Review Article

Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review

Marianne Jensen Hjermstad, PhD, Peter M. Fayers, PhD, Dagny F. Haugen, MD, PhD, Augusto Caraceni, MD, Geoffrey W. Hanks, DSc (Med), MB, Jon H. Loge, MD, PhD, Robin Fainsinger, MD, Nina Aass, MD, and Stein Kaasa, MD, PhD, on behalf of the European Palliative Care Research Collaborative (EPCRC)

Regional Center for Excellence in Palliative Care (M.J.H., N.A.), Department of Oncology, Oslo University Hospital-Ullevål, Oslo, Norway; European Palliative Care Research Center (M.J.H., P.M.F., D.F.H., J.H.L., S.K.), Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology and Trondheim University Hospital, Trondheim, Norway; Department of Public Health (P.M.F.), University of Aberdeen, Scotland, United Kingdom; Regional Center of Excellence for Palliative Care (D.F.H.), Western Norway, Haukeland University Hospital, Bergen, Norway; Palliative Care, Pain Therapy and Rehabilitation Unit (A.C.), Fondazione IRCCS, National Cancer Institute, Milan, Italy; Department of Palliative Medicine (G.W.H.), Bristol Haematology and Oncology Centre, Bristol, United Kingdom; National Resource Centre for Late Effects after Cancer Treatment (J.H.L.), Oslo University Hospital, Oslo, Norway; Division of Palliative Care Medicine (R.F.), University of Alberta, Edmonton, Alberta, Canada; Faculty of Medicine (N.A.), University of Oslo, Oslo, Norway; and Palliative Medicine Unit (S.K.), Department of Oncology, St. Olavs University Hospital, Trondheim, Norway

Abstract

Context. The use of unidimensional pain scales such as the Numerical Rating Scale (NRS), Verbal Rating Scale (VRS), or Visual Analogue Scale (VAS) is recommended for assessment of pain intensity (PI). A literature review of studies specifically comparing the NRS, VRS, and/or VAS for unidimensional self-report of PI was performed as part of the work of the European Palliative Care Research Collaborative on pain assessment.

Objectives. To investigate the use and performance of unidimensional pain scales, with specific emphasis on the NRSs.

Methods. A systematic search was performed, including citations through April 2010. All abstracts were evaluated by two persons according to specified criteria.

Results. Fifty-four of 239 papers were included. Postoperative PI was most frequently studied; six studies were in cancer. Eight versions of the NRS (NRS-6 to NRS-101) were used in 37 studies; a total of 41 NRSs were tested. Twenty-four different descriptors (15 for the NRSs) were used to anchor the extremes. When

Address correspondence to: Marianne Jensen Hjermstad, PhD, Department of Oncology, Regional Center for Excellence in Palliative Care, Oslo University Hospital, Ullevål, P.O. Box 4956, Nydalen, Oslo 0424, Norway. E-mail: m.j.hjermstad@medisin.uio.no

Accepted for publication: August 17, 2010.

0885-3924/$ - see front matter
doi:10.1016/j.jpainsymman.2010.08.016

compared with the VAS and VRS, NRSs had better compliance in 15 of 19 studies reporting this, and were the recommended tool in 11 studies on the basis of higher compliance rates, better responsiveness and ease of use, and good applicability relative to VAS/VRS. Twenty-nine studies gave no preference. Many studies showed wide distributions of NRS scores within each category of the VRSs. Overall, NRS and VAS scores corresponded, with a few exceptions of systematically higher VAS scores.

Conclusion. NRSs are applicable for unidimensional assessment of PI in most settings. Whether the variability in anchors and response options directly influences the numerical scores needs to be empirically tested. This will aid in the work toward a consensus-based, standardized measure. J Pain Symptom Manage 2011;41:1073–1093.

Key Words

Pain assessment, pain intensity, Numerical Rating Scale, Visual Analogue Scale, Verbal Rating Scale, review

Introduction

There is an extensive literature regarding the use of Numerical Rating Scales (NRSs), Verbal Rating Scales (VRSs), and Visual Analogue Scales (VASs) dating from the 1950s. Nearly all of this literature is from the social sciences, notably census and surveys, public opinion polls, and marketing research. Two particular themes emerge in this literature. The first is a focus on determining the optimal number of response options when using NRSs or VRSs; the second relates to the comparative value of VASs and NRSs.

In the area of cancer pain assessment, the main emphasis of most authors has been on comparing VAS scores, the most common measure for pain intensity (PI) in cancer research,[1,2] to the scores obtained on 10-step or 11-step NRSs (NRS-10 and NRS-11, respectively). Fewer papers seem to focus on comparisons involving VRSs. Despite the vast body of papers, few articles recommend the use of one scale over the other. Furthermore, the use of terms is often ambiguous. For the purpose of the present paper, we consistently use the abbreviations and terms outlined in the Appendix.

Two combined expert surveys/literature reviews[3,4] of cancer pain assessment agreed about the top three dimensions to include in a multidimensional assessment of cancer pain: *intensity*, *temporal pattern*, and *treatment-related factors* (exacerbation/pain relief). This is in line with other reports.[2,5–7] The recommendations from consensus meetings on cancer pain conclude that PI should be assessed by unidimensional scales.[2,5,8] Well-validated instruments, such as the Brief Pain Inventory[9] or the short-form McGill Pain Questionnaire,[10] are recommended for more comprehensive pain assessment. At present, there is no consensus concerning the terminology for temporal factors/breakthrough pain.[11–13]

The literature shows that NRSs provide sufficient discriminative power for chronic pain patients to describe their PI.[7] Invariably, authors either report that the NRS and VAS are equally efficient for assessment of cancer pain;[5] that the NRS may be preferred for assessment of PI in chronic nonmalignant pain in the clinic because of ease of use and standardized format;[7,14] and that the NRS is preferred by the majority of patients in different cultures.[5,15,16]

The European Palliative Care Research Collaborative (EPCRC) aims to design a computer-based tool for self-report of frequent cancer symptoms.[17] The first version, primarily focusing on pain, was used in the EPCRC-Computerized Symptom Assessment of pain, depression, and cachexia, an international data collection study including more than 1000 advanced cancer patients (www.epcrc.org).[17] The present systematic review is one step of the systematic, iterative process[18] toward the development of the computerized tool. We have reviewed studies with a specifically stated objective of comparing the use of

the NRS, VRS, and/or VAS for unidimensional self-report of PI, in cognitively intact adults. Because of the widespread use of the NRS for the assessment of PI in many disease groups and the fact that it constitutes a major part of more comprehensive assessment tools, specific focus was put on this scale. The major study aim, therefore, was to examine the results from comparative studies on unidimensional assessment of PI using the NRS, VRS, or VAS. The following points were investigated:

- What was the objective of comparing scales, and which scales were most frequently compared?
- Did compliance and usability differ between scales?
- Were different modes of scale administration compared, that is, plastic rulers, computers?
- Did the number of response options, verbal anchor descriptors, and time frames vary?
- What kind of statistics was used to report the results?
- Were patients' preferences for scales examined?
- Did the results from cancer patients differ from results in other patient groups?
- Were any of the scales recommended over the other(s) for research purposes and/or clinical use, and if so, why?

Methods

The literature search was performed in the following databases; MEDLINE (1950–2010, May week 2), PsycINFO (1806–2010, May week 3), and EMBASE (1980–2010 week 20) through OvidSP, and the Social Science Citation Index in Web of Science (1956–2010 update May 22) through ISI Web of Knowledge. Search-term *groups* representing 1) the NRS/VRS/VAS, 2) evaluation (including assessment and measurement), validation, comparison, clinimetry (including the clinimetric filter for PubMed/MEDLINE of Terwee et al.[19]), and 3) pain, were applied in all the combinations and adaptations according to the specific database and search engine requirements. Two limitations—"adults" and "English"—were used. The detailed search profiles can be obtained from the corresponding author on request.

All abstracts were read, and papers were selected for further reading if the abstract contained any information related to explicit comparisons of all or any two of the NRSs, VRSs, and VASs for assessment of PI. For inclusion in the present report, the publication had to meet the following criterion: A study with a specifically stated primary or secondary objective of comparing NRS/VRS/VAS for self-report of PI in adults.

Thus, case reports, editorials, letters, commentaries, reviews, and overviews were excluded, as were conference abstracts, and clinical studies simply using different scales for PI assessment, without aiming to compare the use and properties of scales. Pure validation studies of new tools or tools translated from the original language into another also were not included. Specific versions of the scales, that is, Faces Pain Scales (close to the VRSs) and the box variant with horizontal or vertical boxes for each value of the NRS, were not included, nor were studies comparing two types of the same scale, for example, the pen and paper version vs. the plastic version of the VAS, unless also comparing them to another scale (NRS/VRS).

In line with the study objectives and because of the plethora of pain assessment tools available, only the NRSs, VRSs, and VASs used for unidimensional assessment of PI in the included studies are described in detail; other pain tools are listed only in the tables.

The review process was conducted in two steps. First, two independent raters (M. J. H. and I. B.) examined all abstracts according to the eligibility criteria, consulting the full-text papers if in doubt about inclusion. In cases of uncertainty, a third independent classification was performed by a third reviewer (D. F. H.) and subsequently discussed. Second, all full-text papers of the selected abstracts were read to finally decide about inclusion. The *Related Articles* function in PubMed and the reference lists of the included papers were examined for additional relevant publications meeting the inclusion criteria.

