The Sensation and Pain Rating Scale: easy to use, clear to interpret, and responsive to clinical change

Corresponding author: Victoria J Madden, Pain team, Department of Anaesthesia and Perioperative Medicine, University of Cape Town, D23 Groote Schuur Hospital, Main Rd, Observatory, 7925, Cape Town. HIV Mental Health Research Unit, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town. IIMPACT in Health, University of South Australia, Adelaide, Australia. torymadden@gmail.com. ORCID: 0000-0002-5357-4062.

Peter Kamerman
Affiliation(s): Brain Function Research Group, School of Physiology, Faculty of Health Sciences, University of the Witwatersrand, South Africa
Email: peter.kamerman@wits.ac.za
ORCID: 0000-0002-3103-5295

Hayley B. Leake
Affiliation(s): IIMPACT in Health, Allied Health and Human Performance, University of South Australia, Kaurna Country, Adelaide, SA, Australia.
Email: Hayley.leake@unisa.edu.au
ORCID: 0000-0002-6133-2186

Mark J. Catley
Affiliation(s): IIMPACT in Health, Allied Health and Human Performance, University of South Australia, Kaurna Country, Adelaide, SA, Australia.
Email: mark.catley@unisa.edu.au
ORCID: 0000-0002-1582-4390

Lauren C. Heathcote
Affiliation(s): Health Psychology Section, Department of Psychology, Institute of Psychiatry Psychology and Neuroscience, King's College London
Email: lauren.heathcote@kcl.ac.uk
ORCID: 0000-0003-2515-3102

G Lorimer Moseley
Affiliation(s): IIMPACT in Health, Allied Health and Human Performance, University of South Australia, Kaurna Country, Adelaide, SA, Australia.
Email: lorimer.moseley@gmail.com
ORCID: 0000-0002-3750-4945

Number of pages of manuscript: 36
Number of figures: 8
Number of tables: 4

Abstract

The Sensation and Pain Rating Scale (SPARS) allows rating of non-painful as well as painful percepts. Although it performs well in the experimental context, its clinical utility has not been tested. This prospective, repeated-measures study mixed qualitative and quantitative methods to examine the utility and performance of the SPARS in a clinical context, and to compare it with the widely used 11-point NRS for pain. Patients (n = 121) provided ratings on the SPARS and NRS at 6 time points: at first consultation, before and after sham and active clinical interventions, and at follow-up consultation. Clinicians (n = 9) reported each scale's usability and interpretability using Likert-type scales and free text questions, and responded to other questions with free text. Each data type was initially analysed separately: quantitative data were visualised and the ES II metric was used to estimate SPARS internal responsiveness; qualitative data were analysed with a reflexive inductive thematic approach. Data types were then integrated for triangulation and complementarity. The SPARS was well received and considered easy to use, after initial familiarisation. Clinicians favoured the SPARS over the NRS for clarity of interpretation and inter-rater reliability. SPARS sensitivity to change was good (ESII=0.9; 95% CI: 0.75-1.10). The greater perceptual range of the SPARS was deemed especially relevant in the later phases of recovery, when pain may recede into discomfort that still warrants clinical attention. The SPARS is a promising tool for assessing patient percept, with strong endorsement from clinicians for its clarity and superior perceptual scope.

Introduction

In experimental settings, participants are frequently asked to rate the perceived intensity of stimuli across a range of stimulus strengths. The perception of these stimuli may range from no sensation through to the worst pain imaginable. To capture this full range of percepts, we developed the Sensation and Pain Rating Scale (SPARS), an analogue scale with left-most anchor of -50 ('no sensation'), central anchor of 0 ('the exact point at which what you feel transitions to pain'), and right-most anchor of +50 ('the worst pain you can imagine'). Thus, the range from -50 to 0 reflects non-painful percepts; 0 to +50 reflects painful percepts. The SPARS has good psychometric properties [27]. Several studies have used [4; 8; 25; 36] or adapted [16] it as an outcome measure.

