Pain assessment in younger and older pain patients: psychometric properties and patient preference of five commonly used measures of pain intensity

PAIN MEDICINE
Volume 8 • Number 7 • 2007

Pain Assessment in Younger and Older Pain Patients: Psychometric Properties and Patient Preference of Five Commonly Used Measures of Pain Intensity

Madelon L. Peters, PhD,* Jacob Patijn, MD, PhD,† and Inge Lamé, MSc†

*Department of Clinical Psychological Science, University Maastricht, Maastricht, The Netherlands;
†Pain Management and Research Center, Department of Anesthesiology, University Hospital Maastricht, Maastricht, The Netherlands

ABSTRACT

Objective. To study the psychometric properties and preference for five different pain intensity scales (horizontal visual analog scale [VAS], vertical VAS, Box-11, Box-21, and verbal descriptor scale) across different age groups.

Design. Chronic pain patients rated their present, average, weakest, and strongest pain on five different scales, and indicated scale preference.

Setting. Outpatient pain facility.

Results. The number of mistakes on all scales increased with increasing age, and the VAS appeared to be most prone to making mistakes. All scales appeared to be sufficiently valid, but the verbal descriptor scale was less related to the common underlying pain factor than the other scales. The Box-21 was the most preferred scale overall, although patients aged >75 years especially preferred the verbal descriptor scale.

Conclusion. The numerical Box-21 scale is an excellent choice for pain intensity assessment in heterogeneous patient groups. The verbal descriptor scales may be considered when the study population consists of a majority of older persons.

Key Words. Pain Assessment; Older Pain Patients; Validity; Preference

Introduction

During the next decades, health care services in developed countries will be increasingly called upon to treat older patients with pain symptoms. The proportion of older people in North American and West European countries is increasing, and pain is a prevalent complaint among older people [1–3]. Although the exact prevalence figures differ, the majority of epidemiological studies report a significant increase of chronic pain with advancing age [4–7]. Therefore, the percentage of older patients in pain treatment and management facilities can be expected to rise considerably in the coming years, and this may have implications for pain assessment, intervention, and management. Reliable and accurate pain assessment is a prerequisite for successful pain management, and it is thus important that pain facilities use instruments with proven reliability and validity in both younger and older pain patients. The most frequently used scales for measuring pain intensity are visual analog scale (VAS), numerical rating scale (NRS), and verbal descriptor scale (VDS). Reliability and validity of these scales have been extensively tested in adults younger than 65 years, and in general appeared to be good [8]. Measures with a relatively large number of response alternatives (e.g., VAS) may have the advantage of greater sensitivity in comparison with instruments with fewer response alternatives (e.g., VDS; [9]). How-

Reprint requests to: Madelon L. Peters, PhD, Department of Clinical Psychological Science, University Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands. Tel: 00 31 43 3881603; Fax: 00 31 43 3884155; E-mail: madelon.peters@dep.unimaas.nl.

 doi:10.1111/j.1526-4637.2007.00311.x

ever, a VAS may be difficult to understand for some people, as demonstrated by a higher number of errors and a lower validity coefficient (i.e., association with other pain assessment tools) than found for numerical scales [10–14]. Therefore, numerical scales with 11 or 21 response alternatives have been proposed as good candidates for pain assessment in most (clinical and research) situations [10,13,15]. Although a scale with 11 levels may be less sensitive than a VAS, a scale with 21 levels has been shown to provide a sufficient level of discrimination for chronic pain patients to describe their pain [16].

There have been only a few studies comparing the usefulness and psychometric properties of the various pain intensity instruments specifically in older persons. Because sensory and cognitive capacities may decline with advancing age, but also because of lack of experience with psychometric tests, some instruments may be more appropriate than others for older persons [1,17,18]. A rather consistent finding in previous studies is that older persons show a higher failure rate on a VAS in comparison with various other pain intensity scales [12,19–24]. Kremer et al. [10] reported that VAS noncompleters were significantly older than patients who did complete the VAS (75 vs 54 years). Moreover, it has been suggested that older patients may have difficulty with a horizontal VAS in particular, and that they would be more comfortable in using a vertical VAS [18–20].

