CrossMark

Response scale selection in adult pain measures: results from a literature review

Shima Safikhani[1], Katharine S. Gries[2], Jeremiah J. Trudeau[2], David Reasner[3], Katja Rüdell[4], Stephen Joel Coons[5], Elizabeth Nicole Bush[6], Jennifer Hanlon[3], Lucy Abraham[4] and Margaret Vernon[1*]

Abstract

Background: The purpose of this literature review was to examine the existing patient-reported outcome measurement literature to understand the empirical evidence supporting response scale selection in pain measurement for the adult population.

Methods: The search strategy involved a comprehensive, structured, literature review with multiple search objectives and search terms.

Results: The searched yielded 6918 abstracts which were reviewed against study criteria for eligibility across the adult pain objective. The review included 42 review articles, consensus guidelines, expert opinion pieces, and primary research articles providing insights into optimal response scale selection for pain assessment in the adult population. Based on the extensive and varied literature on pain assessments, the adult pain studies typically use simple response scales with single-item measures of pain—a numeric rating scale, visual analog scale, or verbal rating scale. Across 42 review articles, consensus guidelines, expert opinion pieces, and primary research articles, the NRS response scale was most often recommended in these guidance documents. When reviewing the empirical basis for these recommendations, we found that the NRS had slightly superior measurement properties (e.g., reliability, validity, responsiveness) across a wide variety of contexts of use as compared to other response scales.

Conclusions: Both empirical studies and review articles provide evidence that the 11-point NRS is likely the optimal response scale to evaluate pain among adult patients without cognitive impairment.

Keywords: Pain, Outcome measurement, Response options, Rating scales, Response scales

Background

Patient-reported outcome (PRO) instruments are important for assessing the symptoms and/or impact of symptoms on patients' lives [1–6]. Part of developing or adapting a PRO instrument includes the selection of a response scale (e.g., verbal rating scale [VRS], numeric rating scale [NRS], visual analogue scale [VAS]). Despite the importance of response scale selection for PRO instruments, there seems to be little empirical basis for the type of response scale selected, the response options, the visual orientation of the scale (i.e., vertical vs. horizontal), or the response scale verbal anchors. When considering a response scale for a PRO instrument, it is important to consider the context of use including

population of interest, therapeutic area, and study implementation. Such considerations are central because response scales contribute to the precision as well as the performance of an instrument in the clinical trial setting, such as the ability to detect changes with treatment [1, 7].

Response scale selection is a critical aspect of PRO instrument development and has broad downstream implications for the usability of the measure from the patient perspective, the level of precision with which the construct of interest is measured, and the quantitative properties of the outcome score including range, standard deviation, scoring, and score interpretation guidelines, as well as the responsiveness of the measure to detect change [8]. Additional complicating factors such as placement and exact wording of response anchors, cultural comparability/translatability of the format and wording, and ability to migrate the scale to various

* Correspondence: margaret.vernon@evidera.com
[1]Evidera, 7101 Wisconsin Ave., Suite 1400, Bethesda, MD 20814, USA
Full list of author information is available at the end of the article

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 2 of 9

modes of data collection (paper, electronic) should also be considered.

Pain is subjective in nature and is measured by patient report of intensity, among other subjective qualities. There is an abundance of existing literature on pain measurement, from expert opinions to consensus guidelines to empirical studies, all of which compare the suitability of response scale types within a particular context of use. In the PRO assessment field, 5- and 7-point VRS, 11-point NRS (particularly recommended for use in pain measurement but used in other areas as well [9]), and 10 cm/100 mm VAS are commonly used for adult assessments. VAS, NRS, and VRS response scales are generally reliable and valid, and are usually part of primary outcome measure(s) in clinical trials of chronic pain treatments. For these reasons, it is important to consider the empirical evidence as well as precedent, especially when considering the development of a new instrument.

Although response scales in pain measurement are well-explored, the existing literature should be further compared and analyzed for key themes and recommendations. Existing Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) guidelines and review articles indicate the NRS as the gold standard in measurement of pain intensity [9, 10]; however, much of the key research was published prior to the 2009 issuance of the Food and Drug Administration's (FDA) Guidance for Industry titled "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims" [1]. As such, a review to determine the state of the evidence was necessary in order to guide the ongoing research at the Critical Path Institute's PRO Consortium [11], as well as to inform the PRO assessment field during a period of expanding instrument development. The purpose of this paper is to describe a comprehensive review of the scientific literature aimed at identifying response scale options common in pain measures for the adult population, to summarize the empirical evidence in support of response scale selection, and to provide recommendations for this context of use.

Methods
Search procedure, terms, and strategy
The literature review was part of a larger study to summarize the available empirical evidence to support response scale selection during the development of new PRO instruments. The targeted search strategy included formal guidelines or review articles on the selection of response scales and response scale methodology (not specific to PRO instruments) and evidence on the selection of response scales for use in PRO instruments. Evidence was assembled and collated based on pre-determined categories such as selection of response scales for use in PRO

instruments based on therapeutic area (e.g. pain). This paper focuses on the identification and review of literature that addressed response scale selection and use for PRO-based pain assessment in adults. To provide a comprehensive review, references included review articles, consensus guidelines, and expert opinion pieces as well as primary research studies (from the last 10 years) on response scale choice for pain assessment. The search database included Embase and MEDLINE (2004–2014; only articles published in English) and presentation abstracts from two recent annual meetings/conferences of the International Society for Pharmacoeconomics and Outcomes Research (2013–2014) and International Society for Quality of Life Research (2012–2013). The types of included references were PRO instrument articles with information on response scale selection in adults. Additional supplementary searches (e.g., reviewing reference lists of articles that met the initial search criteria) were also conducted to ensure no relevant publications were missed. These supplementary sources were not limited by publication date, and included the reference lists of key articles, publications not included in the search databases, and websites for major PRO-related working groups and consortia (e.g., PROMIS, NIH Toolbox, Medical Outcomes Study, Neuro-QoL, ASEQ-ME, EORTC, EuroQol Group, and FACIT Measurement System).