Although the SPARS was developed for experiments, its broad perceptual range may be useful in the clinical setting. As a person recovers from a painful episode, pain evoked by certain tasks or movements may recede into non-painful percepts that still reflect deviation from a state of 'fully recovered'. Clinically, these percepts are important in guiding ongoing rehabilitation. For example, reports of stiffness, pulling, non-painful sensitivity to movement, and not feeling 'quite right' are all potentially important in pacing return to full loading, or assessing non-tissue-related contributions to pain and disability. Quantifying 'the extent to which the percept deviates from asymptomatic' may improve clinical reasoning, reporting, and planning. Conventional numerical rating (NRS) and visual analogue scales for pain do not provide this utility. When using a 0-100 scale anchored with at 0 'no pain', a quarter of patients used numbers greater than 0 to rate clinical sensations that they also classified as non-painful [24]. Anecdotally, clinicians report variable NRS thresholds for the transition to pain. This improvised, unstandardised, and idiosyncratic scale use has the potential to obscure clear communication between patient and clinician, between clinicians, and between clinicians and the stakeholders they report to, such as compensation providers.

This mixed-methods study aimed to examine the utility and performance of the SPARS in a clinical context, and compare it to the widely used 11-point (0-10) NRS. We investigated clinicians' experiences of the SPARS with respect to usability, interpretability, and responsiveness to clinical change. We also compared the SPARS and NRS for clinical utility, relative coverage of the range of clinically relevant percepts, and agreement. In light of the historical data referenced above [24], we were interested in how events rated as non-painful on the SPARS would be rated on the NRS. We anticipated that clinicians would favour the non-painful range of the SPARS for its greater coverage of clinically relevant percepts, and that this preference for the SPARS would be more apparent as patients recovered than when patients presented with recent-onset pain. We also anticipated that clinicians would report some difficulty orientating to the SPARS anchors. Finally, we anticipated that, for events rated as painful on the SPARS, agreement between ratings on the SPARS and ratings on the NRS would be good.

Methods

Study overview

This study was a prospective, repeated-measures study in multiple physiotherapy practices in Australia. Physiotherapy practices were chosen as the ideal context for the study, given that many

patients continue their physiotherapy rehabilitation toward full function, return to work or play. The study was approved by the University of South Australia's Human Research Ethics Committee (Application ID: 202696). Participants included clinicians and patients, with each group having a distinct participant role. Figure 1 shows the overall structure of the study. Data collection comprised two phases. In Phase 1, participating clinicians administered both a conventional 0 – 10 NRS and the SPARS to participating patients, at six time points: on first presentation of the patient; immediately before and after a therapeutic intervention expected to produce pain relief – herein called an 'active' intervention); immediately before and after an intervention' expected to *not* produce pain relief – herein called a 'sham' intervention); and at a follow-up appointment anticipated to be the last consultation (ranged from 5-62 days after first presentation). The active and sham interventions were independently selected by the clinicians, according to their clinical judgement in the context of each patient's presentation. In Phase 2, clinicians administered the SPARS but not the NRS, at the same time points. Once both Phases 1 and 2 had been completed, the participating clinicians completed an online form to provide their perspectives on the SPARS and NRS.

** Figure 1 approximately here**

The protocol for this study was locked on Open Science Framework [26] (at: https://tinyurl.com/spars-clinical1), after the data collection had been completed and before the data were inspected or analysed. The protocol was intended to achieve accountability about the intended quantitative analysis. As such, it did not specify plans for the qualitative analysis, nor for integration of the data. Table S1 of the supplementary file presents the questions we aimed to answer, the quantitative analysis plan for each, notes on any deviations from those plans, and the availability of qualitative data where applicable. The two most important deviations were for blinding of the data analyst (see Blinding section below) and the approach to estimating sensitivity to change.

Blinding

Patients were blinded to the aim of the study and the use of active and sham interventions, but participating clinicians were not. Data entry investigators were blinded to the aim of the study, the range and nature of the SPARS, the order in which assessments were undertaken, and the order of the active and sham intervention. The data analysts were not blinded (protocol deviation; Table S1). No assessments were undertaken to identify potential loss of blinding.

