Int J Public Health 53 (2008) 96–103
1661-8556/08/020096-8
DOI 10.1007/s00038-008-6097-2

Original Article

The test-retest reliability of a questionnaire on the occurrence and severity of back pain in a German population sample

Susanne Schlademann, Thorsten Meyer, Heiner Raspe

University of Lübeck, Institute of Social Medicine

Submitted: 8 August 2006; Revised: 14 June 2007; Accepted: 5 December 2007

Summary

Objectives: Back pain is considered as one of the most frequent disturbances of health. The available study aimed at the evaluation of the test-retest reliability of fundamental questions on the occurrence and severity of back pain that have been widely used in international population surveys.

Methods: Four hundred and eighty-seven inhabitants of Lübeck/Germany aged 18 to 74 were mailed a first questionnaire on back pain. All respondents were resent the identical questionnaire two weeks after initial mailing. Additionally, they received a questionnaire on selected back pain items covering the time interval between the test and retest survey to identify cases with instable answering patterns due to back pain related changes, exclusively.

Results: One hundred and seventy-nine respondents took part in both the test and retest survey (response rate 36.8 %) and were included in our reliability analyses. Respondents of the first questionnaire were of higher social status than the non-respondents. Dependent on scale level, Cohen´s kappa, quadratic weighted kappa, and intraclass correlation coefficients were computed, respectively.

Conclusions: All evaluated questions reached good up to very good reliability scores.

Keywords: Back pain – Population-based survey – Test-retest reliability – Kappa – Intra-class correlation coefficient.

Introduction

Background

Besides structured interviews, questionnaires are a frequently applied method for the collection of health related data in epidemiological research. Accessibility to a broad population and easy administration concur with the necessity of a careful development and evaluation of psychometric characteristics, as the quality of a questionnaire directly affects the results. Ideally, an instrument gets extensively tested within a pilot study, especially with regards to its reliability, validity, economy and acceptance.

Questionnaires applied in population-based surveys on the occurrence and severity of back pain enable researchers to estimate point-, one year- and lifetime prevalence. In several German regions these surveys were regularly conducted (e. g. Hannover 1989, Bad Säckingen 1991, Lübeck 1991 and 2003) and put into an international framework (e. g. [1]). Although an adequate validity is assumed, the reliability on a single item basis has rarely been challenged. Earlier reports on disease-specific scales for low back pain revealed a mixed picture of reliability: Based on the work of Deyo et al. [2], Mannion et al. [3] examined a core-set of six questions on pain, function, symptom-specific well-being, satisfaction, social disability and work disability related to low back pain. They reported moderate to excellent intraclass correlation coefficients (ICC = .67 to ICC = .95) in patients with low back pain. A comparison of five low back pain disability questionnaires showed test-retest results ranging from poor to excellent with ICC from ICC = .37 to ICC = .84 [4]. The test-retest data on a back pain specific version of the Hannover Functional Ability Questionnaire were highly correlated with r = .91 (retest after three days) and r = .75 (retest after one week), respectively [5; 6]. We present a study on the test-retest evaluation of central items applied to international epidemiological back pain research [1; 7].

Int J Public Health 53 (2008) 96–103

Reliability

Next to objectivity and validity, the reliability of a measurement constitutes a cardinal quality factor of an instrument. Theoretically, reliability means the part of the total variance explained by true variance. Reliability comprises different aspects, e. g. the extent of reproducibility (stability) or consistency. By this means reliability evaluations may include repeated measurements of one subject (test-retest reliability). Inter-item consistency (often described by Cronbach's alpha) implies the extent to which different items are indicators of one and the same dimension.

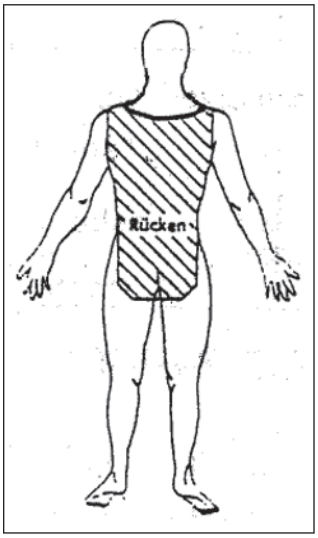

Figure 1 Region of interest (Note: The German word "Rücken" means "back")