For abstracts and full-text articles, the search results were examined by two independent reviewers to determine whether the article should be included in the review; in the case of non-agreement, a third senior reviewer made the final judgement. Articles were excluded if they provided no direct or indirect evidence relevant to the search objectives, were not applicable to PRO instrument development, or addressed a therapeutic area not pre-specified for inclusion. Once articles fitting the search criteria were identified, the relevant data were extracted and summarized. The extraction tables included data on the study objective, study design, study population, therapeutic area, name of PRO instrument, type of response scale, and empirical evidence for response scale selection. To further characterize the quality of included articles, several different efforts were employed. First, included articles were categorized as providing either direct evidence or indirect evidence. Direct evidence was defined as evidence that provided a direct answer to the research question of interest. For example, direct evidence articles compared empirically the relative robustness or merits of two different response scale types within the same study/population. Indirect evidence was defined as important, relevant evidence that should be considered in the review and overall conclusions, but that does not directly answer the research question or hypothesis. Second, an assessment was made of the quality of the data presented and the strength of results and recommendations for each article included in the review. Each article was

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 3 of 9

assigned a grade based on the type of article and strength of the data, from A (primary research; reflects the strongest empirical evidence for response scale recommendation) to D (review or expert opinion; reflects the weakest empirically-based evidence). In summary, the methods were comprehensive and covered several search objectives; the detailed literature review methods and processes are summarized elsewhere [12].

Results

Our review of the literature providing an empirical basis for response scale selection yielded pain (adult population) as the most prevalent area (Gries K, Safikhani S, Pease S, Harrington M, Rudell K, Berry P, Crescioni M, Patel M: Comprehensive literature review to characterize response scale types in patient-reported outcome measures, Unpublished observations). As a result of the abstract review, we identified 18 review articles/consensus guidelines/expert opinion pieces and 38 primary research studies for full text review, providing insights into the optimal response scale selection for pain assessment (Fig. 1). Further examination of the full text articles narrowed these down to 14 and 29, respectively.

Response scale types used in adult pain studies

In the adult population, there were 14 review articles/consensus guidelines/expert opinion pieces and 29 primary research studies, providing insights into optimal response scale choice for pain assessment. There were seven individual response scales identified in this comprehensive review across all search objectives (i.e., VAS, VRS, NRS, Faces, other graphical, binary, and Likert or Likert-type scales). Based on review, it appears that adult pain studies have historically used simple single-item measures of pain with a VAS, NRS, or VRS as their primary outcome, albeit sometimes in the context of a multi-item instrument (e.g., the Brief Pain Inventory). There was no widely accepted standard for clinical pain assessment that would facilitate the comparison of response scale performance across trials (Table 1). Further descriptions of the response scale types are summarized elsewhere (Gries K, Safikhani S, Pease S, Harrington M, Rudell K, Berry P, Crescioni M, Patel M: Comprehensive literature review to characterize response scale types in patient-reported outcome measures, Unpublished observations).

Summary of key evidence regarding numeric rating scales in the adult pain population

In the review of the 29 direct empirical evidence, we identified 11 articles that directly compared NRS with other type of response scales in either a validation study or a direct methodological comparison (Table 2). The following observations regarding the NRS suggest that the NRS is potentially superior to the VRS or VAS. The NRS was preferred over the VAS because it was easier to administer and score [13]; was preferred by patients [14]; was found to be more responsive [15]; and had both higher patient acceptability and better psychometric properties [16]. The NRS was found to be more responsive than a six-point VRS [17]; the most feasible and discriminative self-report scale as compared to a VAS or 5-point VRS [18]; and more sensitive and discriminative than a binary scale [19]. One study that compared an NRS to a *Faces* scale (a type of graphical scale that uses photographs or pictures showing a continuum of facial expressions) found both scales to be equally well understood [20], although a population of Swahili-speaking patients preferred the Faces scale over the NRS [21].
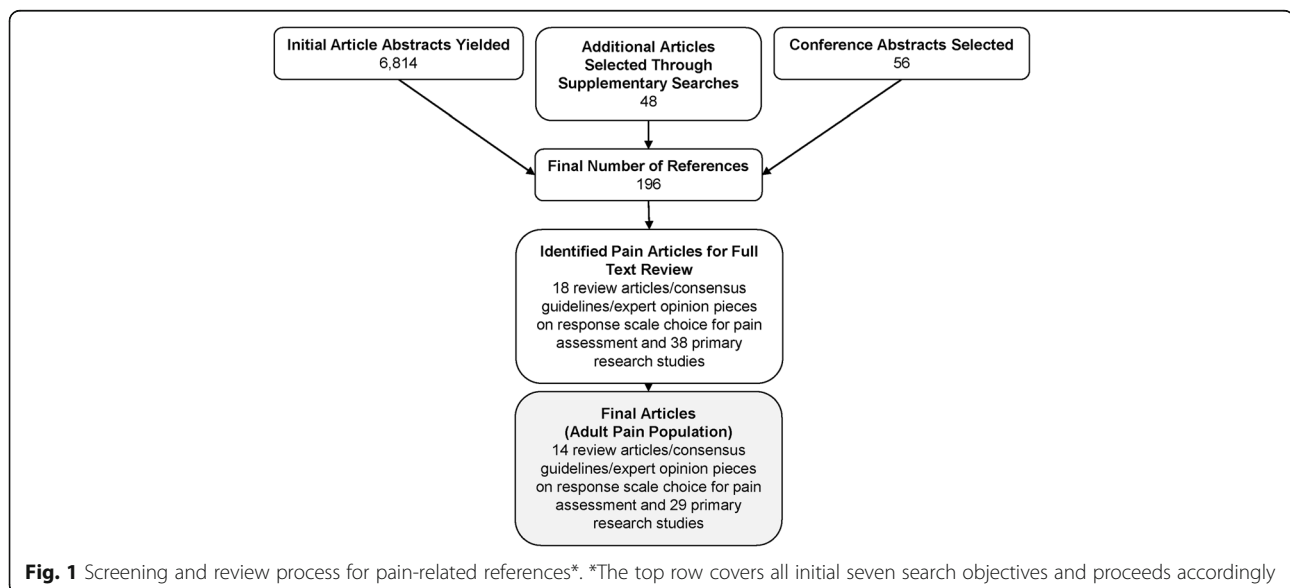