Participants

A convenience sample of practising clinicians was recruited via a clinical advisory network established through the Pain Adelaide Stakeholders' Consortium. Volunteer clinicians were asked to inform consecutive eligible patients about the study. Patients were eligible if they were aged over 18 years, were presenting for treatment from one of the participating clinicians, and were judged by the participating clinician to have sufficient English language and cognitive proficiency to provide verbal ratings using each of the two numerical rating scales. Consenting volunteer patients were administered the two scales by the clinician in the manner described above. The clinician would then submit the data to a researcher (GLM) via email. At the completion of data collection, the clinicians were asked to respond to an online questionnaire to reflect on their experiences. Each clinician and each patient was given a unique study ID and all submitted data were anonymised.

Sample size and power calculations

Sample size was determined on the basis of precision estimates of a representative distribution of SPARS ratings taken from a pilot trial in a clinical setting, and on extrapolation of published accounts of SPARS data obtained in the experimental context [25; 27; 36]. For this experiment, the analysis requiring the largest sample was that to determine whether the SPARS yields different ratings from the NRS for percepts within the painful range. We used clinical pilot data obtained by one clinician (GLM), who administered the NRS and the SPARS with the question 'What was your average pain over the last two days?' Data were collected from 20 consecutive patients who presented for care with a chronic pain disorder. Using an online sample size calculator (HyLown.com), mean NRS ratings, mean SPARS ratings/5 (to convert the 0 to +50 range of the SPARS to a rating out of 10), a standard deviation across both of 1.5, an acceptable difference of 0.1, 80% power and alpha of 0.05, we required 87 NRS and SPARS rating pairs. We allowed for 10% missing data, which meant we aimed recruit 96 patient participants for Phase 1. We aimed to recruit 25 patients for Phase 2, in which patients completed only the SPARS (to verify that the utility and integrity of ratings on the SPARS are not dependent on, or confounded by, also obtaining ratings on the NRS).

Outcomes

In Phase 1, patients provided ratings on both the SPARS and the conventional NRS for pain, under six conditions. The order in which the two scales were administered for each participant was determined the study team, who used a random number table in MS Excel. This order was held consistent within participants. In Phase 2, patients provided ratings on the SPARS only, under the same six conditions. These ratings were collated by double-entry into an electronic spreadsheet, alongside the following descriptive information:

- Clinician study ID (anonymised)
- Patient study ID, age, gender (from options: male; female; other), diagnosis (from options: spinal; limb; pelvic; complex regional pain syndrome; widespread or multiple sites; not listed)
- Order of ratings: NRS or SPARS first
- Type of pain provocation task chosen by the clinician (options: functional; pressure; manual therapy; movement)
- Type of 'active' intervention chosen by the clinician (options: manual therapy; education or advice; an active exercise; active or passive stretch)
- Type of 'sham' intervention chosen by the clinician (options: manual therapy; education or advice; an active exercise; active or passive stretch)
- Order of interventions (randomised *a priori* via a random number table in Excel)
- Time between initial assessment and follow-up or final assessment (clinician discretion)
- Patient perspective at follow-up or final assessment on the need for further intervention (yes/no)
- Patient perspective at follow-up or final assessment of whether they consider themselves to have completely recovered (yes/no).

Primary outcome – the SPARS

All patients completed the SPARS under six different conditions. For each consenting patient, the clinician briefly explained the SPARS and asked the patient to rate the feeling in a specific body part (the site of a patient-reported pain complaint) in six situations.

At initial interview (one condition): the patient was asked: 'On a scale from minus 50 to 50, where minus 50 means you feel nothing, where plus 50 is the worst pain possible, and where zero is the exact point at which you feel pain, how would you rate the average feeling in your [affected body part] over the last two days?'
In initial assessment (four conditions; before/after an active/sham intervention): The clinician selected a pain provocation test (e.g. manual assessment, functional movement) on the basis of their own discretion. They then selected two brief interventions – one that they predicted to decrease pain and that would be part of their normal assessment or treatment (the 'active' intervention), and one that they predicted would have no effect on pain based on their own discretion, but that might be part of their normal assessment (the 'sham' intervention).

