” (0), “moderate” (50) and “very good” (100).

The most favourable judgement with regard to the response format was given for a pain mannequin for the assessment of pain localization (“the only thing that was really good in this questionnaire”). In this item there was a close correspondence between the experience of the person and the assessment method. A problem related to symptom measurement in general is an unclear distinction between the pain experience and its psychological appraisal: “severity (of pain) is one thing, suffering from pain something else”. For this person 2 different dimensions were confounded in the pain items.

An example (translated for this report) of an uninformative response format is given in Fig. 1. Consumption of alcoholic beverages and smoking behaviours were subsumed under the heading “habits”, as can be seen in Fig. 1. We did ask about alcohol and cigarette use during the guided open-ended interview, so as to compare the results of both the questionnaire and interview (without claiming the interview results to be the gold standard, especially because of social desirability bias). First, we refrained from using the term “alcohol” after a number of interviews, because for a couple of interviewees this term was restricted to spirits, and excluded beer and wine.

The verbal labels for the response categories referring to the amount of beverage consumption appeared to be equivocal. What do the different response options mean to the individual patient? Fig. 2 shows that only a few patients referred to the “more often” category.

How much does a person have to drink, in order to cross the threshold from “sometimes” to “more often”? In order to measure this we asked the patients the following question: “How often do you have to consume alcoholic beverages on average in order to qualify for a “more often” response?” The responses varied markedly, as can be seen in Table III.

habits			
How often do you consume the following beverages on average?			
spirituous beverages	[1] never	[2] sometimes	[3] more often
wine	[1] never	[2] sometimes	[3] more often
beer	[1] never	[2] sometimes	[3] more often
no alcohol since	_____		
withdrawal treatment	[1] no	[2] yes – when? _____	

Fig 1. Example of an uninformative item for the assessment of alcohol consumption (translation and reconstruction of the item)

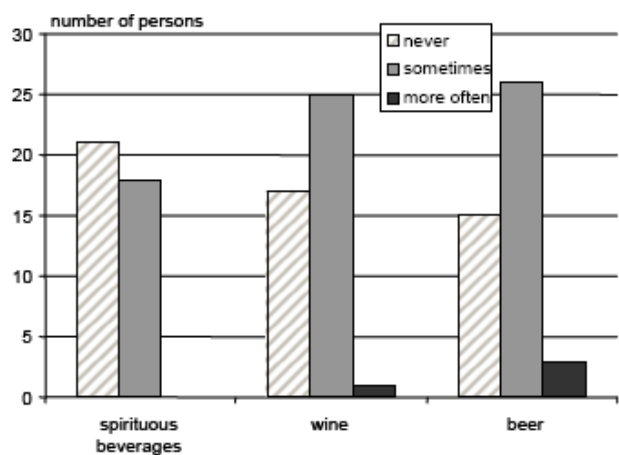


Fig 2. Consumption of alcoholic beverages as has been reported in the questionnaire prior to admission.

<sup>1</sup>“Sometimes” is a translation of the German word “manchmal”; “more often” is a translation of “öfter”.

It became clear that “sometimes” was regarded as the “normal” category, while “more often” (the most extreme category) stood for being somewhat deviant. In those patients who drank alcoholic beverages on a regular basis, there were quite a few for whom “more often” was just a bit more than what they consumed themselves. Although “never” and “more often” comprise some useful information, the middle category “sometimes” had been ticked across the complete spectrum of alcohol consumption, which renders this response format largely uninformative.

#### *Responses in cases of uncertainty*

Only a few patients refrained from completing certain items if they felt they were not able to provide an answer. People repeatedly referred to the middle category as a way to cope with this situation, sometimes even phrased as a general rule: “...didn’t think much about it. Take the middle and you don’t make a mistake.” In one questionnaire uncertainty emerged at the point of change from “objective” questions (e.g. prior diagnosis) to more “subjective” questions (e.g. degree of “suffering” from different symptoms).