Fig. 1 Screening and review process for pain-related references*. *The top row covers all initial seven search objectives and proceeds accordingly

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 4 of 9

Table 1 Summary of key evidence to support response scale selection in pain

Reference	Article type, evidence type [a], and grade [b]	Article recommendations and rationale
Ong and Seymour 2004 [35]	Expert opinion, Direct, D	The VAS is simple, widely used, easily understood by most patients, and has strong validity. In general, the VAS is probably the most reliable and sensitive tool for measuring pain.
Dworkin et al. 2005 [9]	Consensus guideline, Direct, D	VRS and NRS are preferred over VAS by patients. VAS measures typically have greater amounts of missing data and are more difficult for older patients and patients taking opioids to understand. NRS ratings may be difficult in the presence of cognitive impairment; VRS might be easier in these circumstances. An 11-point NRS measure of pain intensity is recommended with 0 meaning "no pain" and 10 meaning "pain as bad as you can imagine." An additional pain intensity VRS can be considered (none, mild, moderate, severe).
Williamson and Hoggart 2005 [10]	Review article, Direct, B	All three of the pain-rating scales explored in this review are reliable and valid (VAS, VRS, and NRS). As a tool for pain assessment the NRS is probably more useful than the VRS or the VAS.
Schofield et al. 2006 [36]	Review article, Direct, B	Among fifteen reviewed papers of pain tools in the adult population, one examined five different scales across a range of care home settings in the UK. It suggests that VRS was the most successful, followed by the NRS.
Litcher-Kelly et al. 2007 [37]	Review article, Direct, B	Authors researched which pain assessments were most frequently used in clinical trials. The most frequently used assessments were the single-item VAS and the NRS.
Phillips 2007 [38]	Review article, Direct, B	No recommendation provided: Tailoring the pain assessment to the individual patient and knowing when and how to use different assessment tools will substantially increase the probability of a comprehensive pain assessment. Combining the appropriate tools to identify pain intensity, location, and pain behavior is an appropriate strategy to improve assessment outcome.
Khorsan et al. 2008 [31]	Review article, Direct, B	The NRS provides intensity estimates relatively quickly, is highly patient-centered, and has the most value when looking at change within individuals. The simplest and least time-consuming measure is the NRS.
Grimmer-Somers et al. 2009 [28]	Review article, Direct, B	There is considerable detail regarding the psychometric properties of pain severity scales. VAS, VRS, and NRS are useful for quick initial assessment of one pain dimension.
Hjermstad et al. 2011 [29]	Review article, Direct, B	NRS-11, VRS-7, or VAS work well and are applicable in pain intensity assessment. These response scale types exhibit ease of use, strong psychometric properties, work well in clinical use, and/or are sensitive to symptoms.
Vela et al. 2011 [32]	Expert opinion, Direct, D	We suggest using the NRS because of its ease of use.
Cook et al. 2013 [30]	Review article, Direct, B	A 0–10 point NRS ("no pain" to "worst imaginable pain") was chosen for the

Table 1 Summary of key evidence to support response scale selection in pain *(Continued)*

Reference	Article type, evidence type [a], and grade [b]	Article recommendations and rationale
		National Institutes of Health (NIH) Toolbox to measure pain intensity in adults because it is commonly used in clinical studies and has strong validity.
Riddle 2013 [39]	Review article, Direct, B	The McGill pain questionnaire and the pain VAS are the two most commonly used pain instruments in the clinical setting. Previous research has shown them to be valid and reliable in measuring acute pain.
Wolters et al. 2013 [33]	Review article, Direct, B	Our extensive review of pain intensity measures has led us to suggest that the NRS-11 be used to assess this endpoint domain in neurofibromatosis trials. It is brief, reliable, valid, and widely used. Additionally, it is highly recommended in clinical trials from pain experts and other consensus groups, is widely used in research and has strong reliability, validity, sensitivity to change, and feasibility in ages 8 years and older.
Green et al. 2014 [40]	Review article, Direct, B	The VAS-Usual pain and VAS-Worst pain were moderately correlated ($r = 0.63$), providing evidence of their structural validity.

[a] Direct evidence: Primary research that compares different response scales within study. Indirect evidence: Review or expert opinion based on empirical evidence

[b] Grade Key: A) Primary research; compares different response scales within the study; B) Review or expert opinion; based on an empirical evidence base; C) Primary research; evaluates a single response scale type within the study; and D) Review or expert opinion, based on expert consensus, convention or historical experience

Summary of key evidence regarding visual analog scales in the adult pain population

Seven studies directly compared the use of the VAS with other response scales. The following observations were made to support the use of the VAS. The VAS was found to be more sensitive than a four-point VRS [22]; a VAS with intermediate verbal anchors enhanced scale comprehension compared to a traditional VAS [15]; and the VAS was found to be equally well accepted and understood as a Faces scale [23]. One study found that a four-point VRS had lower scale failure rates (i.e., one or both tests were not completed correctly), than a VAS [24]. Table 3 summarizes the key evidence regarding VAS in the assessment of pain for adults.