Results

The searches produced 359 hits (MEDLINE 208, Embase 89, PsycINFO 30, SSCI 32) of

which 120 were duplicates. After screening the 239 abstracts, 59 were retained for further reading. The main reason for noninclusion at this stage (69%, 125/180) was that comparing scales for unidimensional assessment of PI was not a specifically stated study objective (Fig. 1). Reading of the 59 full-text articles resulted in another 13 papers being excluded, whereas eight additional papers were identified from the reference lists and/or the *Related Articles* function. Fifty-four papers were finally retained (Fig. 1).

Country of origin showed a wide spread: 13 papers (24%) were from the United States, six from the United Kingdom, three from Australia/New Zealand and Canada, respectively, two from Africa, one from Mexico, and one from China. The 28 remaining papers were from European countries other than the United Kingdom, including 12 from the Nordic countries. The majority, 35 (65%), were published in 2000 or later.

Objectives of Comparing Scales and Study Samples

Most of the 54 studies compared different pain rating scales to find the most applicable scale for clinical use in a given population, as reflected in the samples studied (Table 1). Thirteen studies evaluated postoperative PI,[20–32] one of these in the elderly.[31] Another eight were conducted in the emergency room/intensive care unit (ICU).[33–40] Six studies focused on cancer pain,[15,41–45] whereas one study used results from cancer patients for comparison.[46] Five studies examined pain in rheumatoid arthritis.[47–51] Four studies evaluated pain assessment in the elderly[52–55] in addition to the one mentioned above.[31] Three were experimental studies in volunteers, looking at ratings of pain that was induced by electric or heat/cold stimulations,[56–58] whereas the remaining 14 publications[59–72] encompassed different populations and various age spans.

Sample size varied from 12 to 1387. In 32 studies, repeated pain assessments were performed, in addition to one study with repeated assessment in a subsample only for test-retest purposes.[45] The different scales were presented to the patients in random order in 25 studies (one of which also used fixed order in a subsample for test-retest purposes) and in fixed order in 16; the order of presentation was not specified in 13 studies.

Overall, the VAS was by far the most frequently used scale. A total of 59 VASs were administered

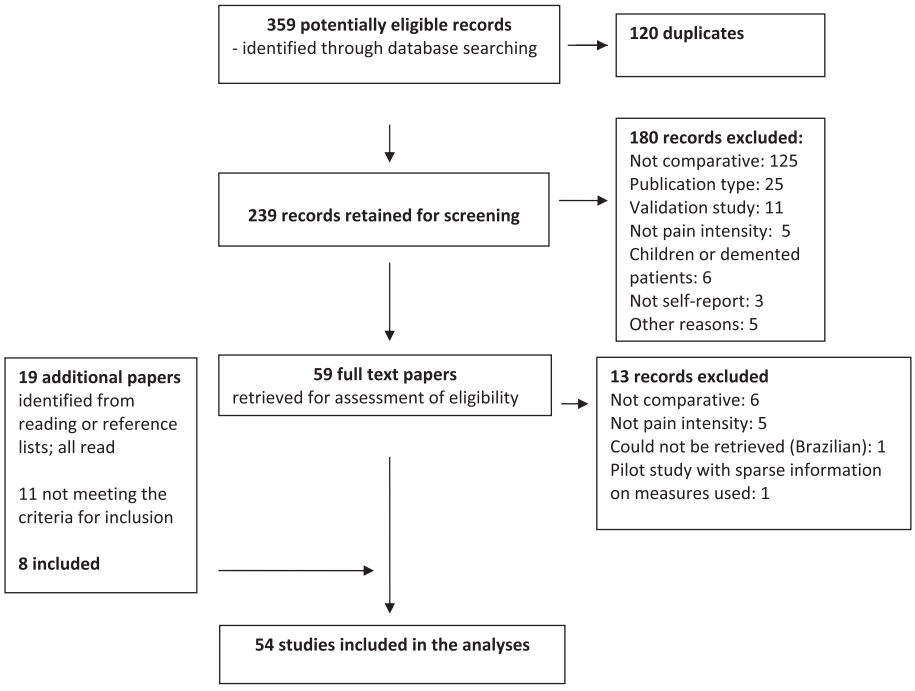

Fig. 1. Flowchart of the literature review and selection of papers.

Table 1

Overview of the 54 Studies Comparing Assessments of PI, Study Objectives, and Conclusions

First Author	Study Objectives	Population	Sample Size	Scales for PI + Other Scales[a]	Statistics[b]	Results	Conclusion and Preference for Scale Use[c]
Ahlers, 2008[33]	Compare scales in ICU patients, inter-rater reliability, compare scores of observers and patients	Critically ill ICU	113	NRS-11 VAS-100 mm, ruler +BPS	Kappa coefficients, Spearman rank	High reliability across NRS/VAS patients (0.84). Good inter-rater reliability. Observers often underestimate pain, especially with NRS ≥ 4.	No preference. Self-report important.
Akinpelu, 2002[20]	Study relationship between scales, influence of education	Women Caesarean section	37	VAS-100 mm VRS-11 +Box Numerical Scale	ANOVA, Pearson's	High correlation coefficients across scales, increasing with higher education.	No preference.
Banos, 1989[21]	Assess the usefulness of VAS for postoperative pain	Postsurgery	212	VAS-100 mm VRS-5	Spearman rank, Pearson's	High correlation VAS/VRS in patients. Lower VAS correlations between patients and observers with higher pain levels.	No preference. No differences, VAS valid.
Bergh, 2000[52]	Examine applicability of scales in older patients	Geriatric clinic	167	NRS-10 VAS-100 mm VRS-7 (GRS) +Verbal questions	Spearman rank, logistic regression, pair-wise comparisons	High correlation NRS/VAS/GRS. Lower accomplishment of scales with higher age, especially with VAS.	No preference, all scales useful, in-depth measures necessary with higher age.
Bergh, 2001[53]	Compare the verbally reported effect of analgesics with changes in pain scores	Geriatric clinic, nonpathological fractures	53	NRS-10 VAS-100 mm VRS-7 (GRS) +PRS	As above	High correlation NRS/VAS/GRS scales decreasing with age. Often in contrast with verbally reported analgesic effect.	No preference, all scales useful, must be supplemented with scales for pain relief.
Berthier, 1998[34]	Determine the most effective method for self-report of acute PI	ER, with/without trauma	290	NRS-11 VAS-100 mm, VRS-5, ruler	Pearson's, *t*-test, pair-wise comparisons, repeatability	NRS more reliable for trauma patients, equivalent to VAS without trauma. NRS/VAS better discriminant power for all the patients.	NRS preferable due to lower nonresponse rate.
Bolton, 1998[59]	Compare the responsiveness of scales	Chiropractic outpatients	79	NRS-11 VAS-100 mm VRS-5	Wilcoxon, Spearman rank	NRS most responsive for current pain. For usual pain, responsiveness of all measures was enhanced.	NRS preferable due to ease of use. Assessment of usual pain better than current pain.
Breivik, 2000[60]	Examine agreement, estimate differences in sensitivity between scales	Oral surgery outpatients	63	NRS-11 VAS-100 mm VRS-4	Stochastic simulation techniques	Large variability in VAS scores within each VRS-4 or NRS-11 category, between patients. Simulations showed VAS was more powerful than VRS.	No preference. Selection of NRS-11 or VAS to be based on subjective preferences.
Briggs, 1999[22]	Compare relationship between scales, examine characteristics of noncompliant patients	Orthopedic surgery, second postoperative night	417	VAS-100 mm VRS-5	Spearman rank	VAS and VRS scores highly correlated, but a wide range of VAS scores corresponding to each VRS category. Lower VAS completion rate with various impairments.	VRS preferred, due to compliance and ease of use.
Brunelli, 2010[45]	Compare NRS and VRS for breakthrough pain exacerbations	Advanced cancer patients	240	NRS-11 VRS-6	Percentage consistent ratings, weighted kappa	NRS higher discriminatory capability between background and peak PI, lower proportion of inconsistent ratings, higher reproducibility in PI exacerbations.	NRS preferred, due to higher discriminatory capability and reproducibility.