Based on a series of studies, Herr and colleagues advocate the use of a VDS in older persons with pain. Older people preferred the VDS to a (horizontal or vertical) VAS [19], and the completion rate for the VDS appeared to be high [18]. In a more recent study in which younger and older volunteers participated in an experimental pain test, a 7-point VDS, a vertical VAS, an 11-point NRS, a 21-point NRS, and the "Faces of Pain Scale" (five pictures of progressively distressed facial expressions) were compared [23]. The VDS proved to be the most sensitive and reliable pain assessment instrument. In contrast, Chibnall and Tait [25] propose that a 21-point NRS is the scale of choice for older persons. In a comparison between the numerical Box-21 scale, the Gracely Box scale (a 21-point NRS with verbal label at 12 of the 21 points), a 5-point VDS, and the "Faces of Pain Scale," they found that the numerical Box-21 had superior psychometric properties and led to fewer mistakes than the other scales in hospitalized older persons with no to moderate cognitive impairment. Finally, Gagliese et al. compared a horizontal VAS, a vertical VAS, an 11-point NRS, a 7-point VDS, and the McGill Pain Questionnaire to assess postoperative pain in younger and older surgical patients, and concluded that the NRS should be the first choice for older persons [24].

In sum, at present it is still undecided which pain intensity scale is most suitable for older patients. There seems to be agreement on the fact that the VAS is the least suitable scale to score pain intensity in older adults, although some have argued that this only applies to a horizontal version of the scale. There is less agreement on what would be the most appropriate pain intensity scale for older people; both the VDS and numerical scales have been proposed as the best choice. Another issue is that in many situations, it would be preferable to have an instrument that is appropriate for younger and older patients alike. In clinical practice, it is not feasible to select different instruments for individual patients and, therefore, it is important to have available an instrument that can be used for initial assessment in patients of varying ages. The present study was designed to compare the utility of five widely used pain intensity scales (horizontal VAS, vertical VAS, two different numerical scales, and a VDS) as an initial assessment instrument in a heterogeneous patient population varying in age and educational level as seen in pain treatment and management facilities.

In outpatient pain facilities, it is common practice that patients receive questionnaires to be filled out at home without control on how and when they do this and without professional assistance available. To reflect this clinical situations, in the present study we also mailed the questionnaires to study participants and did not offer assistance in filling them in. We compared errors, reliability, and validity of a vertical VAS, a horizontal VAS, an 11-point NRS (Box-11), a 21-point NRS (Box-21), and a VDS with six increasing levels of pain intensity. We focused on three questions:

1. Which scale has the least incorrect responses?
2. Are the scales valid measures of the construct "pain intensity"?
3. Which scale do patients prefer?

We hypothesized that all five scales would be valid representations of the construct "pain intensity," but that more incorrect responses would be made on the horizontal VAS than on the numerical and verbal scales, especially in older people. We further expected that the number of incorrect responses on the vertical VAS would be lower than on the horizontal VAS, but higher than on the

other three scales. Finally, we expected that both younger and older patients would show the lowest preference for the VAS (either horizontal or vertical). We had no specific hypothesis for differences in preference of the other three scales.

Methods

Participants

Participants for the study were recruited from the outpatient pain management facility of the University Hospital Maastricht. All participant were referred to the facility for assessment and management of recurrent or chronic pain. We strived for an approximately even distribution of participants across six age brackets (<35, 35–44, 45–54, 55–64, 65–74, and >75 years), with a minimum of 50 patients in each bracket. To accomplish this, we needed 5 months of unselected sampling, during which all referrals to the facility were asked to participate in the study, followed by 3 months of selective sampling in the youngest and two oldest age groups, and another 2 months of sampling for the oldest age group. Finally, a total of 338 patients participated in the study. All participants were informed on the purpose of the study and signed an informed consent form.

Measures

Participants received a questionnaire consisting of: (1) questions on general demographic information (age, sex, and education) and location of major pain complaint; (2) five different scales to measure four aspects of pain intensity (present pain, strongest pain, weakest pain, and average pain last week); and (3) three questions related to scale preference.

Horizontal and Vertical VAS

For the present study, we used a VAS consisting of a 10-cm horizontal line, with the labels "no pain" (left) and "the worst pain imaginable" (right). The same [21] or similar descriptors (i.e., "no pain – the most intense pain imaginable"; [23]) have been used in other studies. Because previous research has suggested that a vertical VAS may be more appropriate for older persons [18–20,24], we also included a vertical version. The same labels were used, with "no pain" on the bottom of the scale and "the worst pain imaginable" at the top.

Box-11 and Box-21

Box scales are a special form of NRS, in which the numbers are presented in a horizontal row of boxes. Participants place an X in the box with the number that best reflects their pain intensity [11,26]. The Box-11 uses numbers from 0 to 10, with the extremes labeled "no pain" (0) and "worst pain imaginable" (10). The Box-21 is identical to the Box-11, but presents 21 boxes in a row, with numbers labeled from 0 to 100 in increments of 5 [15,25].