Summary of key evidence regarding verbal rating scales in the adult pain population

Three studies provided empirical evidence regarding the direct comparison of the VRS with the NRS and VAS. The following observations were made to support the use of the VRS. In 25 patients with chronic low back pain, intra-rater correlations between a VRS and VAS showed no significant difference between raters, thus demonstrating appropriate reliability [25]. A five-point

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 5 of 9

Table 2 Summary of key evidence to support the numeric rating scale (NRS) in pain

Reference	Study type, evidence type [a], and grade [b]	Article recommendations and rationale
Grotle et al. 2004 [41]	Cross-sectional study, Direct, A	The NRS showed statistically higher standardized response mean SRM when compared with the VAS. The VAS format is difficult for some patients and time-consuming to score. The NRS is easier to understand and quicker to score.
Gagliese et al. 2005 [16]	Prospective study, Direct, A	The NRS should be the first choice across age groups. Compared to the other scales, it had low error rates, high face validity, and high convergent, divergent, construct and criterion validity. Results suggest that use of the VAS should be discouraged, unless practice is provided to patients.
Jackson et al. 2006 [14]	Instrument development and/or validation study, Direct, A	Most participants preferred using an NRS to the continuous VAS. Some found numbers easier to use whereas others found the scale of pain intensity (referred to by authors as the SPIN) more helpful.
Ritter et al. 2006 [13]	Cross-sectional study, Direct, A	The NRS (with graphical component) is a valid measure and was as successful as the VAS in measuring the underlying pain variable. The NRS was easier to administer and code than the VAS.
Chanques et al. 2010 [18]	Prospective study, Direct, A	The NRS should be the tool of choice for the intensive care unit (ICU) setting, because it is the most feasible and discriminative self-report scale for measuring critically ill patients' pain intensity.
Huang et al. 2012 [21]	Instrument development and/or validation study, Direct, A	The Faces scale and NRS were adapted and translated for a population of Swahili-speaking patients in western Kenya, and demonstrated the face validity, acceptability, and field-readiness of these scales through cognitive interviewing of hospitalized patients. In this population the Faces scale was preferred over the NRS.
Van Dijk et al. 2012 [42]	Prospective study, Direct, A	A lack of agreement is found between the patients and Acute Pain Nurses on what constitutes 'bearable' pain in relation to the reported NRS score. These findings suggest variable interpretation of scores on an NRS.

Table 2 Summary of key evidence to support the numeric rating scale (NRS) in pain *(Continued)*

Reference	Study type, evidence type [a], and grade [b]	Article recommendations and rationale
Chien et al. 2013 [17]	Instrument development and/or validation study, Direct, A	The NRS is potentially more sensitive to clinical changes in comparison to the VRS
Huang et al. 2013 [20]	Prospective study, Direct, A	The Faces scale and the NRS were easily understood and well accepted by participants and should be implemented for daily use in the inpatient setting in order to gauge patients' pain and response to pain treatment.
Rothaug et al. 2013 [19]	Prospective study, Direct, A	Binary answer format was proven to be a practical alternative to the NRS format for a screening instrument, however may not be as sensitive to discriminate levels of pain intensity as compared to the NRS.
Gonzalez-Fernandez et al. 2014 [15]	Cross-sectional study, Direct, A	The NRS (referred to in the article as general Labeled Magnitude Scale; essentially a VAS with the addition of numbers along the scale) has the potential to replace the VAS in the measurement of pain intensity in the clinical setting. As a minimum it may serve as an important tool in the management of patients with chronic pain.

[a] Direct evidence: Primary research that compares different response scales within study. Indirect evidence: Review or expert opinion based on empirical evidence

[b] Grade Key: A) Primary research; compares different response scales within the study; B) Review or expert opinion; based on an empirical evidence base; C) Primary research; evaluates a single response scale type within the study; and D) Review or expert opinion, based on expert consensus, convention or historical experience

VRS and VAS were administered in a cross-sectional study of patients with chronic, nociceptive, and neuropathic pain [26]. The VAS cut-off positions relative to the VRS labels and non-linear properties indicated different meaning of the rated pain intensity [26]. The VRS, NRS, and VAS were all sensitive to change for pain assessment, but the VAS was more difficult for patients in the study to understand [27]. The VRS showed substantial discrimination between pain words and was not dependent on the level of education of the patients [27]. Table 4 summarizes the key evidence regarding VRS in the assessment of pain for adults.

Summary of results in pain

Several review articles support the usefulness, reliability, and validity of each of the scale types (i.e., the VAS, VRS,

Safikhani et al. Journal of Patient-Reported Outcomes (2018) 2:40

Page 6 of 9

Table 3 Summary of key evidence to support the visual analog scale (VAS) in pain

Reference	Study type, evidence type [a], and grade [b]	Article recommendations and rationale
Skovlund et al. 2005 [22]	Cross-sectional study, Direct, A	The VAS is consistently more sensitive than the four-point VRS.
Chan et al. 2005 [43]	Prospective study, Direct, C	The VAS as a numeric measure of pain intensity allowed for documentation of pain over time and assessment of the relationship between pain and disability and function post-hand injury.
Loos et al. 2008 [24]	Cross-sectional study, Direct, A	Because of lower scale failure rates and overlapping VAS scores per VRS category, the VRS should be favored over the VAS in future post-herniorrhaphy pain assessment.
Boomershine et al. 2011 [44]	Prospective study, Direct, C	The VAS is simple to score for longer instruments and has been shown here to be valid, sensitive to change, and able to identify with score cut offs patients with clinically significant symptoms.
Bergh et al. 2011 [45]	Clinical study, Direct, A	The VAS is more sensitive than the Pain Matchers (PM) for recording changes in pain intensity when assessing the effects of treatment on labor pain.
Dogan et al. 2012 [23]	Prospective study, Direct, A	No patients had difficulty completing either pain scale (VAS versus Faces scale) (there were illiterate patients in the sample).
Gonzalez-Fernandez et al. 2014 [15]	Cross-sectional study, Direct, A	The VAS (with some intermediate verbal anchors) has the potential to replace the VAS in the measurement of pain intensity in the clinical setting.