(*Continued*)

Table 1
Continued

First Author	Study Objectives	Population	Sample Size	Scales for PI + Other Scales[a]	Statistics[b]	Results	Conclusion and Preference for Scale Use[c]
Carpenter, 1995[41]	Compare pain and mathematical equivalence, examine nurses' responses to ratings	Cancer inpatients	50	NRS-6 VAS-100 mm		Lower VAS ratings than NRS. >3/4 ratings not mathematically equivalent. Nurses provided with fictitious patient scenarios did not provide the same pain medication for equivalent ratings.	No preference. Research into interpretation of scales necessary.
Clark, 2003[47]	Explore patient preferences for scales, validation	Rheumatoid arthritis outpatients	113	VAS-100 mm, ruler VRS-5	Pearson's, Spearman rank, χ^2, ICC	High correlation of scales (>0.79). 53% preferred the VRS, 28% the VAS, 19% had no preference. VRS viewed as easier to understand. Patients with lower education (<6 years) preferred the VRS.	No preference. Both scales valid, choice to be based on setting, clinical goal, level of education.
Collins, 1997[23]	Compare the equivalence of PI scores on scales	Postoperative pain, sampled from analgesia trials	1080	VAS-100 mm VRS-4	Mann-Whitney	VAS exceeding 30 mm corresponds to moderate pain or above on the VRS-4, including 85% of those reporting moderate pain.	No preference.
Daoust, 2008[35]	Recommend the best method for assessing PI in the ED	ED convenience sample	1176	VNRS-11 VAS-100 mm VAS-100 mm, ruler (VASp)	Bland-Altman, ICC	High correlation VNRS/VAS (0.88) and the VASp/VAS (0.92). VASp is probably valid to estimate acute pain, VNRS seems less valid due to wide limits of agreement and variable bias (mainly lower scores).	No preference, VASp valid.
De Conno, 1994[42]	Describe scaling properties, compare unidimensional and multidimensional pain scales	Chronic cancer pain patients	53	NRS-11 VAS-100 mm vertical VRS-6 +IPS, PRI, IRS	Principal factor analysis, logistic regression	High correlation between NRS, VRS, and VAS. They also were more strongly associated with the IRS than were with the PRI and IPS.	No preference.
DeLoach, 1998[24]	Examine the use of VAS in the early postoperative period	Postoperative patients	60	NRS-11 VAS-100 mm VRS-5[d] +DSST	Bland-Altman, ICC, Spearman rank, regression analyses	VAS valid in postoperative pain, correlated well with NRS, accuracy should be considered ±20 mm.	No preference, VAS useful.
Downie, 1978[48]	Investigate degree of correlation between pain scales	Two series of rheumatoid arthritis patients	1: 100 2: 104	NRS-11 VAS-100 mm horizontal and vertical VRS-4	Correlations, unspecified	Good correlation between scales, also with breaks between assessments. The NRS-11 performed better than the other scales.	NRS preferred due to lower measurement error.
Ekblom, 1988[36]	Compare pain scales before and after afferent stimulation/placebo	Acute orofacial pain patients, EU	80	NRS-101 VAS-100 mm VRS-6 +2 Modified NRS/VAS + pain relief scales	Pearson's, logistic regression, ANOVA	Except for VRS, significant correlation between scales at both assessments, good reliability. VRS changes did not correspond to equally large changes on the other scales.	No preference, but VRS did not perform well.

Fauconnier, 2009[37]	Compare methods for measuring pelvic PI	Consecutive sample, gynecologic EU	177	NRS-11 VAS-100 mm VRS-5 +2 Nonverbal pain indexes	Cronbach's alpha, Pearson's, factor analyses, ROC curve	Less missing data for NRS, VRS, and VAS than for the two behavioral scales, all methods sensitive to the pain physiology, location, severity.	No preference.
Ferraz, 1990[49]	Evaluate the reliability of three pain scales in literate and illiterate patients	Rheumathology outpatients	91	NRS-11 VAS-100 mm VRS-5	Student's *t*-test, Pearson's, Fisher's *Z*	NRS with highest reliability in both literate and illiterate patients, VAS more difficult to complete.	No preference.
Gagliese, 2005[25]	Compare feasibility and validity of scales for assessment of PI across the adult lifespan	Postoperative pain, older vs. younger patients	504	NRS-11 VAS-100 mm horizontal and vertical VRS-5[e] +Mc Gill Pain Questionnaire	χ^2, ANOVA	NRS was the preferred scale by patients, also showed low error rates, higher face, convergent, divergent, and criterion validity regardless of age. VAS difficult in the elderly.	NRS preferred, as properties were not age related.
Grotle, 2004[61]	Compare responsiveness of functional and pain scales in the clinical course of disease	Acute and chronic low back pain outpatients	104 50 acute 54 chronic	NRS-11 VAS-100 cm +4 Functional scales	K-S Lillefors, Student's *t*-test, standardized response means, ROC effect size	Both NRS and VAS appropriate, NRS significantly more responsive than VAS in the chronic pain group.	NRS preferred for chronic back pain, but both NRS and VAS valid.
Heikkinen, 2005[26]	Explore congruency of patients' and nurses' ratings, evaluate use of a pain tool in the recovery room	Postoperative pain	45	NRS-11 VAS-100 mm +Verbal assessment	Spearman rank, Pearson's, multiple regression analyses	Patients' ability to use different tools varied. Assessments correlated with each other and with nurses' estimations. Nurses both underestimated and overestimated patients' pain. Patients' verbal pain assessments varied widely.	No preference, not totally clear whether pain tools are usable in the recovery room; further research necessary.
Herr, 1993[54]	Determine relationship among measures, examine the ability to use tools correctly, determine tool preferences	Elderly with leg pain	Phase 1: 49 Phase 2: 31	VAS-100 mm VRS-6[e] NRS-20 VAS-100 mm horizontal and vertical +Pain thermometer	Spearman Brown, Tukey's post hoc, ANOVA	Phase 1: Higher correlation between tools when using same verbal anchors; Phase 2: VDS preferred overall, but had higher failure rates. VAS vertical preferred to VAS horizontal. All tools appropriate.	VAS may be preferred in research due to better sensitivity. Patients' preferences important in the clinic.
Herr, 2004[56]	Determine the psychometric properties of 5 pain scales in older and younger adults, examine preferences	Young and old volunteers, quasi-experimental (thermal stimuli)	175	NRS-21 VNS-11[f] VAS-100 mm vertical VRS-11 +FPS	Factor analyses, Cronbach's alpha, Pearson's, χ^2, ANOVA	All scales psychometrically sound, effective in discriminating different levels of pain. VDS was most sensitive and reliable in older. Low failure rates, except for the VAS. NRS preferred by patients.	VDS preferred, due to psychometric properties and patients' preference.
Herr, 2007[50]	Evaluate sensitivity and utility of scales in younger and older	Rheumathology patients, quasi-experimental	61	NRS-21 VNRS-11[f] VAS-100 mm +FPS, IPT	RR of failure to respond, χ^2, Poisson regression, GLM method for scale sensitivity	The IPT lowest failure rate, highest for the VNS and the VAS. Cognitive impairment significantly related to failure on VAS/NRS. All scales sensitive for PI changes. IPT, followed by the FPS most preferred by patients.	IPT preferred.

(*Continued*)

Table 1
Continued

First Author	Study Objectives	Population	Sample Size	Scales for PI + Other Scales[a]	Statistics[b]	Results	Conclusion and Preference for Scale Use[c]
Holdgate, 2003[38]	Test agreement between pain scales, calculate minimum clinically significant change	Convenience sample with acute pain, ED	79	VNRS-11 VAS-100 mm, ruler +VRS-4 pain relief	Mann-Whitney, Wilcoxon, Spearman rank, multiple regression	The VAS and VNRS highly correlated, but cannot be used interchangeably. Large differences between VNRS/VAS in paired observations, significantly higher scores on the VNRS.	No preference, VNRS useful.
Huber, 2007[46]	Determine if sensory or affective pain dimensions predicted unidimensional PI scores	General cancer, acute postoperative pain, chronic musculoskeletal pain, females	109	NRS-6 VAS-100 mm +MAPS	Student's *t*-test, MANOVA, Fisher's Z, Pearson's, multiple regression	Unidimensional PI scores mainly reflect sensory pain dimensions, supporting the discriminant validity of the NRS/VAS. Separate scales should be used to rate PI and emotions.	No preference.
Jensen, 1986[62]	Compare PI measures on selected criteria; correct response, relationship between scores	Chronic pain	75	NRS-101 VAS-100 mm VRS-4 VRS-5 +Behavioral Rating Scale (BRS-6) + Box-11	χ^2, correlation coefficients, principal factor analyses	High correlation across scales, similar rate of correct responses and utility, similar predictive validity. NRS easier to use and offers more response options.	NRS-101 may be preferred based on ease of use, sensitivity, and applicability across age group. All scales useful.
Jensen, 2002[27]	Compare the relative sensitivity of three outcome measures and one composite measure for pain relief in two RCTs	Postoperative pain	247	VAS-100 mm VRS-4 +VRS relief	ANOVA, *F* scores	Variability in the sensitivity of the pain ratings, VAS better than VRS. Pain relief was related yet distinct from changes in PI. The composite score did not increase the sensitivity of the pain assessment.	No preference, choice to be based on the specific dimension that relates to treatment.
Jones, 2007[55]	Examine the equivalency of pain ratings	Nursing home residents	135 + 135 validation sample	NRS-11 VRS-4[e] +FPS	Agreement percent, linear regression	Pain levels highly correlated, lower pain scores reported on the FPS, greater agreement with a modified FPS.	No preference.
Kenny, 2006[57]	Explore if people assign similar levels of numerical PI to verbal descriptors	Volunteers	207	VAS-100 mm VRS-15, self ranked	χ^2, correlations	High-correlation VRS/VAS, but respondents were idiosyncratic in the use of pain words/ descriptors.	No preference. Pain scales should supplement pain descriptions.
Kunst, 1996[28]	Compare pain ratings on VRS and VAS in a diamorphine study	Postoperative pain, lower abdominal gastrointestinal surgery	22	VAS-100 mm VRS-5	Variance/covariance models used for ordinal and interval data	VAS/VRS conveyed broadly similar information, however, VAS in individual patients varied about the patients' median.	No preference.
Langley, 1984[51]	Investigate relationship between scales and sensitivity to change	Rheumatology patients	37	VAS-21 cm VRS-7	Pearson's, Wilcoxon	Significant linear relationship, but better approximated by a curve. VAS better than VRS to detect PI changes, but warrants further investigations.	No preference.
Larroy, 2002[63]	Compare scales for assessment of menstrual pain	Healthy women	1387	NRS-11 VAS-100 mm	Spearman rank	Both scales useful, high correlation.	NRS preferred due to ease of use and interpretation.