Methods

Sample

As in previous surveys on back pain, a systematic sample of German inhabitants of Lübeck aged 18 to 74 was drawn from the population registry of the city of Lübeck (N = 487) [7; 8; 9]. Apart from the address, data on age, sex and residential district were transferred by the registry office. With these data it was possible to determine if the subjects had taken part of an earlier survey on low back pain. The subjects were sent the study questionnaire in August 2004 (test survey). Three weeks after the initial mailing all non-respondents received a postal reminder. Two weeks after a response all participants were sent the identical questionnaire for a second time. An additional questionnaire was sent covering the two week time interval between both surveys (retest survey, without reminder). At the time of the test and retest, all addressed subjects were informed about the methodological focus of the survey and the following retest by a cover letter.

To appraise the social status, registry information and data from the statistical yearbook 2004 of the city of Lübeck were used. First, we ranked the ten residential districts of Lübeck in terms of the proportion of unemployed residents. A second rank order of the ten districts was conducted with regard to the proportion of residents on social welfare. For each residential district, the sum of both ranks was computed. The three lowest sum scores were rated a "high social status", the three highest sum scores were interpreted as indicators for a "low social status", the four sum scores in between implied an "intermediate social status".

Instruments

The items of the study questionnaire have been reapplied in several population-based surveys on back pain since 1984 [7–12]. Due to a lack of a consistent definition of "back", a diagram displaying the region of interest was shown prior to the questions (Fig. 1).

Reliability analysis was conducted using a test-retest design with a time interval of two weeks to prevent carry-over effects [13]. Back pain symptoms could have changed during the period of time between test and retest and by this means influence reliability measurement [14]. For that reason the retest survey included an additional questionnaire covering the time interval between test and retest. It contained exclusively items from the study questionnaire that were sensitive to change over time with regard to the two weeks between test and retest. For example, both test and retest comprised the item "During the past 3 months, did you experience back pain? (yes/no)". Within the retest survey the additional questionnaire collected data on the item "If you think about the period of time since you have completed the first questionnaire: During that time, did you experience back pain? (yes/no)?". This allowed for a substantiated interpretation of potential disagreement of test and retest data due to fluctuating pain symptoms.

Twenty-one items previously used in a back pain survey 2003 were selected [7]. With regard to content these items were assumed to evoke a stable answering pattern. They referred to a period of three months in which the subjects were able to recollect. Data collection comprised the general health status (SF-36; [15]), the occurrence of back pain, bodily pain, pain intensity, functional disability due to back pain, and the back pain related utilization of medical health services during the past three months. Detailed item descriptions are given in Table 2.

Data analysis

Missing values were excluded test by test (pair-wise analysis) from analyses. The occurrence of different types of bodily pain was asked for using a list format. As reported in earlier studies, some subjects affirm only applicable items of a list,

Table 1 Baseline characteristics (data from population registry)

Socio-demographics	sample by registry N = 487	respondents TEST N = 217	respondents RETEST N = 179
age (years) – M (SD)	46.8 (15.7)	48.3 (16.0)	48.9 (15.6)
female – N (%)	258 (53.0 %)	114 (52.5 %)	93 (52.0 %)
social status – N (%)	243 (49.9 %)	121 (55.8 %)	
• high	168 (34.5 %)	70 (32.3 %)	100 (55.9 %)
• moderate	76 (15.6 %)	26 (12.0 %)	58 (32.4 %)
• low			21 (11.7 %)

while items not applicable are disregarded and thus produce missing values. This so-called "checklist misconception effect" is characterized by 1) at least one missing value, 2) at least one valid answer, and 3) the absence of negative answers. In the case of this answering pattern missing values were imputed by the response category indicating absence of symptom [16; 17].

If no back pain had occurred during the past three months, items on back pain symptoms (e. g. pain intensity or back pain related disability) were to be skipped. In case of disregarded skipping instructions values were recoded "not applicable", and they were excluded from analyses.

In the event that uncertainty exists about the equidistance of a scale, it is recommended to complement the calculation of reliability measures for metric data by additional reliability calculations at an ordinal level [18]. If both analyses show similar (at least moderate) reliability, the discussion of scale level is dispensable. In case of extensive deviations of reliability interpretations an evaluation of scale level is advised.

In a first step of reliability analysis, test-retest coefficients were calculated based on all respondents without taking into account the source of disagreement between test and retest (e. g. changes in back pain symptoms).