[a] Direct evidence: Primary research that compares different response scales within study. Indirect evidence: Review or expert opinion based on empirical evidence
[b] Grade Key: A) Primary research; compares different response scales within the study; B) Review or expert opinion; based on an empirical evidence base; C) Primary research; evaluates a single response scale type within the study; and D) Review or expert opinion, based on expert consensus, convention, or historical experience

Table 4 Summary of key evidence to support the verbal rating scale (VRS) in pain

Reference	Study type, evidence type [a], and grade [b]	Article recommendations and rationale
Olaogun et al. 2004 [25]	Cross-sectional study, Direct, A	The VRS and the VAS for low back pain are reliable for clinical use
Lund et al. 2005 [26]	Prospective study, Direct, A	Authors favor the VRS in pain intensity assessments over the VAS. The VAS data should be analyzed using statistical methods suitable for ordinal data.
de C. Williams et al. 2011 [27]	Prospective study, Direct, A	The VRS and NRS appeared to yield useful information in a post-conflict situation compared to the VAS which was difficult to explain and understand among participants in this study.

[a] Direct evidence: Primary research that compares different response scales within study. Indirect evidence: Review or expert opinion based on empirical evidence
[b] Grade Key: A) Primary research; compares different response scales within the study; B) Review or expert opinion; based on an empirical evidence base; C) Primary research; evaluates a single response scale type within the study; and D) Review or expert opinion, based on expert consensus, convention or historical experience

NRS) for the self-assessment of pain in adults [28, 29]. The majority of the reviewed articles, 7 out of 13, identified the NRS as the most appropriate response scale for the assessment of pain in adults, including the IMMPACT guidelines for pain measurement in clinical trials and the National Institutes of Health Toolbox pain assessment [9, 10, 30–33].

The NRS circumvents problems for translations, administration, and scoring that may occur with the VAS on paper administration by ambiguous lines being drawn. The electronic format of the VAS may now avoid some of those issues stemming from paper administration;

however, potential issues remain, stemming from various formats of screen sizes, specifically that the length of the VAS line may be variable across devices.

Another consideration is that a VAS is not appropriate for telephone interview-based or interactive voice response (IVR) system-based data collection. The VAS relies on a respondent's ability to view the VAS and mark the position on it (via a pencil/pen, touch with a stylus or finger, or a mouse click) that represents his or her answer to the specific question. Hence, the VAS is limited to paper and screen-based electronic modes of data collection that enable that to occur.

In summary, both types of articles, reviews and empirical studies, provide evidence that the 11-point NRS is likely the optimal self-report response scale to evaluate pain among adult patients without cognitive impairment. Furthermore, the 11-point NRS is the easiest to implement and utilize on electronic data collection systems that time- and date- stamp the respondent's data entry, which is recommended within FDA's PRO Guidance for supporting PRO-based labelling claims [1].

Discussion

In this study, we conducted a comprehensive evaluation of the scientific literature to identify response scale option types for pain studies and to review and summarize available empirical evidence for each type of response scale by context of use, to enhance response scale selection for newly developed self-reported pain instruments

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 7 of 9

in the adult population. With the publication of the IMMPACT guidelines in 2005 [9], the 11-point (i.e., 0 to 10) NRS was recommended as a core outcome measure of chronic pain treatment trials. Subsequently, the NRS became the gold standard for pain assessment in clinical trials; however, it is by no means universal. For example, some analgesics may have differential efficacy across pain types, and neuropathic versus musculoskeletal pain demonstrate different characteristics. However, extensive empirical evidence has been generated since the 2005 publication of the IMMPACT guidelines, and our review demonstrated that the recommended NRS response scale still appears to have the best performance when directly compared to other commonly used response scale types in the self-assessment of pain. NRS measures tend to be preferred over VAS measures by patients [9].VAS measures often result in more missing and incomplete data than NRS measures; this is most likely due to the easier to understand and less abstract nature of NRS measures [9]. Our findings suggested that pain assessment typically consists of single-item questions which typically use either an NRS or a VAS.

The 2014 FDA Guidance for Industry, "Analgesic Indications: Developing Drug and Biological Products" [34], provides recommendations to sponsors on the development of prescription drugs that are the subject of new drug applications for the management of acute and chronic pain as well as the management of breakthrough pain. The FDA recommends the use of a well-defined and reliable PRO instrument to assess the subject's pain intensity. Because pain is a subjective experience, the choice of an adequate instrument to measure the primary endpoint is critical to demonstrating the efficacy of an analgesic. Therefore, it is important to consider whether a well-defined and reliable instrument already exists or can be developed for an analgesic study. It is also important that measures be based on scales or instruments that have been adequately developed for use in the population to be studied and that the instruments are appropriate for use in the setting of a clinical trial to measure change over time. The FDA's recommendations highlight the importance of choosing an adequate response scale when selecting a pain instrument for a clinical trial.

Of course, this study has limitations. First, this review does not address the considerations for measuring the various phenotypic manifestations of pain. Pain can be categorized according to its duration, acute or chronic, as well as other characteristics, such as breakthrough pain and episodes of acute pain that occur in the context of otherwise well-controlled, chronic pain. The recommendations emerging from this review speak broadly to the concept of pain and do not reflect an examination of those intricacies due to the lack empirical evidence in the literature. In addition, this review did not discriminate between the choice of response scale for a single stand-alone item or for more than one item in a multi-dimensional pain inventory. Also, the review and merit of multi-domain PRO instruments which include pain were not included in this review. Nor does this review address the question of whether pain should be measured unidimensionally or across multiple domains.