Lasheen, 2009[44]	Evaluate fluctuation of symptoms, compare symptom scales	Cancer, hospice inpatients	125	VAS-100 mm VRS-4	χ^2, ANOVA, regression analyses	Significant differences between VRS categories and corresponding VAS scores, but overlap too wide to accurately assign cut-off points. VAS less reliable.	VRS may be better due to large variability of VAS.
Li, 2007[29]	Determine the psychometric properties and applicability of scales in China	Postoperative pain	173	NRS-11 VAS-100 mm VRS-6[e] +FPS revised	Spearman rank, ICC, ANOVA, McNemar, Bonferroni corrections	All four scales with good reliability/validity, high correlation, good sensitivity, all useful.	FPS preferred.
Loos, 2008[30]	Evaluate the optimal tool after hernia repair	Postoperative pain, outpatients	706	VAS-100 mm VRS-4	Pair-wise comparisons, kappa coefficients	Higher failure rates with VAS, not influenced by age. Overlapping VAS scores within each VRS category.	VRS preferred due to lower failure rates.
Lund, 2005[64]	Evaluate the quality of the intraindividual pain assessment and the equivalency of scale cut-offs	Musculoskeletal pain, outpatients	80	VAS-100 mm VRS-5	Pair-wise comparisons, coefficient of monotonic agreement	VAS/VRS not to be used interchangeably, low intrascale agreement, the meaning of the rated PI dependent on pain etiology. Probable underestimation of PI when the VAS was categorized. Overlapping VAS scores within each VRS category.	VRS may be preferred, but pain etiology should be considered.
Lundeberg, 2001[65]	Evaluate the intraindividual disagreement in pain ratings	Chronic pain patients	69	NRS-21 VAS-100 mm +Pain matcher for magnitude matching	Rank-order agreement coefficient, ROC curve	All tools reliable and responsive to pain relief, only random disagreement, Pain matcher may be useful.	No preference.
Magbagbeola, 2001[66]	Compare and validate pain measures in Nigeria	Patients referred to physiotherapy for painful conditions	100	VAS-10 cm VRS-4 VRS-5	Correlation coefficients	High correlation across scales, regardless of education. VAS/VRS can be used together with a good pain history.	No preference.
Marquie, 2008[39]	Investigate the use and correlation of two pain scales in French patients	Emergency inpatients with pain	198	VNRS-11 VAS-100 mm	Pearson's, Bland-Altman agreement	VAS/VNRS ratings highly correlated both for patients and physicians, VNRS recommended as the tool of choice in ED acute pain.	VNRS preferred due to ease of use.
Paice, 1997[15]	Investigate use and validity of VNRS-11 in cancer	Convenience sample, cancer pain	50	VNRS-11 VAS-100 mm VRS-5[g]	χ^2, Mann-Whitney, Spearman rank	High correlation of scales, lower compliance with VAS regardless of age, gender. VNRS preferred by patients.	VNRS preferred due to ease of use.
Pesonen, 2008[31]	Investigate feasibility of tools for assessment of acute postsurgical pain in elderly	Elderly inpatients with acute pain after cardiac surgery	160	VAS-100 mm VRS-5 +FPS-7, RWS	Student's *t*-test, Cochran, Fisher's exact, Spearman rank	Lower compliance on VAS and FPS. Pain assessment most reliable with VRS and RWS. VAS, FPS not ideal in patients > 75 years.	VRS preferred in the elderly, VAS unsuitable.
Peters, 2007[67]	Study the psychometric properties and patients' preferences	Chronic pain outpatients	338	VAS-100 mm horizontal and vertical VRS-6[e] +Box-11, Box-21	Factor analyses, multilevel logistic regression analyses, logistic regression	All scales valid, but more mistakes with increasing age, most on the VAS. Box scales most preferred, the VDS in the older. In mixed population, box scale is the method of choice.	Box-21 preferred.

(*Continued*)

Table 1
Continued

First Author	Study Objectives	Population	Sample Size	Scales for PI + Other Scales[a]	Statistics[b]	Results	Conclusion and Preference for Scale Use[c]
Price, 1994[68]	Examine and compare scale characteristics and ease of use	Orofacial pain and chronic pain outpatients	33	NRS-11 VAS-100 mm horizontal and vertical M-VAS-100 mm, ruler (mechanical)	Triangulation method, regression, Pearson's	High correlation between NRS/VAS/M-VAS, all can be used for PI assessment. Only M-VAS provides ratio scale measurement.	M-VAS may be preferred due to ease of use. Needs further investigation.
Rodriguez, 2004[43]	Compare the effectiveness of 3 tools for postoperative pain in older adults	H&N cancer patients, ≥55 years old, with communication impairment	35	NRS-11 VAS-100 mm + FPS	MANOVA	High correlation between tools, all appropriate in this population. NRS the preferred scale, VAS the least preferred.	NRS may be preferred based on patients' and nurses' views, but individual needs to be considered.
Seymour, 1982[32]	Examine sensitivity and reproducibility of scales, related to analgesic effect	Postoperative pain after dental surgery	12	NRS-11 VAS-100 mm VRS-4	Wilcoxon's	High correlation between scales, especially VAS/NRS. VAS most sensitive and discriminated better between small changes in PI.	VAS may be preferred due to better sensitivity.
Singer, 2001[40]	Compare acute pain ratings with one-week recall	Convenience sample of ED patients	50	VNRS-11[h] VNRS-101[h] VAS-100 mm	Linear regressions, Pearson's	High correlation between scales and between initial scores and recalled initial pain after one week.	No preference.
Skovlund, 1995[69]	Compare statistical power for treatment success/failure	Migraine patients, at the beginning and four hours after medication in acute attack	268	VAS-100 mm VRS-4	Stochastic simulation model, Wilcoxon's, C^2 test with Yats distribution	Similar reliability and power of VAS and VRS, both scales useful.	No preference.
Skovlund, 2005[70]	Compare the sensitivity of two common pain scales	Healthy individuals with pain from endoscopic screening	168	VAS-100 mm VRS-4	χ^2, Student's t-test, stochastic simulation model, two-sample method, Wilcoxon's	VAS consistently more sensitive.	VAS may be preferred in mild to moderate pain, in people with no impairment.
Svensson 2000[71]	PI, scale concordance, statistical modeling for research	Long-term undefined pain, prior to body awareness course	43	NRS-7 VAS-100 mm	Statistical modeling of distributions of paired assessments (details in paper)	A certain point on the VAS did not relate to a numerically labeled PI on the NRS. Continuous VAS/NRS offer a false impression of reliable measures expressed in millimeters or numerals.	VRS with clearly described response categories preferred for research.
Williams, 2000[72]	Examine patients' use, description, and interpretation	Chronic pain inpatients + volunteer sample	78	NRS-11 NRS-20 NRS-101 VAS-100 mm +Interviews	Descriptive statistics only	Anchor point seemed to affect use, ratings incorporate various dimensions of pain; a range of internal/external factors, not only PI.	No preference.

Yakut, 2003[58]	Assess reliability and validity of three pain scales	Volunteers, experimental pain induced by trigger pressure	51	NRS-11 VAS-100 mm VAS-100 mm reverse (RVAS)	Student's *t*-test, ICC, SEM, Pearson's	No difference in reliability between RVAS and VAS, equally efficient. RVAS slightly better with high pain. Replication in patients necessary.	No preference.

ANOVA = analysis of variance; PI = pain intensity; ICU = Intensive Care Unit; ED = emergency department; ER = emergency room; Box Numerical Scale = 10 vertical boxes, *no pain, worst pain*; ICC = intraclass correlation coefficient; BPS = Behavioral Pain Scale; GRS = Graphic Rating Scale; PRS = Pain Relief Scale; IPS = Integrated Pain Score; PRI = Pain Rating Index; IRS = Pain Relief Scale; DSST = Digit Symbol Substitution Test; FPS = Faces Pain Scale; IPT = Iowa Pain Thermometer; MAPS = Multidimensional Affect and Pain Survey; RWS = Red Wedge Scale; MANOVA = multivariate analysis of variance.

[a]To facilitate reading; VAS-10 cm and VAS-100 mm are described as VAS-100 mm, regardless of the description in the actual paper.

[b]Traditional descriptive statistics; mean, percent, median, etc. are not listed.

[c]Details presented for the NRS/VRS/VAS scales only.

[d]VRS named VPS in paper.

[e]VRS named VDS in paper.

[f]VNRS called VNS in paper.

[g]VRS named SDS in paper.

[h]Called NRS-10 and NRS-100 in paper.

in 52 of the 54 studies, relative to 39 VRSs in 37 studies. In seven studies,[25,42,48,54,56,67,68] a vertical version of the VAS was used, together with the traditional horizontal VAS in four of these. Traditional NRSs were included in 32 studies (33 NRSs), whereas the verbal version, the VNRS (see Appendix), was used in another five (eight VNRSs), plus in two of those that also used the NRS,[50,56] yielding a total of 41 NRS scales.