Verbal Descriptor Scale

Verbal descriptor scale (VDS) consists of adjectives describing increasing levels of pain intensity [8]. For the present study, we used the six-level pain rating scale from the SF-36 because validated translations in many languages are available for this scale (for an overview, see [9]). The English labels for this scale are (from top to bottom): none, very mild, mild, moderate, severe, and very severe.

All scales and the instructions were presented in Dutch. To avoid order effects, five different versions of the questionnaire booklet were made, with each type of scale once occurring first and once occurring last. Each patient could receive, at random, one of the five different versions. All of the scales were presented on separate sheets and were preceded by a clear instruction. Four different intensity ratings were asked for each scale, in fixed order: (1) How much pain do you experience *right now*? (2) How much pain did you experience when it was *at its worst* last week? (3) How much pain did you experience when it was *at its least* last week? (4) How much pain did you experience *on average* last week? Patients were requested to complete the scales in the order as presented in the questionnaire. Thus, they first answered all five questions for one scale, and then could go on to the next scale.

Following Herr and Mobily [19], on the last page of the questionnaire booklet, participants rated which scale best described their pain, which scale was most easy to fill out, and which scale they preferred. Below each question all five scales were listed, and participants had to circle one of the scales. To avoid incorrect responding on the basis of nonspecific factors (e.g., impaired visual abilities), throughout the questionnaire a large font (Arial 14 pt) was used (cf. Herr and Mobily [19]).

Procedure

The study was approved by the Medical Ethical Committee of the University Hospital Maastricht. All referrals to the university outpatient pain facility received a booklet with questionnaires to be filled out before their first appointment as part of the standard intake procedure. For the duration of

this study, the envelope contained a letter explaining the study, an informed consent form, and a separate questionnaire with the five pain scales and the questions regarding scale preference. The first page of this separate questionnaire clearly stated that these additional scales were for research purposes only, and that patients could freely choose to participate. Patients were asked to fill out all questions during a single occasion, and not to skip any questions. Moreover, it was stressed that this was not a knowledge test, and that there were no right or wrong answers. Patients could return the informed consent and the research questionnaire at the reception desk during their first appointment.

All 20 pain ratings (4 ratings on 5 scales) per patient were reviewed and coded for incorrect responses. Incorrect responses could take several forms, e.g., two responses on one scale or a range of responses, a nonexisting response or a rating in-between two response categories, or no response at all. For each type of scale, only one incorrect response was counted per participant; for example, if a participant skipped all four pain intensity questions on the Box-11, this was counted as one incorrect response. Similarly, if a participant circled two response alternatives on the first VDS, and none on the others, this was coded as one incorrect response for the VDS. In this way, a theoretical number of 1,690 (338 participants × 5 scales) incorrect responses were possible.

There was also another type of mistake that participants could make when filling out the scales, i.e., improper use of the ordinal qualities of the scales. Per scale, four assessments were made: momentary pain, weakest pain, strongest pain, and average pain. When a scale and its labels are understood properly, one should expect strongest pain to be rated as stronger than or at least equal to average pain, and average pain to be rated as stronger than or equal to weakest pain. When either of the comparisons strongest minus average pain or average minus weakest pain yielded a negative outcome, this was considered a mistake in ordinal understanding of the scale.[1]

Data Analyses

We calculated the number of incorrect responses per scale and per age bracket. Next, for each scale the total number of incorrect responses was listed as a function of the position of this scale in the questionnaire. This enabled us to examine whether there was an order effect (i.e., whether more mistakes were made on either the first or the last scale that a participant had to fill out). To examine whether the number of incorrect responses or mistakes in scale ordinality are significantly related to age, education, scale type, or their interaction, multilevel logistic regression analyses were performed using Mlwin (The Centre for Multi-level Modelling, London, UK) with "incorrect response" or "mistake in scale ordinality" (yes/no) as the dichotomous outcome variables, "age" and "education" as between-subject predictor variables, and "scale type" as within-subject predictor variable. Age was entered as a continuous variable, education was dummy coded as a dichotomous variable (highest education completed: none, elementary school, or lower vocational training vs high school, advanced vocational training, or university degree). Scale type was coded with four dummy variables. An interaction effect of age with scale type was examined by entering the products of age and the four dummies coding scale type in the analyses. In order to test the significance of the predictors in the model, z-values were calculated (beta/SE of beta).

Following Jensen et al. [11] and Chibnall and Tait [25], construct validity was assessed by means of factor analysis, using the principal axis extraction method. It was predicted that a single pain intensity construct would emerge, and that all scales would significantly load on this factor. The magnitude of the loadings is an indication of how well a specific scale captures the underlying pain intensity construct. The factor analyses were performed for each of the four pain scores separately (momentary pain, strongest pain, weakest pain, and average pain). Invalid scores (e.g., when an ordinal mistake was made, or when an entry was not interpretable) were excluded from these analyses.