A second data analysis focused on disagreement between test and retest data on a single item basis: For each item analysis only respondents reporting no change during the period of time between test and retest were included.

Reliability measures

Cohen´s kappa was computed in case of unordered categories [19]. Nominal data in rank order were analysed using quadratic weighted kappa. The interpretation of kappa values follows Altman [20]. Thus, kappa values $\kappa \leq .20$ indicate poor, between $\kappa = .21$ and $\kappa = .40$ fair, $\kappa = .41$ to $\kappa = .60$ moderate, $\kappa = .61$ to .80 good, and $\kappa = .81$ to $\kappa = 1.00$ very good reliability. The agreement of metric data was examined by applying „one-way, random" intraclass correlation coefficients [21; 22]. As recommended by Fleiss [23], ICC ≤ .40 refer to poor, ICC between .41 and .75 to fair to good and ICC > .75 to excellent reliability.

Quadratic weighted kappa values were computed using a free demo version of MedCalc (www.medcalc.be). All other calculations were run using SPSS 14.0.

Results

Non-response analysis, sample characteristics

None of the contacted subjects (N = 487) had participated in previous back pain surveys. All N = 487 persons were included (age M = 48.9; SD = 15.6, range 19–74). The response rate of the test survey was 44.6 % after one postal reminder (N = 217). Respondents (R; N = 217) and non-respondents (NR; N = 270) did not differ significantly in terms of age (R: M = 48.3, SD = 16.1; NR: M = 45.7, SD = 15.4; T = -1.81, p = .070) and sex (R: 52.6 % female; NR: 53.3 % female; Chi-squared = .03; p = .869). The social status indicator seemed to be higher in respondents (R: 55.8 % high, 32.1 % intermediate, 12.1 % low social status; NR: 45.2 % high, 36.4 % intermediate, 18.4 % low social status; Chi-squared = 6.4, p = .041).

One-hundred and seventy-nine persons (36.8 %) completed both the test and the retest survey. The mean number of days between the date of completion of the test and the retest survey was 14.6 (SD = 5.6). Respondents (N = 179) and non-respondents (N = 38) of the retest survey differed in both the average number of disability days (p = .013) and the average impact of back pain on recreational activities (p = .041), which were higher in non-respondents. Taking into account the inflation of type 1 error (Bonferroni-adjustment;   = 0.05; p < .001), these differences cannot be regarded as significant. Table 1 shows the baseline characteristics of the sample and subgroups of respondents. Figure 2 illustrates the participation development.

The study sample was statistically comparable to a sample of a regional back pain survey in 2003, except from "current bodily pain" which was more pronounced in the sample of 2003 [7].

Results of test-retest analyses

In the first step, the test and the retest responses of all N = 179 participants were related without regard to the sources