The NRSs/VNRSs were compared with the VAS in 16 studies, with the VRS in two, and with both the VAS and VRS in 18 studies; the VAS and VRS were compared in the remaining 18 studies. As indicated in Table 1, several other assessment tools for pain or other symptoms also were included in some studies.

The nomenclature used by authors was consistent for the NRSs/VNRSs, with one exception: the acronym VNS used for a VNRS in one paper. Full consistency was found for the VAS scales, although specific acronyms were used for the plastic or mechanical devices substituting for the traditional paper VAS in some studies. For the VRS scales however, four different abbreviations were used; Verbal Pain Scale, Verbal Descriptor Scale, Simple Descriptor Scale, and Graphic Rating Scale. Some of these variants had a number connected to each verbal descriptor. For consistency, Table 1 uses only the standard terms NRS, VRS, and VAS.

Compliance and Usability

When reported, better compliance was reported for the NRSs/VRSs relative to the other scales in 15 studies, whereas 16 studies did not provide any such information. Lower compliance on the VAS was found in nine studies, associated with higher age, degree of trauma, or other impairments. Compliance results were based on the number of patients who were able to perform the ratings, the number of correct answers, and error rates percentages. In some studies, test/retest scores and discriminant validity between patient groups also were used to indicate compliance.

Different Modes of Administration

Six studies used a plastic or mechanical VAS version with a moveable cursor along a line, with anchors at the extremes only, as a substitute for the traditional paper VAS.[33–35,38,47,68]

Four of these studies were performed in the emergency room or with ICU patients. The two studies using both the paper and the ruler version[35,68] concluded that the two versions correlated highly, and that the mechanical version seemed easier and more practical in the emergency room when compared with the VNRS.[35] The experimental study of induced pain/unpleasantness[68] used a mechanical VAS that also provided an option for judgment of ratios of perceived PI, which was regarded as a feasible method for research and clinical work.

One study compared the intraindividual variation in repeated scores on traditional tools (NRS/VAS) with electric skin stimulation as the matching stimulus,[65] and concluded that none of the methods demonstrated systematic disagreement. None of the identified studies aimed to compare electronic or web-based appliances (handheld devices, laptop computers, cell phones, etc.) with traditional paper and pencil versions for PI assessment.

Response Options, Anchor Descriptors, and Time Frames

The NRS-11/VNRS-11 was most frequently used ($n = 26$), but six other versions also were used: NRS-6,[41,46] NRS-7,[71] NRS-10,[52,53] NRS-20,[54] NRS-21,[50,56,65,72] and NRS-101.[36,40,62,72] One study[38] allowed the patients to give their score as half integers on the VNRS-11, which may then be regarded as a 21-point scale. One study used two NRSs (0−10 and 0−100) that were erroneously labeled as NRS-10 and NRS-100,[40] whereas one NRS-101 was used as a VNRS.[36]

Five different versions of VRS answer categories were used; 12 used a VRS with four response categories (VRS-4), 15 used a VRS-5, seven used a VRS-6, three used a VRS-7, and one used a VRS-11; one study used a 15-category version[57] where the patients were asked to assign their own verbal descriptors to the numbers between the two anchors "none" and "severe." All studies using VAS scales used the VAS-100 mm version, also labeled as VAS-10 cm.

As shown in Table 2, the descriptors used for the extremes varied, with 24 different adjectives being used. "No Pain" and "Worst Pain Imaginable" were most frequently used; the terminology was not given in five studies. Twenty-two studies used the same verbal anchors for all scales being compared (three scales or more in nine studies, two scales in 13 studies), and 14 studies used different descriptors for all scales being compared. Among the studies that used the same labels for two of three or more scales, the VRSs were most often labeled differently.

The exact wording of the probe questions that were used for PI assessment was not reported in all papers, nor were the time spans being covered. However, 36 studies specifically asked for "current pain," "present pain," or "pain right now," supplemented in seven studies with specified ratings of weakest, worst/strongest, recalled, anticipated, or average pain over different periods of time. Another seven papers did not specify the wording, but it was deduced from the objectives and patient samples that current pain was being evaluated. Other formats were PI at rest and when moving/coughing (2); maximal pain last hour (1); worst pain ever experienced (1); and average pain last week (3), last 24 hours (2), last night (1), and last month (1). One study supplemented the 24-hour PI rating with a rating of the most severe PI in the last 24 hours to specifically address pain exacerbations.[45]

Use of Statistics

Descriptive statistics were used in all the studies (not tabulated). Apart from the statistical modeling papers that used stochastic simulation techniques and other advanced statistics, the majority of studies used various forms of correlation statistics for comparing scale scores, inter-rater reliability, and evaluation of treatment effect, depending on primary study outcomes. However, confidence intervals for the differences between scales were rarely presented and intraclass correlation coefficients were used in five papers only. Most papers reported good correlation between scales (Table 1), particularly so between the NRS and VAS. In cases of discrepancy, the NRS scores were higher than the equivalent VAS scores, particularly so for the verbal NRSs.[38] One study found that more than 75% of the patients provided ratings that were not mathematically equivalent on NRS and VAS.[41]

Some studies reported a marked variation between numerical and verbal scales, but in

Table 2
Overview of Anchor Labels Used with the NRS/VNRS, VAS, and VRS

	NRS/VRNS	VRS	VAS
	37 Studies	37 Studies	52 Studies
	41 Scales	39 Scales	59 Scales
Wording of Anchor Labels	*n*	*n*	*n*
No pain, worst pain	1	3	5
No pain, worst pain possible	2	—	3
No pain, (the)[a] worst possible pain	3	3	8
No pain, worst pain imaginable	6	3	11
No pain, worst pain ever	1	1	3
No pain, pain cannot be worse	—	—	1
No pain, worst pain experienced	—	1	—
No pain (at all),[a] unbearable pain	4	4	5
No pain, pain as bad as it could be	4	—	4
No pain, very intense pain	—	—	1
No pain, the most intense pain imaginable	3	1	4
No (pain) (at all),[a] (severe) pain	2	10	3
No pain (at all),[a] very severe pain	—	2	—
No pain (at all),[a] the most severe pain possible	—	1	—
No pain, pain, which could not be more severe	—	1	—
No pain, the most severe pain you can possibly imagine	1	—	1
No pain sensation, the most intense pain sensation imaginable	1	—	3
No pain, maximum pain	4	—	3
No pain, maximal amount of pain	3	—	—
No pain, intolerable pain	1	—	—
No pain, excruciating pain	—	5	—
Mild, excruciating pain	—	1	—
No pain, horrible pain	—	1	—
Least possible pain, worst possible pain	—	—	1
Wording not specified in paper	4	1	3

[a]Indicates that the words in brackets were used in some tools, not in others.

different directions. One study reported that VAS scores above 30 mm corresponded to moderate pain or above on the VRS-4, thereby including 85% of those reporting moderate pain,[23] yet another study found that the stepwise change in the VRS did not correspond to equally large changes on the other scales,[36] and multiple studies found that there was a wide range of VAS and NRS scores within each VRS category[22,26,44,60,71] or that patients' own pain descriptors varied widely regardless of scale scores.[26] Two of the four papers used statistical modeling of data from various patient samples and reported that VASs were most sensitive to changes.[60,70]

Evaluation of Patient Preferences

Six studies examined patients' preferences for scales:[43,47,54,56,67,72] in rheumatoid arthritis (1), geriatric (2), chronic pain (2), and cancer (1) patients, respectively. All studies used a VAS and different VRSs, supplemented by the NRS-11, NRS-21, or NRS-101 in three studies, and by other scales (Table 1). Although patient preferences reflect the tools being used and the population under study, the VRS was preferred by the less educated[47] and the elderly,[54,67] and the NRS was the instrument of choice in an age-mixed population,[56] in chronic pain patients,[72] and in head-and-neck cancer patients.[43]

Two studies assessed patient preferences for different versions of the VAS scales. The elderly preferred the vertical to the horizontal version.[54] No preference was demonstrated for the traditional horizontal VAS over the reversed version with the "no pain" on the right side.[58]

Studies in Cancer Populations

Six studies were done in cancer patients, five in samples with mixed cancer diagnoses,[15,41,42,44,45] and one in head-and-neck cancer.[43] Four studies used NRSs/VRNSs and VASs, supplemented by VRSs in two,[15,42] a supplementary measure for PI in one,[43]

and the Italian McGill Pain Questionnaire for multidimensional pain assessment, a scale for pain relief, and an integrated PI and duration measure in one.[42] One study compared the VAS-100 mm and a four-point VRS;[44] another, the NRS-11 and VRS-6.[45]

Study objectives were to compare scales for clinical use with respect to scaling equivalence of the NRS-6 vs. the VAS-100 mm[41] or the VAS-100 mm vs. the VRS-4,[44] to examine the administration of the verbal NRS in general,[15] and in relation to communication impairment[43] (Table 1). One study compared unidimensional ratings of PI with multidimensional scales, including duration and relief,[42] whereas one study compared NRS and VRS for assessment of episodic pain exacerbation in chronic cancer pain.[45]

Although correlations across scales were high in all studies, the recommendations for use in cancer differed. NRS-11 was recommended in three studies based on results and ease of use,[15] patient preferences,[43] and better psychometric properties (lower inconsistency, better discriminatory power, and reproducibility).[45] One study found that the NRS-6 yielded lower within-patient scores than the VAS and that the scales should not be used interchangeably;[41] no specific recommendation for either scale (NRS/VAS/VRS) was given in the study comparing unidimensional and multidimensional scales.[42] One study in hospice patients concluded that the VAS-100 mm showed no superiority over the VRS in assessing fluctuating symptoms, that there were significant differences between VRS categories and corresponding VAS scores, and the VRS was more appropriate for symptom assessment in those with advanced disease. On the basis of the few studies in cancer, it cannot be concluded that results or recommendations differ from those in other populations.