Further, this review does not attempt to address other aspects of pain endpoints, such as recall period or worst vs. average pain intensity that may impact instrument validity. The PRO instrument's recall period for assessing pain should be appropriate for the type of pain studied and the planned study design. The FDA recommends use of an instrument that asks the subject to assess his or her worst pain over a relatively short time period, and no longer than the past 24 h, with the assessment occurring at the same time [34].

Additionally, this review focused on adult pain and does not attempt to address other populations (e.g., cognitively impaired, pediatric), or other reporters (e.g., observer- or clinician-rated). In cases of young children or subjects who cannot provide self-report, observers (e.g., parents, caregivers) can report on observable indicators (e.g., events, behaviors, or signs) of pain (e.g., wincing, crying, or squirming). However, an observer cannot validly rate a subject's pain intensity; and the FDA does not consider an instrument that requires an observer to do so to be well-defined or reliable [34].

Lastly, this literature review was conducted in early 2015 and was limited to articles published in English during a 10-year timespan from 2004 through 2014. Although a brief scan of the literature since 2014 did not find key evidence that would change the conclusions, the authors recognize that ending the structured review at 2014 is a limitation of this manuscript.

Conclusions

Overall, the literature supports the NRS as the preferred scale for pain measurement; the available empirical evidence demonstrates its superior performance to other response scale types in this context of use. It is simple, straightforward, and easy to interpret. This review includes evidence from a wide range of painful conditions, including acute pain as well as chronic neuropathic and musculoskeletal indications, suggesting a relatively broad generalizability of the findings. Recommendations for future research include well-designed head-to-head comparisons of response scales used in pain measures, structured reviews of pain assessment in other populations (e.g. pediatric, geriatric), and further exploration of the measurement considerations for the various phenotypic manifestations of pain.

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 8 of 9

Availability of data and materials
This article is entirely based on data and materials that have been published, are publicly available (thus, accessible to any interested researcher), and appear in the References list.

Authors' contributions
Each author participated in drafting the manuscript or revising it critically. Concentrating on the study concept and design were KG and SS: JT, LA, and DR handled data acquisition; KR, ENB, and MV focused on the analysis and data interpretation. All authors have given their approval of the final version or the manuscript. All the authors have agreed to be accountable for all aspects of the work, particularly for ensuring that any questions of the work's accuracy or integrity are promptly investigated and resolved.

Ethics approval and consent to participate
This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication
Not applicable.

Competing interests
SJC reports no employment in a pharmaceutical company nor does he hold stocks, shares, or stock options in a pharmaceutical company. LA is an employee and owns stocks, shares, and stock options for Pfizer Ltd. KG reports she is a current employee at Janssen but reports no stocks, shares, nor options. ENB works for Eli Lilly & Company and holds stocks or shares there, but no stock options. JTH and DSR are salaried employees and own stocks, shares, and stock options at Ironwood Pharmaceuticals. KR is an employee of Shire and owns stock options there. JT is a current employee at Janssen but reports no stocks, shares, nor options. SS and MV are employees of Evidera, a research and consulting firm to the biopharma industry and, as such, are not allowed to accept remuneration from any Evidera clients. None of these authors report any other arrangements that could be perceived as conflicts of interest.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Evidera, 7101 Wisconsin Ave., Suite 1400, Bethesda, MD 20814, USA.
[2]Janssen Global Services LLC, 700 Route US 202, Raritan, NJ 08869, USA.
[3]Ironwood Pharmaceuticals, 301 Binney Street, Cambridge, MA 02142, USA.
[4]Pfizer LTD, Outcomes & Evidence, Global Health & Value, Pfizer Ltd, Tadworth, Surrey, UK. [5]Patient-Reported Outcome Consortium, Critical Path Institute, 1730 E River Rd, Tucson, AZ 85718, USA. [6]Global Patient Outcomes and Real World Evidence, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA.