Finally, whether scale preference was related to sociodemographic variables was tested by means of logistic regression for each of the scales separately. The dependent variable was "scale preferred" (yes/no), and independents were age, sex, and education.

Results

Description of Sample

The mean age of patients was 54 years (range 18–92), and 63% were female (212 female, 123 male,

[1]For the VAS, a margin of ±10 mm was employed to decide whether the outcome of the comparison indicated ordinal problems. Respondents cannot be expected to place a mark on the exact same position of a line when scoring the same pain level twice.

Table 1 Sample size, demographics, and number (and percentage) of incorrect ratings and ordinal mistakes across the various age brackets

Age Bracket (years)	N	% Female	% Low Education*	No. Incorrect Ratings (%) All scales	No. Ordinal Mistakes (%) All scales	Number of Ordinal Mistakes Per Scale VDS	Box-11	Box-21	Horizontal VAS	Vertical VAS
<35	53	83	33	2 (0.8)	7 (2.6)	1	0	3	1	2
35–44	55	58	40	5 (1.8)	5 (1.8)	1	0	1	0	3
45–54	73	57	51	11 (3.0)	8 (2.2)	1	2	2	2	1
55–64	55	55	59	4 (1.5)	5 (1.8)	1	1	1	0	2
65–74	52	67	67	10 (3.8)	13 (5.0)	1	2	2	1	7
>75	50	62	64	15 (6.0)	9 (3.6)	2	1	2	0	4
Total	338	63	52	47 (2.8)	47 (2.8)	7	6	11	4	19

* Low education is defined as none, elementary school, or lower vocational training.
VAS = visual analog scale; VDS = verbal descriptor scale.

and 3 unknown). Major pain locations were back (39%), legs (16%), and neck/shoulder (12%). In total, 52% of patients had completed elementary school or lower vocational training as their highest education; 15% had completed university or higher vocational education. All age brackets consisted of at least 50 participants, with a little overrepresentation in the age bracket of 45–54 years (N = 73; all other brackets: N = 50–55). Table 1 presents an overview of the demographics per age bracket.

Incorrect Responses

The data showed that few incorrect responses were given: an overall error rate of 47 (i.e., 2.8% of the potential number of 1,690 errors) was found across age groups and scale type. First, it was checked whether there was an order effect on mistakes, i.e., whether scales that were presented later in the questionnaire were more (or less) likely to show mistakes. This appeared not the case. There were only few differences in number of incorrect responses on each of the scales, with the Box-21 showing the lowest number of mistakes (7) and the vertical VAS showing the highest number of mistakes (12). Absolute number and percentage of incorrect responses on any scale per age bracket are shown in Table 1 (column 5). Number of incorrect responses appears to increase in the two oldest age groups, although absolute numbers remain low.

Multilevel logistic regression analysis was used to test whether number of incorrect responses was related to age, education, scale type, or their interaction. In the first step, age and education were entered in the model. Age proved to be significantly related to making a mistake ($z = 2.36$, $P = 0.02$), with older patients making more mistakes. Education was nonsignificant ($z = 0.18$, $P = 0.86$), but was retained in the model as a control variable.[2] In the next step, the four dummies coding scale type were entered to the equation. There was no significant difference in number of mistakes on the five scales. Finally, interaction terms (age * scale type) were entered, but these neither proved significant.

Mistakes in Understanding of the Ordinal Properties of a Scale

Overall, 47 mistakes were made in scale ordinality (i.e., 2.8% of the potential number of 1,690 mistakes). There were notable differences in this kind of mistakes between scales: most mistakes were made on the vertical VAS and least on the horizontal VAS (Table 1). As apparent from Table 1, again a slight increase with age was noted, but again total number of errors remained low.

To test the contribution of age, education, and scale type to the total number of ordinal mistakes on all scales, similar multilevel logistic regression analyses as described above for incorrect responses were performed. In step 1, age and education were entered in the equation. Both proved to be significantly related to making an ordinal mistake (age: $z = 2.36$, $P = 0.018$; education: $z = 2.06$, $P = 0.039$), older people and, in contrast to our expectations, higher-educated patients made more mistakes. In step 2, the four dummies that coded scale type were entered. Together they significantly improved the model ($\chi^2 = 14.22$, df = 4, $P = 0.007$), indicating that number of ordinal mistakes differed between scales. Only the dummy that coded for the contrast between horizontal VAS and vertical VAS reached significance on its

[2]In general, older people had had less education.

own ($z = 2.91$, $P = 0.004$). Finally, in step 3, the interaction terms were entered, again significantly improving the model ($\chi^2 = 9.639$, df = 4, $P = 0.047$). This interaction could be traced back to the fact that the age effect was most apparent on the vertical VAS.