Study Recommendations

The majority of papers, 29, did not conclude with a preference for one tool over the other(s) (Table 1). Three papers recommended tools other than the NRS/VRS/VAS. The NRS was considered superior in 11 studies,[15,25,34,39,43,45,48,59,61–63] primarily because of ease of use and high compliance, although some papers expressed a slight reservation, claiming that the tool may be best suited for a subset of the population only. Seven papers recommended using the VRS[22,30,31,44,56,64,71] for ease of use, low age-dependent failure rates, superior psychometric properties, and better responsiveness to fluctuating symptoms, although depending on the pain etiology.[64] Four papers recommended VAS as the preferred tool, also with some reservations.[32,54,68,70] Few papers specifically stated if the preference was for clinical use, but based on the study objectives, this was likely to be the case in most papers. Two of the statistical modeling papers specifically recommended VRS for research.[69,71] The arguments were that although the reliability and power of the VAS and VRS made them equally useful for clinical use, the psychometric properties of the VRS were better for research purposes,[69] and that numerical measures such as the NRS/VAS provide false impressions of being reliable measures.[71]

Discussion

The level of PI at the initial assessment has been shown to be a significant predictor of the complexity of cancer pain management and the time needed to obtain stable pain control.[73] PI is probably the most clinically relevant dimension of the pain experience regardless of disease. The overarching importance of this domain was accentuated in the present review, in that 89% of the identified studies were performed in populations other than cancer patients. According to expert surveys and consensus conferences,[3–5,8] PI should be assessed by unidimensional scales based on self-report. The importance of the latter was evidenced by the incongruence in some studies between patient and proxy ratings, with proxies underestimating high pain levels. Despite the apparent consensus on PI assessment, our review showed that PI is monitored by a wide variety of unidimensional scales. The differences were expressed by the number of response options, scales of variable lengths, different verbal descriptors, and the different time spans covered. Reviews also have shown that the development of new tools for various pain domains, including intensity, is a continuously ongoing process[3] that may further add to this variability.

The objectives in most of the reviewed papers were to find the most applicable scale in the population being studied, but only 25 papers concluded with a specific recommendation. This may be because the statistical methods and sample sizes were insufficient to detect significant differences. The use of correlation coefficients alone is misleading to decide whether one scale performs better than another,[74] and only a few studies provided sophisticated statistical methods. This is presumably because many of the studies were designed to test the applicability of the scale use, not psychometric performance. The latter is supported by the fact that most of the unidimensional scales performed reasonably well in all studies.

Although some studies examined the applicability of mechanical, ruler versions of the VASs, it was a little surprising that none of the identified studies compared computerized and paper versions of the different scales. The rapid development in handheld computer technology provides ample opportunities for self-report of symptoms in most settings and also has been shown as a feasible assessment method in patients with advanced cancer.[75,76] Advanced technology may increase the reliability of pain and symptom assessment, facilitate the transfer of information, and yield immediate scores that are readily available for clinical or research purposes. However, these methods do not enhance the validity and clinical utility of the assessments per se, which are dependent on the psychometric properties of the questions that are being presented to the patients.

The exact number of response options used in a scale is important. A scale with only two (e.g., pain/no pain) or three response options offers little opportunity for discrimination. Most of the reviewed papers used scale versions according to current recommendations, primarily NRS-11s, VRSs up to seven categories, and VAS-100 mm.[2,5,6] An overall conclusion from the general measurement literature is that there is relatively little gain in precision with more than seven options and hardly any above nine.[77,78]

The most frequently used scale in the reviewed studies was the VAS-100 mm, which is relatively seldom subject to variations in length. This scale potentially offers the greatest opportunities for discrimination, although in practice this is illusory if most respondents are unable to discriminate PI with precision beyond nine or 10 distinct levels. Only one study used a VRS with more than seven points. The NRS-11 has been shown to perform well for PI assessment over the central portion of the continuum (2–8)[79] and was the most used version of the NRSs. Four studies in this review used NRSs ranging from 0 to 100 (NRS-101). However, on inspection, these were actually presented as verbal scales, having the patient indicate a number between 0 and 100 rather than marking the appropriate number. Thus, there seems to be some ambiguity in recognizing the scales, calling for standardization. It also has been shown that patients tend to treat the NRS-101 scales as NRS-21 or NRS-11 scales.[7] The NRS-11 was the preferred scale in the few studies investigating patient preferences in line with previous reports,[77,78] but it should be noted that some VRS-6 scales may be scored as 0–2–4–6–8–10, thereby complying with the preference for the 0–10 scales.

As far as we know, no specific recommendations exist with respect to anchor labels, as evidenced by the 24 different descriptors used. Although most scales used "no pain" at the lower end, there were more variations at the upper extremes (Table 2), some directly implying a comparison with previous pain experiences ("worst pain experienced"). One study concluded that the anchor labels incorporated a range of personal values, not only a description of the PI domain.[80] Another study by the same author showed a lack of concordance between patients and of consistency within patients when they were completing VASs and NRSs by using their own descriptions and forced choices.[72] In our opinion, it seems likely that the labels influence the responses, maybe even more at the upper end of the scale than at the lower end, particularly so in different languages and cultures. However, to what extent and in which direction the actual scores are influenced, remains an empirical question that needs further investigation. Nevertheless, standardization with respect to anchor labels is warranted and the optimal description should be aimed for.

The compliance rates were surprisingly high in all studies reporting this, which may be viewed in context to the different settings.

Most papers used the term compliance, regardless of the different statistical or arithmetic methods used to examine this, which actually shows an inconsistent use of concepts, as it covered both compliance and usability in most studies.

It is highly likely that responding to a verbal NRS by saying a number between 0 and 10 or using a plastic ruler held by a nurse in the emergency room yields close to 100% compliance, whereas completion of pen and paper scales in elderly cancer patients is more complicated. It also may be that a selection bias comes into play in these relatively well-controlled studies, with specific emphasis on prompting as many patients as possible to answer. Overall, however, better compliance was reported for the NRS relative to the other scales, whereas the VAS seemed to be more complicated with higher error rates, especially in the elderly or cognitively impaired, as documented previously.[15,81] In relation to this, it may be regarded as a study limitation that our results were not differentiated between study populations, for example, the elderly, the cognitively or physically impaired, etc. Pain assessment in cognitively impaired adults, however, implies challenges other than the ones related to the actual pain tools and their content, in relation to mode of administration, visual limitations, layout, print size, actual size, and format of the paper tool, the need to go through the scores with the patient in more detail than with the cognitively intact, and so forth. Thus, it was decided to limit the literature search to those who were cognitively intact, and tabulate specific results from the elderly in some of the studies as appropriate.

The majority of the reviewed papers showed relatively consistent findings with respect to the correlation between scales, and when assessed, most coefficients between changes in scores over time were high, indicating that the scales tended to measure variations in the same direction.[45,48] However, several authors pointed to the variation in NRS scores within each bracket of the VRS and reported that ratings were not mathematically equivalent, which was taken as an indication of low interchangeability between scales by some authors.[26,38,45,64]

In addition, the expectation of obtaining direct equivalence between mathematically different scales may have been too optimistic. It is probably not realistic that patients provide equal values on scales with different layouts, response options, and anchor labels. The VRS pain assessment scales that are being considered in this article have response options chosen that are ordinal and generally assumed to have approximately equal intervals—although in the past this equal-interval assumption has rarely been tested and many statisticians would argue that ordinal methods of analysis should be applied.

The 54 papers included in the present review constituted only 23% of the papers originally identified by our search terms. The majority of the papers that were not included did not have a specific aim to compare scales. Thus, it was interesting to notice that several of these used both NRSs and VASs, and simply reported the mean values to conclude on the efficacy of analgesic treatment. It may be that some of these studies would have given additional information about the performance of the scales. Despite our thorough reading of abstracts and several articles from treatment studies, it can never be ruled out that we did not identify all relevant papers. However, the included papers cover a broad spectrum of studies comparing PI assessment tools, so we do not think that we failed to include important information on this subject.

Another limitation of this study is related to the heterogeneity of studies, samples, and variety of scales that may restrict the general relevance of our findings. Additionally, most papers based their conclusions and recommendations on descriptive or correlation statistics and were not designed to investigate the psychometric properties of the tools, which may be fundamentally different in, for example, chronic cancer pain vs. acute postoperative pain. However, the 11 studies specifically recommending the use of NRS were performed in different populations (acute, chronic, or cancer pain) and the recommendations were among others based on feasibility criteria, which are important features both for clinical use and research. The results from the studies in cancer patients did not differ from the other studies in any respect. Thus, the heterogeneity of the included studies describes very well this lack of standardization in PI assessment. PI as a dimension is paramount for

all pain management and should follow a standardized assessment methodology, regardless of the patient population or whether it is a part of a multidimensional tool or is used as an unidimensional scale.