References
1. Food and Drug Administration. (2009). Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. *Federal Register, 74*(235), 65132–65133 Available at: Https://www.Fda.Gov/downloads/drugs/guidances/ucm193282.Pdf.
2. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (pro) instruments for medical product evaluation: Ispor pro good research practices task force report: Part 1-eliciting concepts for a new pro instrument. *Value in Health, 14*(8), 967–977. https://doi.org/10.1016/j.jval.2011.06.014.
3. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (pro) instruments for medical product evaluation: Ispor pro good research practices task force report: Part 2-assessing respondent understanding. *Value in Health, 14*(8), 978–988. https://doi.org/10.1016/j.jval.2011.06.013.
4. Leidy, N. K., & Vernon, M. (2008). Perspectives on patient-reported outcomes: Content validity and qualitative research in a changing clinical trial environment. *Pharmacoeconomics, 26*(5), 363–370.
5. Frank, L., Basch, E., Selby, J. V., & The Patient-Centered Outcomes Research Institute. (2014). The pcori perspective on patient-centered outcomes research. *JAMA, 312*(15), 1513–1514. https://doi.org/10.1001/jama.2014.11100.
6. Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: A practical guide to their development and use, fifth edition* (5th ed.). New York: Oxford University Press.
7. Willke, R. J., Burke, L. B., & Erickson, P. (2004). Measuring treatment impact: A review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Controlled Clinical Trials, 25*(6), 535–552. https://doi.org/10.1016/j.cct.2004.09.003.
8. Khadka, J., Gothwal, V. K., McAlinden, C., Lamoureux, E. L., & Pesudovs, K. (2012). The importance of rating scales in measuring patient-reported outcomes. *Health and Quality of Life Outcomes, 10*, 80. https://doi.org/10.1186/1477-7525-10-80.
9. Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P., Kerns, R. D., Stucki, G., Allen, R. R., Bellamy, N., Carr, D. B., Chandler, J., Cowan, P., Dionne, R., Galer, B. S., Hertz, S., Jadad, A. R., Kramer, L. D., Manning, D. C., Martin, S., McCormick, C. G., McDermott, M. P., McGrath, P., Quessy, S., Rappaport, B. A., Robbins, W., Robinson, J. P., Rothman, M., Royal, M. A., Simon, L., Stauffer, J. W., Stein, W., Tollett, J., Wernicke, J., Witter, J., & IMMPACT. (2005). Core outcome measures for chronic pain clinical trials: Immpact recommendations. *Pain, 113*(1–2), 9–19. https://doi.org/10.1016/j.pain.2004.09.012.
10. Williamson, A., & Hoggart, B. (2005). Pain: A review of three commonly used pain rating scales. *Journal of Clinical Nursing, 14*(7), 798–804. https://doi.org/10.1111/j.1365-2702.2005.01121.x.
11. Coons, S. J., Kothari, S., Monz, B. U., & Burke, L. B. (2011). The patient-reported outcome (pro) consortium: Filling measurement gaps for pro end points to support labeling claims. *Clinical Pharmacology and Therapeutics, 90*(5), 743–748. https://doi.org/10.1038/clpt.2011.203.
12. Gries, K., Safikhani, S., Pease, S., Harrington, M., Rudell, K., Berry, P., Crescioni, M., Patel, M., Vernon, M. (2018). Literature review to assemble the evidence for response scales used in patient-reported outcome measures. *Journal of Patient-Reported Outcomes*. https://doi.org/10.1186/s41687-018-0056-3.
13. Ritter, P. L., Gonzalez, V. M., Laurent, D. D., & Lorig, K. R. (2006). Measurement of pain using the visual numeric scale. *The Journal of Rheumatology, 33*(3), 574–580.
14. Jackson, D., Horn, S., Kersten, P., & Turner-Stokes, L. (2006). Development of a pictorial scale of pain intensity for patients with communication impairments: Initial validation in a general population. *Clinical Medicine (London, England), 6*(6), 580–585.
15. Gonzalez-Fernandez, M., Ghosh, N., Ellison, T., McLeod, J. C., Pelletier, C. A., & Williams, K. (2014). Moving beyond the limitations of the visual analog scale for measuring pain: Novel use of the general labeled magnitude scale in a clinical setting. *American Journal of Physical Medicine & Rehabilitation, 93*(1), 75–81. https://doi.org/10.1097/PHM.0b013e31829e76f7.
16. Gagliese, L., Weizblit, N., Ellis, W., & Chan, V. W. (2005). The measurement of postoperative pain: A comparison of intensity scales in

Safikhani *et al. Journal of Patient-Reported Outcomes* (2018) 2:40

Page 9 of 9

younger and older surgical patients. *Pain, 117*(3), 412–420. https://doi.org/10.1016/j.pain.2005.07.004.

17. Chien, C. W., Bagraith, K. S., Khan, A., Deen, M., & Strong, J. (2013). Comparative responsiveness of verbal and numerical rating scales to measure pain intensity in patients with chronic pain. *The Journal of Pain, 14*(12), 1653–1662. https://doi.org/10.1016/j.jpain.2013.08.006.

18. Chanques, G., Viel, E., Constantin, J. M., Jung, B., de Lattre, S., Carr, J., Cisse, M., Lefrant, J. Y., & Jaber, S. (2010). The measurement of pain in intensive care unit: Comparison of 5 self-report intensity scales. *Pain, 151*(3), 711–721. https://doi.org/10.1016/j.pain.2010.08.039.

19. Rothaug, J., Weiss, T., & Meissner, W. (2013). How simple can it get? Measuring pain with nrs items or binary items. *The Clinical Journal of Pain, 29*(3), 224–232. https://doi.org/10.1097/AJP.0b013e31824c5d7a.

20. Huang, K. T., Owino, C., Gramelspacher, G. P., Monahan, P. O., Tabbey, R., Hagembe, M., Strother, R. M., Njuguna, F., & Vreeman, R. C. (2013). Prevalence and correlates of pain and pain treatment in a western Kenya referral hospital. *Journal of Palliative Medicine, 16*(10), 1260–1267. https://doi.org/10.1089/jpm.2013.0080.

21. Huang, K. T., Owino, C., Vreeman, R. C., Hagembe, M., Njuguna, F., Strother, R. M., & Gramelspacher, G. P. (2012). Assessment of the face validity of two pain scales in Kenya: A validation study using cognitive interviewing. *BMC Palliative Care, 11*, 5. https://doi.org/10.1186/1472-684X-11-5.

22. Skovlund, E., Bretthauer, M., Grotmol, T., Larsen, I. K., & Hoff, G. (2005). Sensitivity of pain rating scales in an endoscopy trial. *The Clinical Journal of Pain, 21*(4), 292–296.

23. Dogan, S. K., Ay, S., Evcik, D., Kurtais, Y., & Gokmen Oztuna, D. (2012). The utility of faces pain scale in a chronic musculoskeletal pain model. *Pain Medicine, 13*(1), 125–130. https://doi.org/10.1111/j.1526-4637.2011.01290.x.

24. Loos, M. J., Houterman, S., Scheltinga, M. R., & Roumen, R. M. (2008). Evaluating postherniorrhaphy groin pain: Visual analogue or verbal rating scale? *Hernia, 12*(2), 147–151. https://doi.org/10.1007/s10029-007-0301-9.