Validity

Following Jensen et al. [11] and Chibnall and Tait [25], construct validity was assessed by means of factor analysis, separately for all four pain ratings (momentary, strongest, weakest, and average pain). In all four analyses, a single factor emerged, as was evident from the scree plot and eigenvalues. The eigenvalue for the first factor varied from 4.40 to 4.63 (explained variance: 88.0–92.6%), while all eigenvalues for the second factor were below 0.3. Table 2 shows the factor loadings of each of the five scales on the first factor. Although all scales had a significant loading on the "pain intensity" factor, the Box-11 and Box-21 consistently showed the highest loadings, closely followed by the horizontal VAS. The vertical VAS, and especially the VDS, appeared to be a little less strongly related to the underlying pain intensity construct. To examine whether construct validity of the scales would differ per age group, these analyses were repeated for patients under the age of 65 years and patients of 65 years and older separately. As the lower part of Table 2 shows, the pattern of factor loadings was similar for both groups.

Table 2 Factor loadings of the five pain intensity scales on the primary factor for all patients, patients under 65, and patients aged 65 and over

	Present Pain	Max. Pain	Min. Pain	Average Pain
All patients				
VDS	0.88	0.83	0.89	0.83
Box-11	0.97	0.96	0.98	0.95
Box-21	0.98	0.97	0.98	0.97
Horizontal VAS	0.95	0.94	0.97	0.95
Vertical VAS	0.92	0.92	0.94	0.92
Patients <65				
VDS	0.87	0.81	0.88	0.81
Box-11	0.98	0.95	0.98	0.95
Box-21	0.98	0.97	0.99	0.95
Horizontal VAS	0.97	0.94	0.98	0.95
Vertical VAS	0.92	0.92	0.94	0.90
Patients ≥65				
VDS	0.90	0.87	0.92	0.85
Box-11	0.97	0.97	0.98	0.94
Box-21	0.99	0.98	0.98	0.99
Horizontal VAS	0.91	0.95	0.96	0.95
Vertical VAS	0.91	0.92	0.96	0.95

VAS = visual analog scale; VDS = verbal descriptor scale.

Finally, a test was performed to check whether results were not due to a statistical artifact. The VDS has six response alternatives, while the other scales have more levels (11, 21, or 101), potentially allowing for more variance to be explained. Therefore, the scores of all scales were converted to a 6-point scale (e.g., score 0 through 16 on the VAS received a score of "1," 17 through 35 a score of "2," etc.). A principal axis analysis using these new response alternatives, in essence, gave the same outcome, the VDS still loading notably less on the first factor.

Scale Preference

Three questions were asked at the end of the questionnaire: Which scale best described your pain? Which scale was most easy to use? Which scale do you prefer? It should be noted that 45 to 51 participants did not answer one or more of these questions, leaving about 290 scores for data analyses.

There was little differentiation in answers between the three questions (i.e., most people who endorsed a specific scale as best describing their pain also endorsed this scale as "most easy" and "preferred"). Therefore, only the data for "preferred scale" are presented here. The box scales were most preferred: 49.3% of respondents endorsed either the Box-11 (13.7%) or Box-21 (35.6%). The VDS was preferred by 25% of respondents, while another 25.7% of respondents preferred one of the VAS scales (horizontal 11%, vertical 14.7%). Next, we looked separately at scale preference for respondents under 65 years of age, for respondents aged 65–75 years, and for respondents older than 75 years. Preference did not change much from <65 to 65–75 years, but for the oldest participants, there was a clear shift toward preference for the VDS (VDS: 42.9%, Box-11: 0%, Box-21: 28.6%, horizontal VAS: 11.4%, and vertical VAS 17.1%).

Whether preference was related to age, sex, or education was further tested by means of logistic regression for each of the scales separately. Preference for the VDS was significantly predicted by age (dummy coded as under 75 vs 75 or older) and sex. Patients of 75 years or older and female individuals more often preferred the VDS (age: Wald = 6.09, $P = 0.014$, OR = 2.5, CI: 0.30–0.98; sex: Wald = 4.18, $P = 0.041$, OR = 1.9, CI: 0.19–0.83). There was no age–sex interaction. No significant predictors were identified for the other scales.