This means that the same methodology (same scale, wording, time frame, and format) should be applied when assessing pain over time in the same patient population. Some of the reviewed studies conclude that the choice of scale may depend on factors such as patients' preferences and/or their level of cognitive functioning.[27,47,54,60] We only partly agree to this. In addition, it is obvious that certain population characteristics have to be considered, such as age, frailty, literacy level, and cognitive impairment. For example, the higher number of errors on the VAS with increasing age and other impairments makes this scale less applicable in the cognitively impaired, as documented in the literature.[2,5,82] This is also in line with a recent letter based on a study comparing NRS and VRS emphasizing the need to be selective in the use of scales for clinical use.[83] However, because the psychometric properties largely depend on certain basic characteristics, the selection of scales is better guided by specific consensus-based recommendations rather than left to the judgment of the individual clinicians. Furthermore, a standardization and consensus-based recommendation on the use of scales will facilitate the interpretation of results from studies and make comparisons across studies possible. It also may be necessary to distinguish between PI assessments for clinical use vs. research. We have reason to believe that the recommendations identified in the present review were mostly intended for clinical purposes, because only three papers specifically presented recommendations for research.

It is important to remember that a complex pain experience requires a multidimensional assessment, in line with the general recommendations in cancer.[2,5,8] However, directly combining PI scores with other measurements, such as pain interference scores, may be less relevant in clinical settings, as it may obscure the actual scores of each domain.[2,84] For most clinical purposes, PI is the key dimension guiding treatment[5] and it has been questioned whether cancer patients with multiple symptoms are able to discriminate between pain and other factors that interfere with their functioning.[85]

Although cancer pain may differ from acute, postoperative, and chronic pain in many respects, the common feature of any pain, regardless of cause, is its subjective nature, which makes it necessary to assess patients' pain perception in a standardized manner. In this respect, a promising initiative resulted from a consensus meeting on cancer pain assessment and classification in Milan in September 2009.[8] In relation to PI, the recommendation was that it should be measured by a 0–10 NRS with the standard endpoints "no pain" and "pain as bad as you can imagine," with clinically meaningful time frames. This consensus can be supported by the present review. Although the recommendation above was put forward for PI assessment in cancer specifically, it may well be applied to other populations as well. Key factors to remember in this respect in relation to the patient population are level of cognitive function, which may make a verbal NRS the instrument of choice, and the appropriate time frame for monitoring changes in PI over time.[8] We think it is time to welcome all consensus-based approaches that aim to standardize and facilitate the assessment of the subjective pain experience to improve pain management and promote research.

In conclusion, the results show that NRS-11, VRS-7, or VAS all work quite well. Thus, it is reasonable to say that the most important choice is not the type of scale per se, but the conditions related to its use, which include: a standardized choice of anchor descriptors, methods of administration, time frames, information related to the use of scales, interpretation of cut-offs and clinical significance, and the use of appropriate outcome measures and statistics in clinical trials.

We believe that all these areas can be improved by an international consensus process based on the evidence, which, in our opinion, should include, at least as a first step, perfecting and standardizing the use of NRS-11.

Disclosures and Acknowledgments

The EPCRC is funded by the European Commission's Sixth Framework Programme

(contract no LSHC-CT-2006-037777) with the overall aim to improve treatment of pain, depression, and fatigue through translation research. Core scientific group/work package leaders: Stein Kaasa (project coordinator), Frank Skorpen, Marianne Jensen Hjermstad, and Jon Håvard Loge, Norwegian University of Science and Technology (NTNU); Geoffrey Hanks, University of Bristol; Augusto Caraceni and Franco De Conno, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan; Irene Higginson, King's College London; Florian Strasser, Cantonal Hospital St. Gallen; Lukas Radbruch, RWTH Aachen University; Kenneth Fearon, University of Edinburgh; Hellmut Samonigg, Medical University of Graz; Ketil Bø, Trollhetta AS, Norway; Irene Rech-Weichselbraun, Bender MedSystems GmbH, Austria; Odd Erik Gundersen, Verdande Technology AS, Norway. Scientific advisory group: Neil Aaronson, The Netherlands Cancer Institute; Vickie Baracos and Robin Fainsinger, University of Alberta; Patrick C. Stone, St. George's University of London; Mari Lloyd-Williams, University of Liverpool. Project management: Stein Kaasa, Ola Dale, and Dagny F. Haugen, NTNU.

Authors have no conflicts of interest.

Special thanks to librarian Ingrid Riphagen at the Medical Library, NTNU, Trondheim, for the performance of efficient and precise literature searches, and to Irmelin Bergh (I. B.) at the Regional Center for Excellence in Palliative Care, Oslo University Hospital, for examination of abstracts.

References

1. Caraceni A, Brunelli C, Martini C, Zecca E, De Conno F. Cancer pain assessment in clinical trials. A review of the literature (1999-2002). J Pain Symptom Manage 2005;29:507–519.

2. Jensen MP. The validity and reliability of pain measures in adults with cancer. J Pain 2003;4:2–21.

3. Hjermstad MJ, Gibbins J, Haugen DF, et al. Pain assessment tools in palliative care: an urgent need for consensus. Palliat Med 2008;22:895–903.

4. Holen JC, Hjermstad MJ, Loge JH, et al. Pain assessment tools: is the content appropriate for use in palliative care? J Pain Symptom Manage 2006;32: 567–580.

5. 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: 239–255.

6. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.

7. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387–392.

8. Kaasa S, Apolone G, Klepstad P, Loge JH, Hjermstad MJ. Consensus expert conference on cancer pain and classification, the need for international consensus: working proposals on international standards. J Clin Oncol, in press.

9. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. Pain 1983;17:197–210.

10. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975;1: 277–299.

11. Hagen NA, Stiles C, Nekolaichuk C, et al. The Alberta Breakthrough Pain Assessment Tool for cancer patients: a validation study using a delphi process and patient think-aloud interviews. J Pain Symptom Manage 2008;35:136–152.

12. Haugen DF, Hjermstad MJ, Hagen NA, Caraceni A, Kaasa S. Assessment and classification of cancer breakthrough pain. A systematic literature review. Pain 2010;149:476–482.

13. Mercadante S, Radbruch L, Caraceni A, et al. Episodic (breakthrough) pain: consensus conference of an expert working group of the European Association for Palliative Care. Cancer 2002;94: 832–839.

14. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–158.

15. Paice JA, Cohen FL. Validity of a verbally administered numeric rating scale to measure cancer pain intensity. Cancer Nurs 1997;20:88–93.

16. Sze FK, Chung TK, Wong E, et al. Pain in Chinese cancer patients under palliative care. Palliat Med 1998;12:271–277.

17. European Palliative Care Research Collaborative (EPCRC). Available from http://www.epcrc.org. 2008.

18. Kaasa S, Loge JH, Fayers P, et al. Symptom assessment in palliative care: a need for international collaboration. J Clin Oncol 2008;26:3867–3873.

19. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009; 18:1115–1123.

20. Akinpelu AO, Olowe OO. Correlative study of 3 pain rating scales among obstetric patients. Afr J Med Med Sci 2002;31:123–126.

21. Banos JE, Bosch F, Canellas M, et al. Acceptability of Visual Analogue Scales in the clinical setting: a comparison with Verbal Rating Scales in postoperative pain. Methods Find Exp Clin Pharmacol 1989; 11:123–127.

22. 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;18:438–446.

23. Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain 1997;72:95–97.

24. DeLoach LJ, Higgins MS, Caplan AB, Stiff JL. The visual analog scale in the immediate postoperative period: intrasubject variability and correlation with a numeric scale. Anesth Analg 1998;86: 102–106.

25. 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;117:412–420.

26. Heikkinen K, Salantera S, Kettu M, Taittonen M. Prostatectomy patients' postoperative pain assessment in the recovery room. J Adv Nurs 2005;52: 592–600.

27. Jensen MP, Chen C, Brugger AM. Postsurgical pain outcome assessment. Pain 2002;99:101–109.

28. Kunst G, Chrubasik S, Black AM, et al. Patient-controlled epidural diamorphine for postoperative pain: verbal rating and visual analogue assessments of pain. Eur J Anaesthesiol 1996;13: 117–129.

29. Li L, Liu XQ, Herr K. Postoperative pain intensity assessment: a comparison of four scales in Chinese adults. Pain Med 2007;8:223–234.

30. Loos MJA, Houterman S, Scheltinga MRM, Roumen RMH. Evaluating postherniorrhaphy groin pain: Visual Analogue or Verbal Rating Scale? Hernia 2008;12:147–151.

31. Pesonen A, Suojaranta-Ylinen R, Tarkkila P, Rosenberg PH. Applicability of tools to assess pain in elderly patients after cardiac surgery. Acta Anaesthesiol Scand 2008;52:267–273.

32. Seymour RA. The use of pain scales in assessing the efficacy of analgesics in post-operative dental pain. Eur J Clin Pharmacol 1982;23:441–444.