The order in which they performed these interventions was randomised *a priori* using a random numbers table. Prior to the interventions, the clinician explained, in their own words, that they wanted to determine how pain on the provocation test is affected by different tasks. Clinicians were instructed to keep patients blinded to the use of active and sham interventions. Before and after performing each intervention, the clinician implemented the provocation test and asked the patient: 'On a scale from minus 50 to 50, where minus 50 means you feel nothing, where plus 50 is the worst pain possible, and where zero is the exact point at which you feel pain, how would you rate the feeling in your [affected body part] during that task?'
At final or long-term follow-up assessment (one condition): the patient was asked, 'On a scale from minus 50 to 50, where minus 50 means you feel nothing, where plus 50 is the worst pain possible, and where zero is the exact point at which you feel pain, how would you rate the average feeling in your [affected body part] over the last week?' If this assessment occurred less than one week after the first assessment, the time frame for the question about pain was adjusted to the last 'two days', in order to exclude the same period over which the patient had already provided a rating at the initial interview.

Secondary outcome – 11-point conventional NRS for pain

Patients in Phase 1 also completed a conventional 11-point NRS for pain, in each of the six conditions described above. The wording for the NRS was: 'On a scale from 0 to 10, where zero is no pain and 10 is worst possible pain, how would you rate your pain …'

When the collection of ratings data had been completed, each clinician was asked to complete an electronic form to report their perspectives on the SPARS and NRS. Six questions (quantitative) required a response on a Likert-type scale with options of: Very easy, Somewhat easy, Neither easy nor difficult, Somewhat difficult, and Very difficult. Eight questions (qualitative) allowed free text responses. The questions were:

1. How easy or difficult was it to explain the SPARS to your patients? (Likert scale)
2. How easy or difficult was it to explain the NRS to your patients? (Likert scale)
3. Please add any comments on this AND tell us how we could make it easier for you to explain the SPARS and/or the NRS. (Free text)
4. How easy or difficult was it for your patients to report a SPARS rating? (Likert scale)
5. How easy or difficult was it for your patients to report a NRS rating? (Likert scale)
6. Please add any comments on this AND tell us how we could make it easier for them to use the SPARS and/or the NRS. (Free text)
7. How easy or difficult was it for you to interpret your patients' SPARS ratings? (Likert scale)

8. How easy or difficult was it for you to interpret your patients' NRS ratings? (Likert scale)
9. Please comment on how your confidence in, and competence in, using the SPARS changed over time. (Free text)
10. Please add any comments on this AND tell us how we could make interpretation of the SPARS and/or the NRS easier for you. (Free text)
11. Please compare the SPARS to the NRS in any ways you think are relevant to a clinician working in the context in which you work. (Free text)
12. How likely are you to use the SPARS in your clinical practice from here on? (Free text)
13. How likely are you to recommend the SPARS to other clinicians? (Free text)
14. Please tell us anything else you think we should know about the SPARS, this study or anything at all really. (Free text)

Data Collection, storage, and analysis

All data were collected by treating clinicians via a bespoke form as hard copy or on a device or computer (clinician preference). No patient-specific data, aside from the variables listed above, were collected. Quantitative and qualitative data were collected concurrently. All rating and descriptive data were emailed to a researcher (GLM). Two data extractors, blinded to the study aims, worked independently to enter all clinician-collected data into electronic spreadsheets and then met to compare that their data entries matched. Mismatches were compared to the original files sent from clinicians. For clinician perspectives data, clinicians completed an electronic form hosted by SurveyMonkey®. No individual identifiers were recorded in either data set; only study IDs were used to identify each clinician and patient. However, until the analysis of data began, clinicians could be re-identified if necessary, using a separate, password-protected Excel file.
Data were tidied and wrangled using R (version 4.2.1 [35]) via RStudio (version 2022.07.0 [34]). We used the following packages: arsenal v. 3.6.3 [15], boot v. 1.3.28 [7; 10], cowplot v. 1.1.1 [40], epiR v. 2.0.50 [33], flextable v. 0.7.2 [13], ggdist v. 3.2.0 [20], ggpubr v. 0.4.0 [19], Gmisc v. 3.0.0 [14], gridExtra v. 2.3 [2], janitor v. 2.1.0 [12], kableExtra v. 1.3.4 [46], knitr v. 1.39 [41-43], lme4 v. 1.1.29 [3], patchwork v. 1.1.1 [30], rmarkdown v. 2.14 [1; 44; 45], tidyquant v. 1.0.4 [9], tidyverse v. 1.3.1 [39], and grateful [32].