Discussion

A major finding of the present study is that most pain intensity scales can be safely used for initial assessment in the large majority of patients presenting to an outpatient pain management facility, including older patients. In the present sample, only very few mistakes were made in scale completion and interpretation. This confirms the findings of previous studies reporting only minimal errors across different scales in non- to moderately cognitively impaired older patients [19,23,25]. Nevertheless, our results show that there is a significant increase in incorrect responding with advancing age. Although absolute number of mistakes remained low, older patients more often gave an incorrect or no response across all five scales, and more often did not fully understand the scale labels as shown by mistakes in ordinal judgements (i.e., rating weakest, average, and strongest pain).

We did not find differences in number of errors between the five scales, at least as far as incorrect responses (e.g., mark outside category) or no response at all are considered. Previous studies have indicated that VAS may be more problematic for older persons and/or persons with cognitive impairments than numerical scales or VDS [10,11,20,21,23,24,27]. Some authors suggested that a vertical VAS might be less sensitive to errors because a vertical line orientation would require less abstract thinking (i.e., "highest" pain is scored "high" on the line [18,19]). However, a recent study by Herr et al. [23], in which a vertical VAS was compared with four other pain intensity scales in younger and older participants, still found this scale to be problematic for older participants, and especially for people with cognitive or motor impairments.

Moreover, our study demonstrates that a vertical VAS may induce another type of error. The fact that a person can correctly place a mark on a scale does not necessarily mean that he/she really understands the meaning of this scale, and that his/her rating accurately reflects pain intensity. Therefore, we also took another type of error into account, namely whether a participant rates different levels of pain in the right order on the scale. Especially the vertical VAS proved to be problematic in this respect, with several patients, and especially older patients, giving "strongest pain" a lower rating than "weakest pain." This suggests that it is not self-evident that the top of the scale means "highest" pain. Adding graphic features like colors (e.g., from white to red) as with the Colored Analog Scale (CAS; [28]) may help correct interpretation and prevent this particular problem (cf. [29]). However, the scarce amount of research with the CAS in older persons suggests that adding colors does not resolve all problems attached to the use of analog scales. Older people still experienced more difficulty in completing a CAS than a verbal descriptor or a numerical scale [22].

To summarize, even though most younger and older patients are able to score their pain on either a VAS, NRS or VDS, there is converging evidence that the VAS is most prone to errors, especially in older individuals [11,18,19,22–25]. By contrast, even with moderate to moderately severe cognitive impairment, a large majority of patients may still be able to complete a VDS or a NRS [22,25]. Consequently, if an instrument has to be selected for a heterogeneous patient population of different ages in which no information on cognitive status of the respondents is available, a VAS is less suitable than verbal descriptor or numerical scales.

Another important issue when selecting a scale is its validity. Construct validity, operationalized as the association of a particular scale with the underlying "pain intensity factor" of all scales combined, proved to be sufficient for each of the five scales, in younger and older patients alike. Nevertheless, the box scales and the horizontal VAS showed a somewhat higher agreement with the underlying unifactorial pain construct than the vertical VAS, and especially the VDS. It could be argued that one reason for the lower loading of the VDS is that it has fewer response alternatives than the other scales. Two arguments contradict this explanation. First, the vertical VAS was also less strongly related to the underlying pain dimension, whereas this scale has 101 response alternatives. Second, when all ratings were transformed to a 6-point scale by clustering answer categories, the same result was obtained. Previously, in younger and middle-aged patients, Jensen et al. [11] also reported that especially 11- or 101-point numerical scales showed the highest loadings with the dominant factor (compared with VAS, and three different verbal scales). Studies investigating construct validity of various pain intensity scales specifically in older patients often found the VDS to have a somewhat lower agreement with other pain intensity scales, or with the underlying pain intensity factor, than numerical and VAS scales ([10,22,25]; however, see [23] for similarly strong loading of the VDS). One previous study examined whether loadings of several pain intensity scales on the dominant factor differed between

older and younger participants [24]. Similar to the present results, the pattern of factor loadings was comparable for older and younger participants, and the VDS had the lowest loading in both groups.