Int J Public Health 53 (2008) 96–103

Table 2 Test-retest results

Item $_{a, b}$	**TEST (N = 217)**		**RETEST (N = 179)**		κ κ_W **ICC** $_c$
	N	**M, SD or N, % "yes"**	**N**	**M, SD or N, % "yes"**	
In general, would you say your health is very good/good/satisfying/fair/poor?	216	2.52 (.8)	175	2.49 (.79)	ICC = .834; κ_W = .725
Did you experience back pain during the past 3 months? (yes/no)	214	134 (62.6 %)	175	107 (61.1 %)	κ = .853 κ = .961
During the past 3 months, how severe was your back pain on average meaning „0 = no pain" and „10 = worst pain I can imagine"? (scale 0–10)	135	4.2 (1.9)	114	3.7 (1.9)	ICC = .745; (κ_W = .743)
During the past 3 months, how many days did you experience back pain? (about ___ days) • 0 days • 1–7 days • 8–30 days • 31–89 days • 90 days and more	132	 0 (0 %) 41 (31.1 %) 54 (40.9 %) 22 (16.7 %) 15 (11.4 %)	105	 2 (1.9 %) 35 (33.3 %) 41 (39.1 %) 12 (11.4 %) 15 (14.3 %)	κ_W = .794
During the past 3 months, how many days did you cut down on the things you usually do (work, school, household) because of your back problem? (about ___ days) • 0 days • 1–7 days • 8–30 days	127	 88 (69.3 %) 15 (11.8 %) 24 (18.9 %)	101	 70 (69.3 %) 11 (10.9 %) 20 (19.8 %)	κ_W = .698
During the past 3 months, how much did your back problem interfere with your normal activities (dressing, washing, eating, shopping, etc.) meaning "0 = no interference" and "10 = no activities possible"? (scale 0–10)	134	2.6 (2.5)	115	2.4 (2.3)	κ_W = .841
During the past 3 months, how much did your back problem interfere with your recreational activities and social life meaning "0 = no interference" und "10 = no activities possible"? (scale 0–10)	133	2.4 (2.5)	115	2.2 (2.3)	κ_W = .830
During the past 3 months, how much did your back problem interfere with your working capacity meaning "0 = no interference" and "10 = no activities possible"? (scale 0–10)	133	2.3 (2.5)	114	2.2 (2.5)	κ_W = .892
During the past 3 months, did you take in pain reliever or medicine against rheumatism? (yes/no)	132	43 (32.6 %)	115	33 (28.7 %)	κ = .976 κ = .976
During the past 3 months, did you take in pain reliever or medicine against rheumatism because of your back problem? (yes, on ___ days) • 0 days • 1–7 days • 8–30 days • 31–89 days • 90 days and more	43	 6 (14.0 %) 13 (30.2 %) 20 (46.5 %) 1 (2.3 %) 3 (7.0 %)	29	 0 (0 %) 11 (37.9 %) 14 (48.2 %) 2 (6.9 %) 2 (6.9 %)	κ_W = .618
During the past 3 months, did you consult a doctor or a hospital outpatient clinic because of your back pain? (yes/no)	132	34 (25.8 %)	117	23 (19.7 %)	κ = .786 κ = 1.000
During the past 3 months, did you receive physiotherapeutic treatment by a masseur physiotherapist because of your back pain? (yes/no)	131	24 (18.3 %)	118	17 (14.4 %)	κ = .932 κ = 1.000
During the past 3 months, did you regularly get active to prevent or relieve your back pain, e.g. by participating in a gym course, swimming or going to a fitness centre? (yes/no)	132	63 (47.7 %)	118	53 (44.9 %)	κ = .860 κ = .979

Item a, b	**TEST (N = 217)**		**RETEST (N = 179)**		κ
	N	**M, SD or N, % "yes"**	**N**	**M, SD or N, % "yes"**	κ_W **ICC** c
Below you find 10 different types of pain. For each kind of pain, please mark if you experienced this pain during the past 3 months:					
• headache or migraine	185	97 (52.4 %)	151	90 (56.6 %)	$\kappa = .758$
• pain in face, masseters, jaw joints or ears	178	29 (16.3 %)	140	28 (20.0 %)	$\kappa = .768$
• pain in neck	190	103 (54.2 %)	154	83 (53.9 %)	$\kappa = .691$
• pain in shoulders	192	87 (45.3 %)	152	64 (42.1 %)	$\kappa = .707$
• pain in hands or arms	179	45 (25.1 %)	148	38 (25.7 %)	$\kappa = .743$
• pain in chest	175	21 (12.0 %)	137	12 (8.8 %)	$\kappa = .692$
• pain in stomach	174	56 (32.2 %)	141	44 (31.2 %)	$\kappa = .732$
• pain in hips	192	55 (28.6 %)	153	44 (28.8 %)	$\kappa = .848$
• abdomal pain	182	38 (20.9 %)	144	28 (19.4 %)	$\kappa = .678$
• pain in legs or feet	192	76 (39.6 %)	151	65 (43.0 %)	$\kappa = .677$
Did you ever have had – longer than 12 months ago – back pain? (yes/no)	216	171 (79.2 %)	179	133 (74.3 %)	$\kappa = .729$
If you had experienced back pain earlier, at what age did you have back pain for the first time? (I was about ___ years old.) categorized:					
• up to 10 years					
• 11–20 years	174	8 (4.6 %)	128	4 (3.1 %)	
• 21–30 years		49 (28.2 %)		41 (32.0 %)	ICC = .941
• 31–40 years		48 (27.6 %)		39 (30.5 %)	$\kappa_w = .927$
• 41–50 years		31 (17.8 %)		25 (19.5 %)	
• 51–60 years		20 (11.5 %)		11 (8.6 %)	
• 61–70 years		11 (6.3 %)		4 (3.1 %)	
• more than 70 years		6 (3.4 %)		4 (3.1 %)	
		1 (.6 %)		0 (0 %)	
If you are employed and experienced back pain during the past 3 months, please answer four last questions: During the past 3 months, how many days were you work disabled off work because of your back pain? (categorized yes/no))	99	4 (4.0 %)	94	6 (6.4 %)	$\kappa = 1.000$ $\kappa = 1.000$
Thinking of your back pain and your occupational perspective: Do you think you are able to work until retirement age? (1 = sure/2 = rather yes/ 3 = unsure/4 = rather no/5 = by no means)	110	2.1 (1.0)	95	2.0 (1.0)	$\kappa_W = .819$
Do you think your ability to work is permanently endangered by your back pain? (yes/no)	107	17 (15.9 %)	96	11 (11.5 %)	$\kappa = .766$
Do you currently think about applying for a pension? (yes/no)	110	2 (1.8 %)	98	1 (.6 %)	$\kappa = .1.000$