Visualisation of the data was used to direct the most informative approach to quantitative analyses. Although we had made and recorded *a priori* decisions about the quantitative analytical approach in the locked protocol prior to inspecting the data, the process of data analysis raised questions about the relevance and utility of some of our planned approaches. In response to these discussions, we revised our plans for assessing sensitivity to change as described in Table S1. The analysis script output can be found at [26](https://tinyurl.com/spars-clinical1).
Here we present data visually and in tables. In the tables, unless otherwise specified, p-values are based on a two-sample t-test (for continuous variables) or a chi-square (for categorical variables), as per the 'arsenal' package [15] for R.

Data analysis

This study used both quantitative and qualitative data with the goal of addressing the research questions in a convergent mixed methods design. We mixed the two methods in design, data collection, in a process of integration after initially separate analysis, and in interpretation. Although most of the research questions (Table S1) were designed for specific data types, we had planned to use any available data, even if it provided information for a question for which it was not originally

intended. For example, if the qualitative data unexpectedly provided comments on sensitivity to change, we planned to present those data alongside the quantitative data.

Quantitative analyses

Sensitivity to change

Given the lack of suitable external reference for the perceptual range covered by the SPARS, we were not able to assess external responsiveness. Instead, we focused on internal responsiveness (protocol deviation; Table S1). We first visualised the change in SPARS before and after the active vs the sham intervention. We then used the 'effective size II' (ES II) metric, which is also known as the standardised response mean or the responsiveness-treatment coefficient or an efficiency index. It provides an estimate of change represented by the measure, standardised to the between-participant variability in the change scores, and is not dependent on sample size [17]. An ES II value greater than 0.80 has been suggested to represent large responsiveness, yet confidence intervals may be even more useful than benchmarks for supporting interpretation. We calculated the ES II metric itself and also generated bias-corrected and accelerated bootstrap confidence intervals for the ES II metric.

How do ratings on the SPARS compare to ratings on the NRS?

Relative coverage of perceptual range (SPARS vs NRS)
We visually compared the ranges covered by ratings on each scale, within each condition, between the two scales.

Agreement between SPARS and NRS ratings within comparable ranges
The NRS is expected to capture events that lie within a painful range. Therefore, for this part of the analysis we selected only those events that had been rated between 0 and 50 on the SPARS. We transformed the SPARS values to be comparable to the NRS values by dividing them by 5, and then used Bland-Altman plots and conventional correlation plots to visualise the data. For a quantitative measure of agreement, we selected Lin's concordance correlation coefficient (CCC) [22; 23], which is suited to two methods of measuring the same continuous variable [38]. Lin's CCC has a maximum value of 1, which indicates perfect agreement.

Use of NRS for 'non-painful' events
To understand how patients used the NRS to rate non-painful events, we selected only the events that were rated below 0 on the SPARS and visualised the SPARS and NRS ratings for those events.

Influence of NRS-rating task on SPARS responses

As an internal check of the study methodology, we wanted to understand the stability of SPARS ratings when NRS ratings were also requested (i.e. in Phase 1). For this, we visually compared the central tendency and distributions of SPARS ratings across the two study phases, given that NRS ratings were requested in Phase 1, but not in Phase 2.

Qualitative data analysis

Data were analysed using reflexive thematic analysis [5]. We used an inductive approach in which all the available data were coded, and the codes and resultant themes were developed based on the raw data without an *a priori* coding system. The first author (VJM) familiarised herself with the data by reading and re-reading. Using an iterative process, the data were recursively coded by VJM.

Codes reflected semantic and latent features of the data, to capture both the literal meanings of what participants said, and contextual interpretations of what they said and how they said it. Codes were reviewed and refined via discussions between VJM and HBL, to generate potential and then final themes and to name themes. For context, VJM is a clinical and research physiotherapist with little experience in qualitative data analysis but over 10 years' experience in clinical practice. VJM led the initial development and first experimental testing of the SPARS. HBL is a clinical and research physiotherapist who has led and published mixed methods studies. Neither VJM nor HBL was involved in data collection or in the clinical consultations.