Another criterion for adopting a specific scale may be patient preference, especially when psychometric properties are satisfactory for several instruments. In the present study, patients showed a clear preference for the Box-21, while the horizontal VAS was the least preferred scale across all age groups. This result is very similar to what has recently been found by Herr et al. [23]. These authors reported a clear preference for the Box-21, compared with the VDS, a 10-point numerical scale, the Faces of Pain Scale, and a vertical VAS (in order of preference). No significant differences were found in scale preference in relation to age, sex, education, or cognitive function [23]. We did find an age effect on preference, but only when the oldest age group was compared with the rest. The oldest patient group (age 75+) preferred the VDS to all other scales. Two previous studies including only older participants (>65 years) found a preference for the VDS over other pain assessment instruments. In nursing home residents with varying degrees of cognitive impairment, Closs et al. [22] found a slight preference for the VDS over several other scales (e.g., numerical scale), while again the VAS was clearly least popular of all. Herr and Mobily [19] similarly found a preference for the VDS compared with a vertical and horizontal VAS scale, a NRS, and pain thermometer with verbal labels.

Notably, we found that female participants preferred the VDS more often than their male counterparts. In the study of Herr et al. [23], there was also a tendency for female participants to prefer the VDS (and male participants to prefer the 21-point numerical scale), but this difference did not reach significance. In contrast to our findings, Herr and Mobily [19] found that male participants preferred the VDS more often, and female participants preferred the pain thermometer more often. It should be noted that the pain thermometer is another scale using verbal descriptors. Whether there are sex differences in scale preferences should be confirmed in future studies.

Another finding in the present study was that patients more often selected the Box-21 as their most preferred scale compared with the Box-11. Patients may prefer a numerical scale with more response options over one with fewer options. Indeed, a few patients circled two consecutive numbers on the Box-11, and one even added a number in between two boxes (i.e., $5^1/_2$). Herr et al. [23] similarly found a difference in preference between a 21-point numerical scale and an 11-point numerical scale in favor of the first.

The aim of our study was to investigate the utility and preferences of five different pain scales for initial pain assessment in patients presenting at an outpatients pain facility. Our results suggest that each of the five instruments included in the present study can be reliably used in younger and older persons without known cognitive impairments. Previous studies found evidence that these instruments are also suitable for older patients with mild to moderate cognitive impairment [22,25]. Nevertheless, based on both psychometric properties and preference, we propose the Box-21 as the instrument of first choice in a mixed population of patients. One should be careful to use the VAS because of the poor comprehension and high error rate in some people.

We should be cautious in generalizing the findings of the present study beyond the population of relatively cognitively intact older adults. Studies focussing specifically on cognitively impaired older persons suggest that the VDS may be the most suitable for this group [22,30]. Moreover, because of the higher preference for the VDS in people over 75, one may consider using this scale when the patient population primarily consists of very old people.

Finally, some limitations of this study should be mentioned. We did not obtain data on the cognitive status of our participants. Although this resembles the clinical situation, where pain assessment is routinely made without first assessing cognitive functioning, it excludes inferences on why older patients made more mistakes. Herr et al. [23] have suggested that it is not age per se that impacts on failure to correctly use a particular scale, but the conditions that are often associated with aging, i.e., cognitive and motor impairment. Another possibility may be the fact that older people have been less exposed to questionnaires and other assessment instruments during education and working life. Prior scale exposure may impact the ability to use a scale correctly (cf. [23]). To establish whether a larger number of errors are actually due to aging per se and not to cohort effects, longitudinal studies are necessary.

A second limitation of the study was the fact that all patients filled out the questionnaires at home and we do not have information whether a relative or caretaker may have helped them. If this

has been the case in some instances, the number of mistakes may have been an underestimation of the true error rate. This does not challenge the results, but it could mean that the age-related increase in completing and interpreting scales could actually be more prominent.

Third, in accordance with previous studies, we assessed construct validity of each of the five scales by means of factor analysis [11,25]. This approach assumes that there is a common underlying pain intensity construct that is measured by each of the scales. The magnitude of the loadings of a scale on the common factor is used as an indication of how well this scale captures the underlying pain intensity construct. One could question this approach, as the "true" value of the underlying pain construct is unknown. However, as we do not have a gold standard to measure pain intensity other than self-report, no other means to establish validity were available.

Another methodological concern may be that participants had to give a total of 20 pain ratings. We cannot rule out that this large number of ratings per person introduced some additional unreliability due to fatigue or carryover effects. We tried to control for carryover effects by using five different versions of the questionnaire, each with a different order of scale presentation. Moreover, we reviewed whether the number of incorrect responses on a certain scale was higher when the scale was presented either as the first or the last scale, and found no indication that there was an order effect.

A final limitation is that in assessing the contribution of education to the number of incorrect responses and scale preference, we dichotomized educational level. We cannot exclude that an actual effect of education remained undetected because some of the variance in educational level was missed.

To conclude, when an instrument has to be selected for use in a heterogeneous population of patients, consisting of mixed age groups, and without information on cognitive functioning, a numerical scale, and more specifically the Box-21, is proposed as the instrument of first choice. A VDS may be considered instead when the patient population primarily consists of older persons, especially when there is the suspicion that part of the population may show cognitive decline.