#### *Further problems*

Some further themes emerged that might not be regarded as severe sources of distortion, but that still have the potential to bias results, e.g. symptoms in a depression/anxiety scale that might be related explicitly to back pain symptoms (e.g. Hamilton Anxiety and Depression Scale (HADS), cf. Table I: “I can sit at ease and feel relaxed”). In a questionnaire developed to assess activity limitations in back pain patients (Hannover Functional Status Questionnaire (FFbH-R), cf. Table I) items were identified that have considerable overlap with activity limitations due to other diseases (e.g. being able to run 100 metres in order to catch a bus in a patient with an additional obstructive pulmonary disorder). Having been asked about symptoms of diseases, some patients wondered whether a subsequent question on health status in general should include their reported disease status or be thought of as related only to other health problems.

Table III. *How often do you have to consume alcoholic beverages, on average, in order to qualify for a “more often” response? (coded responses from the cognitive interview).*

	Frequency	Valid (%)
In everyday life, i.e. not occasionally, but less than once per week	1	2
In everyday life, at least once per week, not quite (almost) everyday	10	20
In everyday life, almost everyday	9	18
Every day	18	36
2–3 glasses of wine/beer per day	2	4
5–6 glasses of wine/beer per day	2	4
Up to 8 glasses of wine/beer per day	1	2
In the morning/at work	3	6
Alcoholic/if you need alcohol	2	4
Getting drunk (every week or every month)	2	4
Missing	1	
Total	51	

## DISCUSSION

This study identified different problems that have the potential to pose major threats to content or substantive aspects of validity of standardized questionnaire data in rehabilitation patients. We do not think there are substantive reasons to limit the interpretation of our results to the specific questionnaires used in this study, as they represent a wider spectrum of different health-related questionnaires, ranging from internationally well-known instruments to self-developed questionnaires. Most of the problems reported could be overcome if the instruments themselves were developed carefully and tailored specifically to the motivational states, needs and cognitive capacities of the respondents. This is of special importance, as even highly valued and widespread questionnaires, e.g. the Short-Form 36 (16, 17, 34), have been shown to have substantial drawbacks in this respect. There are excellent accounts of what constitutes a good questionnaire item (35). However, in the development of standardized health status questionnaires the issue of meaning is handled as a minor issue and the testing process is almost exclusively a quantitative activity of producing psychometric indices (16).

The person’s motivation to present an adequate impression of his or her personal state or situation has been emphasized in survey research (13, 17, 35). We found this aspect to be central in the rehabilitation patients prior to admission. How could the persons’ motivation to present an adequate impression be integrated into questionnaire design? For example, if we ask patients about their pain intensity we could first ask about situations in which the pain was worst. Some people were eager to report their worst pain experience and, despite being asked about *average* pain intensity, they still reported their worst state, just to give an impression of their suffering. Afterwards questions on average pain intensity or even about possible pain-free intervals could be asked without a possible bias due to self-presentation motives (36).

In line with results from other medical settings (13, 16) problems of inconsistent time-references and alterations of symptoms were also identified in our study. These problems might be dealt with using a similar approach to that suggested above. First, the patients should be given the opportunity to present the situation that made them apply for rehabilitation services, e.g. a typical situation, regardless of their symptoms status “within the last 7 days”. Secondly, patients should be willing to answer the same questions with respect to the last 7 days, which should yield more comparable results.

We found indications of biased results in questions on general health status in terms of under-reporting, as has been noted previously (14, 16). There should be simple strategies to overcome this possible bias: either to put the question about general health status at the beginning of the questionnaire (battery), or to make the intended meaning clear in the item instructions.

These results should encourage us to develop and evaluate response formats that correspond as closely as possible to the respondents’ perspective. The case of assessment of alcohol consumption exemplifies the possible value of the cognitive

interview technique and the thoughtful phrasing of response alternatives. While an 11-point response scale appears to be advantageous in the light of a high level of variance (i.e. information), we must ensure that the response categories have a *meaning* to the respondents that is also *consistent* across the respondents. With an intellectually heterogeneous sample, the reference group for what constitutes an adequate response format should be selected from the lower part of the distribution. Otherwise we run the risk of producing invalid results for a substantial part of our population of interest.