Integration of data

Integration of data is recommended for mixed methods studies. It allows for deeper insights and capturing nuances in participants' responses that might not be revealed by only one data type. Therefore, we planned to use the qualitative data to corroborate, explain, illustrate, and add nuance to the quantitative results, especially regarding problems and proposed solutions. This broadly aligns with the Bryman [6] justifications of triangulation (corroboration/confirmation) and complementarity (clarification of results from one method by results from another, or expansion). Data were integrated after initial, separate analysis. Where more than one data type shed light on the same topic, the data have been presented using a joint display to identify points of divergence and convergence between the two data types.

Results

Participants

The respondent sample included nine treating clinicians (4M, 5F, average (range) of years practising = 8 (3 – 31)) collectively reporting data from a total of 121 patients (Table 1). There was one SPARS rating of 51 reported (which falls outside the range of the scale), which we retained for analysis on the assumption that the patient had indeed given that rating. The most common diagnosis reported was spinal pain (n = 50), followed by limb pain (n = 30), widespread pain (n = 23), complex regional pain syndrome (n = 9), and pelvic pain (n = 9). The number of patients from whom data were reported per clinician ranged from 23 (clinician 1) to three (clinician 9). These data can be seen in the supplementary file. There were 97 patients included in Phase 1, representing 464 rating pairs for painful events (minimum sample size = 87 rating pairs), and 24 patients included in Phase 2 (target sample size = 25).

Table 1 approximately here

Clinicians reported on the nature of their chosen provocation tests, 'active' interventions, and 'sham' interventions – data shown in Table 2. There were 17 missing data points for active intervention, and 25 missing data points for the sham intervention. The most common active intervention reported was education/advice (reported 50 times), whereas the most common sham intervention reported was manual therapy (reported 26 times).

Table 2 approximately here

The median (interquartile range) period from baseline ratings to follow-up ratings was 30 (22-42) days. The supplementary file includes tables comparing data from those who reported requiring further treatment or not at follow-up, and those who believed they had completely recovered or not at follow-up. At follow-up, 64% of patients (n = 77) said they still needed treatment. Those who reported still needing treatment had a marginally *longer* period to follow-up assessment, higher SPARS and NRS ratings for baseline pain and follow-up pain, and a smaller reduction in SPARS ratings but a counter-intuitively *greater* reduction in NRS from baseline to follow-up, compared to those who said they no longer needed treatment.
Also at follow-up, 88% (n = 107) reported they had not completely recovered. This group had higher SPARS and NRS ratings for baseline pain and follow-up pain, and a smaller reduction in ratings from baseline to follow-up in both scales, than those who said they had completely recovered.

Quantitative results

Ease of explanation, reporting, and interpreting the SPARS

Figure 2 shows clinicians' responses to the six questions asked about the SPARS and NRS. Regarding the SPARS, all the clinicians rated explaining, reporting, and interpreting the SPARS as either very easy or easy. Regarding the NRS, all clinicians rated explaining the NRS as 'very easy' or 'easy'. However, one clinician gave an ambiguous response (neither easy nor difficult) in response to the question about how easy it was for patients to report an NRS rating, and three clinicians reported that it was difficult to interpret their patients' NRS ratings. Overall, these data suggest that the ratings on the SPARS may be easier to interpret than ratings on the NRS.

** Figure 2 approximately here**

How sensitive to change was the SPARS?

We used visual analysis and formal calculations to assess the responsiveness of the SPARS to clinical change elicited by an active intervention. Figure 3 shows SPARS ratings before and after an active or a sham intervention, and shows some spread of ratings into the non-painful range of the SPARS after active intervention that is not seen after sham intervention. The ES II statistic was estimated to be 0.9 (95% CI: 0.75 – 1.10), suggesting excellent internal responsiveness to change.

** Figure 3 approximately here **

How do ratings on the SPARS compare to ratings on the NRS?

Coverage of perceptual range

Figure 4 shows the range of each scale that was used when both scales were used to rate the same event. There were 582 instances of this co-rating, across six different conditions.