References

1 American Geriatric Society. The management of chronic pain in older persons: AGS Panel on Chronic Pain in Older Persons. J Am Geriatr Soc 1998;46:635–51.
2 Fox PL, Raina P, Jadad AR. Prevalence and treatment of pain in older adults in nursing homes and other long-term care institutions: A systematic review. CMAJ 1999;160:329–33.
3 Otis JAD, McGeeney B. Managing pain in the elderly. Clin Geriatr 2001;9:82–8.
4 Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA. The epidemiology of chronic pain in the community. Lancet 1999;354:1248–52.
5 Harkins SW, Price DD. Assessment of pain in the elderly. In: Turk DC, Melzack R, eds. Handbook of Pain Assessment. New York: Guilford Press; 1992:315–31.
6 Helme RD, Gibson SJ. The epidemiology of pain in elderly people. Clin Geriatr Med 2001;17:417–31.
7 Crook J, Rideout E, Browne G. The prevalence of pain complaints in a general population. Pain 1984;18:299–314.
8 Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Melzack R, Turk DC, eds. Handbook of Pain Assessment, 2nd edition. New York: Guilford Press; 2001:15–34.
9 Caraceni A, Cherny N, Fainsinger R, et al. Pain measurement tools and methods in clinical research in palliative care: Recommendations of an Expert Working Group of the European Association of Palliative Care. J Pain Symptom Manage 2002;23:239–55.
10 Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: Patient preference does not confound pain measurement. Pain 1981;10:241–8.
11 Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. Pain 1986;27:117–26.
12 Jensen MP, Karoly P, O'Riordan EF, Bland F Jr, Burns RS. The subjective experience of acute pain. An assessment of the utility of 10 indices. Clin J Pain 1989;5:153–9.
13 Strong J, Ashton R, Chant D. Pain intensity measurement in chronic low back pain. Clin J Pain 1991;7:209–18.
14 Briggs M, Closs JS. A descriptive study of the use of visual analogue scales and verbal rating scales for the assessment of postoperative pain in orthopedic patients. J Pain Symptom Manage 1999; 18:438–46.
15 Jensen MP, Miller L, Fisher LD. Assessment of pain during medical procedures: A comparison of three scales. Clin J Pain 1998;14:343–9.
16 Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387–92.
17 Weiner D, Pieper C, McConnell E, Martinez S, Keefe F. Pain measurement in elders with chronic low back pain: Traditional and alternative approaches. Pain 1996;67:461–7.

18 Herr KA, Garand L. Assessment and measurement of pain in older adults. Clin Geriatr Med 2001;17:457–78.
19 Herr KA, Mobily PR. Comparison of selected pain assessment tools for use with the elderly. Appl Nurs Res 1993;6:39–46.
20 Gagliese L, Melzack R. The assessment of pain in the elderly. In: Mostofsky D, Lomranz J, eds. Handbook of Pain and Aging. New York: Plenum Press; 1997:69–96.
21 Bergh I, Sjostrom B, Oden A, Steen B. An application of pain rating scales in geriatric patients. Aging 2000;12:380–7.
22 Closs SJ, Barr B, Briggs M, Cash K, Seers K. A comparison of five pain assessment scales for nursing home residents with varying degrees of cognitive impairment. J Pain Symptom Manage 2004;27:196–205.
23 Herr KA, Spratt K, Mobily PR, Richardson G. Pain intensity assessment in older adults: Use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. Clin J Pain 2004;20:207–19.
24 Gagliese L, Weizblit N, Ellis W, Chan VWS. The measurement of postoperative pain: A comparison of intensity scales in younger and older surgical patients. Pain 2005;117:412–20.
25 Chibnall JT, Tait RC. Pain assessment in cognitively impaired and unimpaired older adults: A comparison of four scales. Pain 2001;92:173–86.
26 Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, eds. Handbook of Pain Assessment. New York: Guilford; 1992:135–51.
27 Gagliese L, Melzack R. Lack of evidence for age differences in pain beliefs. Pain Res Manage 1997;2:19–28.
28 McGrath PA, Seifert CE, Speechley KN, et al. A new analogue scale for assessing children's pain: An initial validation study. Pain 1996;64:435–43.
29 Scherder EJ, Bouma A. Visual analogue scales for pain assessment in Alzheimer's disease. Gerontology 2000;46:47–53.
30 Ferrell B, Whedon M, Rollins B. Pain and quality assessment/improvement. J Nurs Care Qual 1995;9:69–85.