Note: M = mean, SD = standard deviation;
a) Items sensitive to change over time (shaded) were included in an additional questionnaire covering the time interval between test and retest.
b) Shaded reliability results mark secondary analyses excluding respondents with instable answering due to changes in back pain during the time interval between test and retest. If primary and secondary analyses agree, results of primary analyses are reported.
c) κ = Cohen´s Kappa, κ_W = quadratic weighted Kappa, ICC = intraclass correlation coefficient, model "one-way, random"

of potential disagreement. In the second step, respondents with instable answers due to changes in back pain during the time interval between test and retest were excluded from the analysis of each single item. This applied to two cases: 1) The occurrence of an event was reported in the test survey, but negated explicitly in the retest survey as well as in the additional questionnaire, or 2) The occurrence of an event was explicitly negated in the test survey but reported in both the retest survey and the additional questionnaire.

As expected, reducing cases in terms of these instable respondents due to back pain led to higher reliability scores. However, the resulting differences between the first and the second step of analyses were of a low extent. A list of analysed items and associated results of both steps of analysis are shown in Table 2.

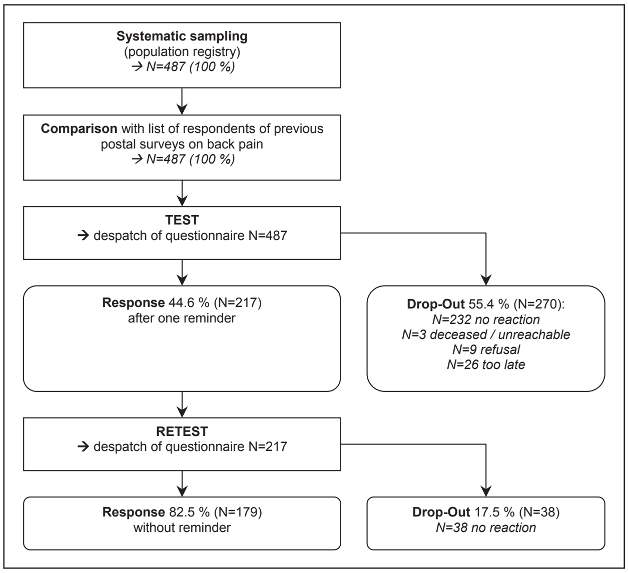

Figure 2 Participation development

Discussion

Earlier studies revealed different results on the reliability of diverse low back pain instruments. A recent review focusing on international cross-cultural adaptations of self-report questionnaires for low back pain showed good to excellent test-retest reliability [24]. Davidson and Keating compared five low back pain disability questionnaires [4]. They analyzed a subgroup of 47 low back pain patients classified as "unchanged" over the six week interval between their test and retest administration. The study provided test-retest results ranging from poor (SF-36 Bodily Pain scale ICC = .37; SF-36 Role Limitations-Physical scale ICC = .39) to fair/good (Roland-Morris-Disability Questionnaire ICC = .53; Waddell Disability Index ICC = .74) and excellent reliability scores (Quebec Pain Disability Scale ICC = .84; Oswestry Disability Questionnaire ICC = .84), respectively.
Our results confirm and complement reliability results of Mannion et al. [3]. They examined a core-set of six questions on low back pain in 45 German low back pain patients completing a questionnaire twice one to two weeks apart: In their study, the question "How severe was your back pain *in the last week*?" was attested good reliability on a visual analogue scale ranging from 0 = "no pain" to 10 = "worst pain I can imagine" (ICC = .71). Referring this question to the *past three months* a similar reliability score (ICC = .75) was found in our analyses. Questions on back function ("During the past week, how much did your back problem interfere with your normal work, including both work outside the home and housework?") and disability ("During the past four weeks, how many days did you cut down on the things you usually do (work, housework, school, recreational activities) because of your back problem?"; "During the past four weeks, how many days did your back problem keep you away from going to work (job, school, housework)?") showed good (ICC = .72) and excellent (ICC = .95) reliability, respectively, on a five-point Likert-scale [3]. Again, referring these questions to the *past three months* our analyses disclosed comparably good to very good kappa values (Tab. 1).
Based on the interpretation of Altman and Fleiss [20; 23] all analysed items presented good to very good test-retest reliability. By interpretation of results one needs to consider limitations of the study indicating both a potential under- and overestimation of reliability:

Selection bias, sample

Comparing the samples of the regional back pain survey 2003 [7] with the test-retest survey 2004, two significant differences occurred: 1) In comparison to the back pain survey 2003 the sample of the test-retest survey reported less "current bodily pain". One reason might be found in the different foci of both surveys. In 2003 the prior epidemiological background of the study might have led to an increased motivation in people suffering from back pain to participate. As a result of the methodological focus of our test-retest survey probably also less affected persons responded, informed by the following cover letter: "Enclosed you will find a first questionnaire. It will take you about ten minutes to complete it. We would like to ask you to fill out and return the questionnaire to the Institute of Social Medicine within the coming few days using the self-addressed envelope enclosed. In this case we would like to resend you an identical questionnaire after two weeks, again asking for completion and return as soon as possible. This procedure might be irritating, but it is necessary to estimate the reliability of our questionnaire". 2) In contrast to the back pain survey in 2003, the methodological focus of our study might have resulted in greater participation in persons of higher social status which should be related to educational status. In both surveys (2003 and 2004) significantly increased participation rates were found in women and elderly persons. The test-retest survey exposed a bias in terms of younger non-respondents that might have caused a reduced variance in our sample and by this an underestimation of reliability.

Answering pattern

Persons reporting back pain in both the test and the retest survey might state an increased (due to a prolonged duration of their current back pain episode) or a diminished (based on a decreased intensity of back pain) affliction in some parameters, respectively. In both cases the underestimation of reliability is possible.

The information on the methodological focus of the survey might have tempted some respondents to make mental notes of the test survey to enable stable answers in the retest survey. However, there was not any single case of identically answered test and retest questionnaires identified. Thus, there is no evidence that respondents knowingly kept their answers in mind.

Calculations of ICC

Due to variance analytical elements the calculation of an ICC requires the normal distribution, the homogeneity of variance, and the equidistance of the metric data analysed. Within the study, it was not possible to display all metric data in a normal distribution. In these cases, the parameters were categorized and analysed using Cohen´s kappa or quadratic weighted kappa, respectively (Tab. 1).

Rating scales (e. g. general health status) should have been examined in terms of equidistance of adjacent categories. In this case the ICC ("one way, random") was calculated and, additionally, quadratic weighted kappa for ordered categories (Tab. 1). The difference of these two reliability measures was marginal. Therefore, possible differences due to the level of measurement were not considered relevant [18].

Missing values

Missing values were excluded from the reliability analyses. A potential overestimation of reliability is pointed out, if non-categorized data of one rater (or in one of the test or retest survey, respectively) are disregarded [18; pp. 50]. The authors advise to generate a special category for non-categorized data that becomes included in reliability analyses. This procedure applies to nominal data, exclusively. For these additional analyses including a category for non-categorized data were performed. Cohen´s kappa values were marginally lower compared to analyses without regard to non-categorized data.

In the calculation of quadratic weighted kappa or intraclass correlation coefficients the equidistance of adjacent categories is assumed. Adding a category for non-categorized data infringes upon this presumption and was considered dispensable.

The limitations of this study clarify the restricted generalizability on similar instruments. The shown reliability scores do not refer to a specific instrument but have to be interpreted considering the underlying sample. Others emphasize the necessity of cross-cultural adaptations of low back pain questionnaires and their psychometric evaluation to allow comparability and integration [24]. Earlier study results [3] and the presented analyses show good to excellent reliability scores encouraging the use and adaptation of the tested questionnaire in further studies.

Acknowledgements

We would like to thank all respondents of the questionnaires for their participation, and Ben Chase and Dr. Nathalie Glaser-Möller for their support in the translation of the manuscript. The study was partly financed by the Federal Ministry for Education and Research funding the German Back Pain Research Network, and by the Institute of Social Medicine (University of Lübeck, Germany).