** Figure 4 approximately here **

Agreement within comparable ranges
Figure 5 (below) shows Bland-Altman plots and simple correlation plots of the relationship between transformed SPARS ratings (SPARS divided by 5) and NRS ratings, for only those events that were rated >=0 on the SPARS and rated on both scales. There were too few data points (n = 25) to provide a useful analysis for the condition of 1-week average at follow-up. In general, the Bland-Altman plots

indicate that the average difference between measurements was less than 5% of the range (which was 0-10) for all conditions.
Table 3 shows concordance correlation coefficient values for the same data, by condition. Although the rules of thumb for interpreting CCC are variable and controversial, none of the values found was concerning.

Table 3 approximately here

** Figure 5 approximately here **

Use of NRS for 'non-painful' events
Figure 6 shows ratings on the SPARS and NRS for events that were rated in the non-painful range of the SPARS, and that were also rated on the NRS. Only three of the six conditions had events that satisfied these criteria. NRS ratings ranged from 0 to 1; SPARS ratings ranged from -50 to -1.

** Figure 6 approximately here **

Influence of NRS-rating task on SPARS responses: are SPARS ratings affected by collecting NRS ratings?

Figures 7 and 8 display SPARS ratings by study phase, to shed light on whether the central tendency and/or distribution of SPARS ratings was affected by also collecting NRS ratings (which took place in Phase 1 only). Although visual inspection of Figure 8 suggests greater spread of SPARS ratings after the active intervention in Phase 1 than in Phase 2 (middle right facet), this was not statistically significant.

** Figure 7 approximately here **

** Figure 8 approximately here **

Qualitative results

We generated two overarching themes related to clinicians' experiences using the SPARS: (1) Easy to use after brief familiarisation, and (2) Avoids several pitfalls of the NRS.

Theme 1: The SPARS was easy to use after brief familiarisation

Overall, clinicians reported positive experiences using the SPARS. Although some clinicians reported an adjustment period from conventional scales to the SPARS, the adjustment was quick.

> *"Really it is no harder to use than the usual NRS, just less familiar." (29)*

> *"I think patients are often very used to using the normal scales and that means they take a little while to get the hang of the new scale. But it really only takes a couple of reminders." (42)*

To help patients adjust to the SPARS, clinicians reported that careful explanation of the SPARS structure was necessary. Most clinicians expressed that, after a practice period, the SPARS structure was easy enough to support rapid adjustment.

> *"The main thing for me was the initial explanation and this was easy for sure." (29)*

> *"Oh it was very quick - a couple of run throughs and then the wording was easy and I am confident now and so is our whole clinic." (69)*

However, some clinicians reported challenges adjusting to the SPARS, irrespective of the speed of adjustment. Some of these challenges stemmed from the deeply ingrained familiarity with the previous pain rating scale – the NRS. Other challenges were attributed to the perceived lack of intuitiveness of the SPARS, in comparison to the NRS.

> *"I found it difficult at first to adjust to the SPARS … once I was a few patients in, it was fine, but I suspect all physios will need to adjust things slightly because asking about pain on a 'no pain' to 'worst pain' scale is so automatic for us." (42)*

> *"All my patients got the hang of it quickly but initially it was not as intuitive as the 0 - 10 scale (32)*

The difference between the perception of intuitiveness between the SPARS and NRS appeared to reflect discomfort with the numbering. That is, clinicians identified the two ranges of the SPARS (which are delineated by the SPARS anchors) as the main feature that had to be clarified for patients.

> *"I really think that once patients understand that it is a -50 to 50 scale and that pain starts at 0, then it is fine (46)*

> *"Once they understand that it is only positive numbers that are about pain then it is easy as pie." (29)*

Visual and electronic aids were proposed to facilitate understanding and use of the SPARS. Specifically, clinicians suggested creating an educational wall poster or a website to serve as a convenient reference for clinicians when communicating with patients. Others suggested adapting the SPARS as an electronic version, to smooth the familiarisation processes. Regardless of format, clearly written instructions – "better instructions above the scale (69)" – were recommended.