25. Olaogun, M. O. B., Adedoyin, R. A., Ikem, I. C., & Anifaloba, O. R. (2004). Reliability of rating low back pain with a visual analogue scale and a semantic differential scale. *Physiotherapy Theory and Practice, 20*(2), 135–142.

26. Lund, I., Lundeberg, T., Sandberg, L., Budh, C. N., Kowalski, J., & Svensson, E. (2005). Lack of interchangeability between visual analogue and verbal rating pain scales: A cross sectional description of pain etiology groups. *BMC Medical Research Methodology, 5*, 31. https://doi.org/10.1186/1471-2288-5-31.

27. de Williams, A. C. C., Rajput-Ray, M., Lassalle, X., Crombie, I., & Lacoux, P. (2011). Assessing pain and mood in a poorly resourced country in a post-conflict setting. *Journal of Pain and Symptom Management, 42*(2), 301–307. https://doi.org/10.1016/j.jpainsymman.2010.11.012.

28. Grimmer-Somers, K., Vipond, N., Kumar, S., & Hall, G. (2009). A review and critique of assessment instruments for patients with persistent pain. *Journal of Pain Research, 2*, 21–47.

29. Hjermstad, M. J., Fayers, P. M., Haugen, D. F., Caraceni, A., Hanks, G. W., Loge, J. H., Fainsinger, R., Aass, N., Kaasa, S., & European Palliative Care Research C. (2011). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: A systematic literature review. *Journal of Pain and Symptom Management, 41*(6), 1073–1093. https://doi.org/10.1016/j.jpainsymman.2010.08.016.

30. Cook, K. F., Dunn, W., Griffith, J. W., Morrison, M. T., Tanquary, J., Sabata, D., Victorson, D., Carey, L. M., Macdermid, J. C., Dudgeon, B. J., & Gershon, R. C. (2013). Pain assessment using the nih toolbox. *Neurology, 80*(11 Suppl 3), S49–S53. https://doi.org/10.1212/WNL.0b013e3182872e80.

31. Khorsan, R., Coulter, I. D., Hawk, C., & Choate, C. G. (2008). Measures in chiropractic research: Choosing patient-based outcome assessments. *Journal of Manipulative and Physiological Therapeutics, 31*(5), 355–375. https://doi.org/10.1016/j.jmpt.2008.04.007.

32. Vela, L. I., Haladay, D. E., & Denegar, C. (2011). Clinical assessment of low-back-pain treatment outcomes in athletes. *Journal of Sport Rehabilitation, 20*(1), 74–88.

33. Wolters, P. L., Martin, S., Merker, V. L., Gardner, K. L., Hingtgen, C. M., Tonsgard, J. H., Schorry, E. K., Baldwin, A., & Collaboration REI. (2013). Patient-reported outcomes in neurofibromatosis and schwannomatosis clinical trials. *Neurology, 81*(21 Suppl 1), S6–S14. https://doi.org/10.1212/01.wnl.0000435747.02780.bf.

34. Food and Drug Administration (2014) Guidance for industry on analgesic indications: Developing drug and biological products (draft guidance). Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM384691.pdf. Accessed 24 June 2016.

35. Ong, K. S., & Seymour, R. A. (2004). Pain measurement in humans. *Surgeon, 2*(1), 15–27.

36. Schofield, P., Clarke, A., Faulkner, M., Ryan, T., Dunham, M., & Howarth, A. (2006). Assessment of pain in older people. Where are we now and what needs to be done? *International Journal on Disability and Human Development, 5*(1), 3–8.

37. Litcher-Kelly, L., Martino, S. A., Broderick, J. E., & Stone, A. A. (2007). A systematic review of measures used to assess chronic musculoskeletal pain in clinical and randomized controlled clinical trials. *The Journal of Pain, 8*(12), 906–913. https://doi.org/10.1016/j.jpain.2007.06.009.

38. Phillips, S. E. (2007). Pain assessment in the elderly. *US Pharmacist, 32*(5), 37–52.

39. Riddle, D. (2013). Instruments to measure acute pain: An integrative review. *Journal of Pain Management, 6*(4), 273–280.

40. Green, A., Liles, C., Rushton, A., & Kyte, D. G. (2014). Measurement properties of patient-reported outcome measures (proms) in patellofemoral pain syndrome: A systematic review. *Manual Therapy, 19*(6), 517–526. https://doi.org/10.1016/j.math.2014.05.013.

41. Grotle, M., Brox, J. I., & Vollestad, N. K. (2004). Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine (Phila Pa 1976), 29*(21), E492–E501.

42. van Dijk, J. F., van Wijck, A. J., Kappen, T. H., Peelen, L. M., Kalkman, C. J., & Schuurmans, M. J. (2012). Postoperative pain assessment based on numeric ratings is not the same for patients and professionals: A cross-sectional study. *International Journal of Nursing Studies, 49*(1), 65–71. https://doi.org/10.1016/j.ijnurstu.2011.07.009.

43. Chan, J., & Spencer, J. (2005). Contrasting perspectives on pain following hand injury. *Journal of Hand Therapy, 18*(4), 429–435, quiz 436. https://doi.org/10.1197/j.jht.2005.07.004.

44. Boomershine, C. S., Emir, B., Wang, Y., & Zlateva, G. (2011). Simplifying fibromyalgia assessment: The vasfiq brief symptom scale. *Therapeutic Advances in Musculoskeletal Disease, 3*(5), 215–226. https://doi.org/10.1177/1759720X11416863.

45. Bergh, I. H., Stener-Victorin, E., Wallin, G., & Martensson, L. (2011). Comparison of the painmatcher and the visual analogue scale for assessment of labour pain following administered pain relief treatment. *Midwifery, 27*(1), e134–e139. https://doi.org/10.1016/j.midw.2009.03.004.