The possible threat to content and substantive aspects of validity, especially in patients of lower intellectual capacity, notably holds true in light of the result that in cases of uncertainty a number of persons (not understanding the question or response format) still chose a “valid” response category, i.e. the middle category. The inclusion of opt out response categories (“don’t know”, “doesn’t apply to me”, “no opinion”, etc.) might reduce this potential bias, providing respondents with the opportunity to refrain from a substantial judgement. However, there is no evidence as to how “opt out” response options could be of help. It is known that “opt out” options lead to a higher percentage of item non-response, although there is a debate on the degree to which this has to be regarded as valid opt-outs (37). Provision of these opt-outs might lead to a more superficial response process (corresponding to “weak satisficing”) compared with the situation in which the respondent is expected to provide a substantial answer (38).

Limitations of study design and conduct have to be taken into consideration. The results of this study are based on the patients’ verbal reports. Cognitive interviews have become a popular feature in survey research. However, data on their evaluation, i.e. their efficacy, is sparse (7, 39). We know that respondents have limited access to the process that determines the responses in a questionnaire (40). In a discussion on the quality of verbal reports in the light of cognitive interviewing, Conrad & Blair (40, p. 69) stated: “Taken together, verbal reports are fragile sources of data, sometimes valid but sometimes not, sometimes independent of the process being reported but sometimes not”. However, verbal data are central in cognitive interview research, as they are considered to be the best source in eliciting meaning issues and providing indications of cognitive processes.

In addition, this has been a pure observational study. Questionnaires were not selected specifically for this study, but were the ones utilized by the different clinics. A more rigorous approach could have made use of systematic variation of questionnaires, e.g. by comparing instruments that were developed by means of patient involvement (focus groups, cognitive pre-tests; 23) vs without patient involvement, or well-validated vs not validated instruments in rehabilitation patients.

Interviews were conducted from a cognitive interviewing perspective and, to a lesser degree, from a traditional qualitative research perspective. The former allowed for analysing a large number of interviews, which is unusual in traditional qualitative research, which is often characterized by in-depth explorations of only a few subjects. Cognitive interviewing in the context of pre-testing questionnaires is applied to identify major problems in different aspects of instrument application, i.e. it developed in

the field of applied quantitative social sciences. Within this field it is common to take written notes and to have the data analysed by a single researcher (7). However, this is in contrast to quality criteria in traditional qualitative research, e.g. integration of multiple perspectives in data analysis, tape-recorded documentation and transcription of interviews, which were not possible in this study. We therefore view our results as a documentation of problems found, not as a true and comprehensive representation of possible pitfalls in questionnaire handling of rehabilitation patients. We should not try to interpret the frequency of problems found as a measure of problem seriousness (8); weighing of problems would also over-interpret our results.

The results of this study shed light on research areas we should focus on in order to improve aspects of content and substantive validity of questionnaire responses. There are several ways to analyse the possible impact of the problems identified above, e.g. the differences between different instructions for the same questions as suggested above could be identified. We could systematically control for the possible social influence on questionnaire responses, as we have suggested elsewhere (15). We already know from research on response-shift or the comparison of self-reports at a defined time-point and an assessment of the status remembered some time later, that retrospective evaluations of one’s health state appears to be biased towards elevated symptom severity (41). So far, we have not identified the (psychological) mechanisms that might lead to these phenomena, despite the existence of applicable psychological theories (e.g. 42).

It should be noted that the foci of this manuscript were potential *problems* in content and substantive validity aspects of questionnaires. In fact, we were impressed by the motivation and sincerity of the patients in completing a comprehensive questionnaire battery, as we have reported earlier (15). Problems of deliberate distortions were implicit only in a few patients. Problems of self-disclosure were related mainly to self-reports of psychological functioning, i.e. dependent on the targeted construct. We therefore concluded that patients’ motivational states, needs and cognitive capacities *per se* did not preclude the use of health questionnaires in rehabilitation patients (15). However, in the application of health questionnaires in these patients we should take into account the motivational basis and peculiarities as well as possible cognitive or literacy limitations of the patients. For example, how does the questionnaire take the motive of self-presentation into account? Do the questions and response formats have unequivocal meanings to the patients? How are patients supposed to report symptoms in the case of fluctuating disease states? Is the time-reference chosen meaningful to the patients’ situation?