33. Ahlers SJGM, van Gulik L, van der Veen AM, et al. Comparison of different pain scoring systems in critically ill patients in a general ICU. Crit Care 2008;12:R15.

34. Berthier F, Potel G, Leconte P, Touze MD, Baron D. Comparative study of methods of measuring acute pain intensity in an ED. Am J Emerg Med 1998;16:132–136.

35. Daoust R, Beaulieu P, Manzini C, Chauny JM, Lavigne G. Estimation of pain intensity in emergency medicine: a validation study. Pain 2008;138: 565–570.

36. Ekblom A, Hansson P. Pain intensity measurements in patients with acute pain receiving afferent stimulation. J Neurol Neurosurg Psychiatry 1988;51: 481–486.

37. Fauconnier A, Dallongeville E, Huchon C, et al. Measurement of acute pelvic pain intensity in gynecology: a comparison of five methods. Obstet Gynecol 2009;113:260–269. [see comment].

38. Holdgate A, Asha S, Craig J, Thompson J. Comparison of a verbal numeric rating scale with the visual analogue scale for the measurement of acute pain. Emerg Med 2003;15:441–446.

39. Marquie L, Duarte LR, Marine C, Lauque D, Sorum PC. How patients and physicians rate patients' pain in a French emergency department using a verbally administered numerical rating scale and a visual analog scale. Acute Pain 2008;10:31–37.

40. Singer AJ, Kowalska A, Thode HC Jr. Ability of patients to accurately recall the severity of acute painful events. Acad Emerg Med 2001;8:292–295.

41. Carpenter JS, Brockopp D. Comparison of patients' ratings and examination of nurses' responses to pain intensity rating scales. Cancer Nurs 1995;18: 292–298.

42. DeConno F, Caraceni A, Gamba A, et al. Pain measurement in cancer patients: a comparison of six methods. Pain 1994;57:161–166.

43. Rodriguez CS, McMillan S, Yarandi H. Pain measurement in older adults with head and neck cancer and communication impairments. Cancer Nurs 2004;27:425–433.

44. Lasheen W, Walsh D, Hauser K, Gutgsell T, Karafa MT. Symptom variability during repeated measurement among hospice patients with advanced cancer. Am J Hosp Palliat Care 2009;26: 368–375.

45. Brunelli C, Zecca E, Martini C, et al. Comparison of numerical and verbal rating scales to measure pain exacerbations in patients with chronic cancer pain. Health Qual Life Outcomes 2010;8:42.

46. Huber A, Suman AL, Rendo CA, et al. Dimensions of "unidimensional" ratings of pain and emotions in patients with chronic musculoskeletal pain. Pain 2007;130:216–224.

47. Clark P, Lavielle P, Martinez H, et al. Learning from pain scales: patient perspective. J Rheumatol 2003;30:1584–1588.

48. Downie WW, Leatham PA, Rhind VM, et al. Studies with pain rating scales. Ann Rheum Dis 1978;37:378–381.

49. Ferraz MB, Quaresma MR, Aquino LRL, et al. Reliability of pain scales in the assessment of literate

and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022–1024.

50. Herr K, Spratt KF, Garand L, Li L. Evaluation of the Iowa pain thermometer and other selected pain intensity scales in younger and older adult cohorts using controlled clinical pain: a preliminary study. Pain Med 2007;8:585–600.

51. Langley GB, Sheppeard H. Problems associated with pain measurement in arthritis: comparison of the visual analogue and verbal rating scales. Clin Exp Rheumatol 1984;2:231–234.

52. Bergh I, Sjostrom B, Oden A, Steen B. An application of pain rating scales in geriatric patients. Aging (Milano) 2000;12:380–387.

53. Bergh I, Sjostrom B, Oden A, Steen B. Assessing pain and pain relief in geriatric patients with nonpathological fractures with different rating scales. Aging (Milano) 2001;13:355–361.

54. Herr KA, Mobily PR. Comparison of selected pain assessment tools for use with the elderly. Appl Nurs Res 1993;6:39–46.

55. Jones KR, Vojir CP, Hutt E, Fink R. Determining mild, moderate, and severe pain equivalency across pain-intensity tools in nursing home residents. J Rehabil Res Dev 2007;44:305–314.

56. 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;20:207–219.

57. Kenny DT, Trevorrow T, Heard R, Faunce G. Communicating pain: do people share an understanding of the meaning of pain descriptors? Aust Psychol 2006;41:213–218.

58. Yakut E, Bayar B, Meric A, Bayar K, Yakut Y. Reliability and validity of reverse visual analog scale (right to left) in different intensity of pain. Pain Clin 2003;15:1–6.

59. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiol Ther 1998;21:1–7.

60. Breivik EK, Bjornsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. Clin J Pain 2000;16:22–28. [see comment].

61. Grotle M, Brox JI, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine (Phila Pa 1976) 2004;29: E492–E501.

62. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986;27:117–126.

63. Larroy C. Comparing visual-analog and numeric scales for assessing menstrual pain. Behav Med 2002;27:179–181.

64. Lund I, Lundeberg T, Sandberg L, et al. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross sectional description of pain etiology groups. BMC Med Res Methodol 2005;5:31.

65. Lundeberg T, Lund I, Dahlin L, et al. Reliability and responsiveness of three different pain assessments. J Rehabil Med 2001;33:279–283.

66. Magbagbeola JA. Pain assessment in Nigerians—visual analogue scale and verbal rating scale compared. West Afr J Med 2001;20:219–222.

67. Peters ML, Patijn J, Lame I. Pain assessment in younger and older pain patients: psychometric properties and patient preference of five commonly used measures of pain intensity. Pain Med 2007;8: 601–610.

68. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 1994;56:217–226.

69. Skovlund E, Flaten O. Response measures in the acute treatment of migraine. Cephalalgia 1995; 15:519–522.

70. Skovlund E, Bretthauer M, Grotmol T, Larsen IK, Hoff G. Sensitivity of pain rating scales in an endoscopy trial. Clin J Pain 2005;21:292–296.

71. Svensson E. Concordance between ratings using different scales for the same variable. Stat Med 2000; 19:3483–3496.

72. Williams AC, Davies HTO, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. Pain 2000;85:457–463.

73. Fainsinger RL, Fairchild A, Nekolaichuk C, et al. Is pain intensity a predictor of the complexity of cancer pain management? J Clin Oncol 2009;27: 585–590.

74. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.

75. Fyllingen EH, Oldervoll LM, Loge JH, et al. Computer-based assessment of symptoms and mobility in palliative care: benefits and challenges. J Pain Symptom Manage 2008;38:827–836.

76. Wilkie DJ, Kim YO, Suarez ML, et al. Extending computer technology to hospice research: interactive pentablet measurement of symptoms by hospice cancer patients in their homes. J Palliat Med 2009; 12:599–602.

77. Bandalos DL, Enders CK. The effect of nonnormality and number of response categories on reliability. Appl Meas Educ 1996;9:151–160.

78. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychol (Amst) 2000;104:1–15.

79. Lai JS, Dineen K, Reeve BB, et al. An item response theory-based pain item bank can enhance

measurement precision. J Pain Symptom Manage 2005;30:278–288.

80. Williams DA, Thorn BE. An empirical assessment of pain beliefs. Pain 1989;36:351–358.

81. Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: patient preference does not confound pain measurement. Pain 1981;10:241–248.

82. Chibnall JT, Tait RC. Pain assessment in cognitively impaired and unimpaired older adults: a comparison of four scales. Pain 2001;92:173–186.

83. Ripamonti CI, Brunelli C. Comparison between numerical rating scale and six-level verbal rating scale in cancer patients with pain: a preliminary report. Support Care Cancer 2009;17: 1433–1434.

84. Fayers PM, Hjermstad MJ, Klepstad P, et al. The dimensionality of pain: palliative and chronic pain patients differ in their reports of pain intensity and interference. Pain 2011 Mar 31. [Epub ahead of print].

85. Stenseth G, Bjørnnes M, Kaasa S, Klepstad P. Can cancer patients assess the influence of pain on functions? A randomised, controlled study of the pain interference items in the Brief Pain Inventory. BMC Palliat Care 2007;6:2.

Appendix

Descriptions and Abbreviations for NRS, VRS, and VAS Used in the Article

NRS	Numerical rating scale, commonly from 0 to 10 (NRS-11) or 1 to 10 (NRS-10). Usually, only the two extreme categories are labeled, for example, "No pain at all" and "Worst imaginable pain." NRS may be called a VNRS/VNS when the scale is explained or shown on paper to the patient, who responds by indicating a number.
VRS	Verbal rating scale. Ordered categorical scale, with each response option consisting of adjectives. For different levels of PI, "no pain," "mild pain," "moderate pain," "severe pain," "extreme pain," and the "most intense pain imaginable" form a six-category VRS scale (VRS-6). VRS scales are commonly of lengths four to seven. The adjectives are scored by assigning numbers (0–6) to each response option. The scale also may be called VPS (Verbal Pain Scale), VDS (Verbal Descriptor Scale), or SDS (Simple Descriptor Scale).
VAS	Visual analogue scale, usually 0–100, a straight line with the extreme categories labeled as for NRS. The distance measured from the "No pain" end to the patient's mark is the VAS score. Usually graduated with labeled marks indicating tens (10, 20, 30, etc.) and sometimes unlabeled marks for the units.