References

1. Raspe H, Matthis C, Croft P, O`Neill T, European Vertebral Osteoporosis Study Group. Variation in back pain between countries: the example of Britain and Germany. Spine 2004;29(9):1017–21.

2. Deyo RA, Battie M, Beurskens AJ, et al.. Outcome measures for low back pain research. Spine 1998;23(18):2003–13.

3. Mannion AF, Elfering A, Staerkle R, et al. Outcome assessment in low back pain: how low can you go? Eur Spine 2005;J 14:1014–26.

4. Davidson M, Keating JL. A comparison of five low back pain disability questionnaires: reliability and responsiveness. Phys Ther 2002;82(2):8–24.

5. Kohlmann T, Raspe H. Die patientennahe Diagnostik von Funktionseinschränkungen im Alltag. Psychomed 1994;6:21–7.

6. Kohlmann T, Raspe H. Zur Messung der Funktionskapazität bei Rückenschmerzpatienten in der Rehabilitation: ein Vergleich standardisierter Fragebogen. Rehabilitation 1996;35:103–8.

7. Hüppe A, Muller K, Raspe H. Is the occurrence of back pain in Germany decreasing? Two regional postal surveys a decade apart. Eur J Public Health 2007;17:318–22.

8. Kohlmann T, Deck R, Raspe H. Prävalenz und Schweregrad von Rückenschmerzen in der Lübecker Bevölkerung. Akt Rheumatol 1995;20:99–104.

9. Raspe A, Matthis C, Héon-Klin V, Raspe H. Chronische Rückenschmerzen: Mehr als Schmerzen im Rücken. Ergebnisse eines regionalen Surveys unter Versicherten einer Landesversicherungsanstalt. Rehabilitation 2003;42: 195–203.

10. Raspe H, Kohlmann T. Rückenschmerzen – Eine Epidemie unserer Tage? Dtsch Arztebl 1993;90(44):A2920–A5.

11. Raspe H, Kohlmann T. Disorders characterised by pain: a methodological review of population surveys. J Epidemiol Community Health 1994;48:531–7.

12. Berger-Schmitt R, Kohlmann T, Raspe H. Rückenschmerzen in Ost- und Westdeutschland. Gesundheitswesen 1996;58:519–24.

13. Bailar JC, Mosteller F, eds. Medical uses of statistics (second edition). Boston: New England Journal of Medicine Press 1992.

14. Staes F, Stappaerts K, Vertommen H, Everaert D, Coppieters M. Reproducability of a survey questionnaire for the investigation of low back problems of adolescents. Acta Paediatr 1999;88:1269–73.

15. Bullinger M, Kirchberger I. Der SF-36 Fragebogen zum Gesundheitszustand. Göttingen: Hogrefe 1998.

16. Meyer T, Schäfer I, Matthis C, Kohlmann T, Mittag O. Missing data due to a 'checklist misconception-effect'. Soz Praventivmed 2006;51:34–42.

17. Mittag O, Meyer T, Deck R, Kohlmann T, Matthis C, Schäfer I, et al. Fehlende Werte in der Somatisierungsskala des SCL-90-R durch „Listenfrage-Effekt“: Häufigkeit in verschiedenen Surveys und Vergleich verschiedener Imputationsmethoden. Gesundheitswesen 2005;67:155–8.

18. Wirtz M, Caspar F. Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen Hogrefe 2002.

19. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46.

20. Altman DG. Practical statistics for medical research. Boca Raton Chapman and Hall/CRC 1991.

21. Wirtz M. Bestimmung der Güte von Beurteilereinschätzungen mittels der Intraklassenkorrelation und Verbesserung von Beurteilereinschätzungen. Rehabilitation 2004;43:384–9.

22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1(1):30–46.

23. Fleiss J. The design and analysis of clinical experiments. New York John Wiley and Sons 1996.

24. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain. Searching for international cross-cultural adaptations. Spine 2007;32(9):1028–37.

Address for correspondence

Dipl.-Psych. Susanne Schlademann
University of Lübeck
Institute of Social Medicine
Beckergrube 43–47
D-23552 Lübeck
Germany
e-mail: susanne.schlademann@uk-sh.de

To access this journal online:
http://www.birkhauser.ch/IJPH