In order to draw valid conclusions from rehabilitation research based on self-report questionnaires, developers and users of the questionnaires must ensure that the patients understand the instructions, questions and response formats as intended. To determine this, the focus should be on those patients who are expected to have the most difficulties responding, and those who have dissimilar educational and social backgrounds to that of the questionnaire developers. These respondents might provide the most unexpected, and therefore valuable, information.

## ACKNOWLEDGEMENTS

This project was supported financially by the "Verein zur Förderung der Rehabilitationsforschung in Schleswig-Holstein e.V" (vffr project no. 63).

## REFERENCES

- McHorney CA. Health status assessment methods for adults: past accomplishments and future challenges. *Ann Rev Pub Health* 1999; 20: 309–335.
- Haigh R, Tennant A, Biering-Sorensen F, Grimby G, Marincek C, Phillips S, et al. The use of outcome measures in physical medicine and rehabilitation with Europe. *J Rehabil Med* 2001; 33: 273–278.
- Streiner DL, Norman GR, editors. *Health measurement scales, 3rd edn. A practical guide to their development and use.* Oxford: OUP; 2004.
- Tourangeau R, Rips LJ, Rasinski K, editors. *The psychology of survey response.* Cambridge: Cambridge University Press; 2000.
- Garland AF, Kruse M, Aarons GA. Clinicians and outcome measurement: what's the use? *J Behav Health Services Res* 2003; 30: 393–405.
- Messick S. Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol* 1995; 50: 741–749.
- Willis GB. Cognitive interviewing revisited: a useful technology, in theory? In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, et al., editors. *Methods for testing and evaluation survey questionnaires.* Hoboken, NJ: Wiley; 2004, p. 23–43.
- Willis GB, editor. *Cognitive interviewing. A tool for improving questionnaire design.* Thousand Oaks: Sage; 2005.
- Biemer PP, Lyberg LE, editors. *Introduction to survey quality.* Hoboken, NJ: Wiley; 2003.
- Krosnick JA. Survey research. *Ann Rev Psychol* 1999; 50: 537–567.
- Schwarz N. Self-reports: how the questions shape the answers. *Am Psychol* 1999; 54: 93–105.
- Schwarz N, Grayson CE, Knäuper B. Formal features of rating scales and the interpretation of question meaning. *Int J Public Opinion Res* 1998; 10: 177–183.
- Ong B, Hooper H, Jinks C, Dunn K, Croft P. "I suppose that depends on how I was feeling at the time": perspectives on questionnaires measuring quality of life and musculoskeletal pain. *J Health Services Res Policy* 2006; 11: 81–88.
- Paterson C. Seeking the patient's perspective: a qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Qual Life Res* 2004; 13: 871–881.
- Meyer T, Deck R, Raspe H. Gültigkeit von Fragebogenangaben in der Rehabilitationsforschung: Unter welchen Bedingungen füllen Patienten Fragebogen aus? *Rehabilitation* 2006; 45: 118–127.
- Mallinson S. Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Social Sci Med* 2002; 54: 11–21.
- Adamson J, Gooberman-Hill R, Woolhead G, Donovan J. "Quester-views": using questionnaires in qualitative interviews as a method of integrating qualitative and quantitative health services research. *J Health Services Res Policy* 2004; 9: 139–145.
- Bullinger M, Kirchberger I, editors. *SF-36 Fragebogen zum Gesundheitszustand. Handanweisung.* Göttingen: Hogrefe; 1998.
- Hunt SM, McKenna SP, McEwen J, Williams J, Papp E. The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med* 1981; 15: 221–229.
- Flor H, Behle DJ, Birbaumer N. Assessment of pain-related cognitions in chronic pain patients. *Behav Res Therapy* 1993; 31: 63–73.
- Franke GH. *SCL-90-R. Symptom-Checkliste von L. R. Derogatis – Deutsche Version, 2nd edn.* Göttingen: Beltz; 2002.
- Radloff LS. The CES-D scale: a self-report depression-scale for research in the general population. *Appl Psychol Meas* 1: 385–401.
- Hautzinger M, Bailer M, editors. *Allgemeine Depressions Skala (ADS).* Weinheim: Beltz; 1993 (in German).
- Hermann C, Buss U, Snaith RP, editors. *Hospital Anxiety and Depression Scale – German version.* Bern: Huber; 1995 (in German).
- Kohlmann T, Raspe H. Hannover Functional Questionnaire in ambulatory diagnosis of functional disability caused by backache. *Rehabilitation* 1996; 35, I–VIII (in German).
- Westhoff G, Listing J, Zink A. Loss of physical independence in rheumatoid arthritis: interview data from a representative sample of patients in rheumatologic care. *Arthritis Care Res* 2000; 13: 11–22.
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* 1991; 86: 1119–1127.
- Dlugosch GE, Krieger W, editors. *Der Fragebogen zur Erfassung des Gesundheitsverhaltens (FEG).* Frankfurt: Swets Test Gesellschaft; 1995 (in German).
- Deck R, Zimmermann M, Kohlmann T, Raspe H. Rehabilitation-related expectations and motivations in patients with nonspecific backache. The development of a standardized questionnaire. *Rehabilitation* 1998; 37: 140–146 (in German).
- Meyer T, Franz M. QLis – a new schizophrenic-specific quality of life scale. *Eur Psychiatry* 2002; 17 Suppl 1: 184.
- Franz M, Lemke MR, Meyer T, Ulferts J, Puhl P, Snaith RP. Deutsche Version der Snaith-Hamilton-Pleasure-Scale (SHAPS-D): Erfassung von Anhedonie bei schizophrenen und depressiven Patienten. *Fortschr Neurol Psychiatr* 1998; 66: 407–413.
- Mayring P, editor. *Qualitative Inhaltsanalyse. Grundlagen und Techniken, 6th edn.* Weinheim: Beltz; 1997.
- Mayring P. Qualitative content analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (on-line journal)* 2000; 1(2). Available from: <http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm>
- Jenkinson C, Peto V, Coulter A. Making sense of ambiguity: evaluation of internal reliability and face validity of the SF-36 questionnaire in women presenting with menorrhagia. *Quality Health Care* 1996; 5: 9–12.
- Fowler FJ, editor. *Improving survey questions. Design and evaluation.* Applied Social Research Methods Series Vol. 38. Thousand Oaks: Sage; 1995.
- Schlenker BR. Self-presentation. In: Leary MR, Tangney JP, editors. *Handbook of self and identity.* New York: Guilford; 2003, p. 492–518.
- Beatty P, Herrmann D. To answer or not to answer: decision processes related to survey item nonresponse. In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, et al., editors. *Methods for testing and evaluation survey questionnaires.* Hoboken, NJ: Wiley; 2004, p. 71–85.
- Krosnick JA. The causes of no-opinion responses to attitude measures in surveys: they are rarely what they appear to be. In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, et al., editors. *Methods for testing and evaluation survey questionnaires.* Hoboken, NJ: Wiley; 2004, p. 87–100.
- Forsyth B, Rothgeb JM, Willis GB. Does pretesting make a difference? An experimental test. In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, et al., editors. *Methods for testing and evaluation survey questionnaires.* Hoboken, NJ: Wiley; 2004, p. 525–546.
- Conrad FG, Blair J. Data quality in cognitive interviews: the case of verbal reports. In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, et al., editors. *Methods for testing and evaluation survey questionnaires.* Hoboken, NJ: Wiley; 2004, p. 67–87.
- Kohlmann T, Raspe H. Zur Messung patientennahes Erfolgskriterien in der Medizinischen Rehabilitation: Wie gut stimmen "indirekte" und "direkte" Methoden der Veränderungsmessung überein? *Rehabilitation* 1998; 37: S30–S37.
- Ross M. Relation of implicit theories to the construction of personal histories. *Psychol Rev* 1989; 96: 341–357.