> *"I think it would be nice to have a nice-looking infographic that we could put on the wall so that we could just remind them about how it works." (42)*

> *"It would be very useful to get a digital version of it that could be integrated into the questionnaires that patients fill in in the waiting room before the appointment." (29)*

> *"I think it would be good to have a website that explains it because that would have saved me the time to teach the rest of the clinic … I would like to see some research on it too*

because I am just presuming that 25 on the SPARS matches 5 on the NRS, but is that how it really works?" (32)

As reflected in the third quote, above, access to a comprehensive explanation and comparative data on the SPARS was suggested, to prevent clinicians from having to make assumptions about how SPARS ratings relate to ratings on other scales. The lower anchor of the SPARS was also flagged as needing clarification.

"I had a few occasions when I was unsure how to define the [-50] end of the SPARS because feeling 'nothing' wasn't quite right. I used 'completely normal' for one patient but that wouldn't be very good for wider use because 'completely normal' might still be uncomfortable or even painful." (69)

Clinicians expressed that they liked that the SPARS allowed patients to describe sensations that were "not painful but don't feel right" (29). Clinicians said that the utility of this to allow tracking of progress after pain had resolved.

"I particularly also liked it when people were better or getting better because it allowed them to communicate the feeling that there was still some improvement to go even though they weren't really having pain anymore." (29)

On the whole, this theme captured generally positive feelings towards the SPARS, a brief period of adjustment, and some possible tools to support implementation of the SPARS in the clinical setting.

Theme 2: The SPARS avoids several pitfalls of the NRS

The second theme captures clinicians' reports that the SPARS offers a greater utility and interpretational clarity than the conventional NRS. Several clinicians noted concerns about the reliability of the NRS, and said they routinely added *ad hoc* anchors to the NRS to clarify a point of transition to pain. They noted that this added anchor was not necessary with the SPARS.

"[on the NRS] when people rate their pain as 1 or 2 and when I question them about it they actually say it is not so much painful as uncomfortable. I can explain it but then the ratings are dependent on me clearly talking them through 0 to 2 … we had an in-service [training session] on it before we started this and nearly decided that we would say that 2 out of 10 was when you are definitely sure it is painful. I think the SPARS solves this problem and I really like it. We all like it here." (69)

As is evident in the preceding quote, some clinicians described that their clinic had developed strategies to manage the lack of clarity associated with the NRS. Indeed, other clinicians described that these strategies had become so established that they referred to them as "policy" and felt the need to explain them when communicating with external stakeholders. They noted that these strategies would not be necessary if the SPARS was adopted.

"The primary implementation strategy that we use to facilitate interpretation and indeed utility of the NRS is to make 1 equal to the point at which it is definitely painful. We tend to articulate that as the pain threshold. I suspect it would assist us and our colleagues if [using SPARS] was a formal modification. This way we would no longer be compelled to provide an explanation of our policy when dealing with external stakeholders. (21)

Clinicians reported that using the SPARS had revealed previously unidentified problems with inter-rater reliability of the NRS. Reliability problems related to the varying ways that different clinicians explained the anchor for "0" and the numbers close to this anchor, on the NRS.

"Some were telling patients that 1 was more discomfort than pain and some were saying that 2 was when it was definitely painful... if the person came back to see someone else, they may score the same thing differently on the normal pain scale, but not on the SPARS." (25)

The SPARS was described by clinicians as a superior tool to the NRS because it resolved problems of inter-rater reliability. Specifically, the SPARS offered distinct categories of "no pain" and "no sensation", offering a more precise distinction between these two states.

"I think this is where the SPARS has a clear advantage over our usual assessment tools. We often use the VAS [visual analogue scale] for pain and we still get questions such as 'is this point (pointing to the left-hand end) where I feel absolutely nothing or is it where I feel some discomfort or stiffness?'" (08)

"[about 'conventional scales':] Even individual patients change how they are using them from appointment to appointment… I think the SPARS is much easier in this way because it makes very clear where the pain threshold is" (42)

Some participants even suggested a formal modification for the NRS, to improve interpretational clarity:

"It is worth suggesting that a modification be implemented for the NRS. This modification would be to label 1 as something akin to the zero point on the -50 to 50 scale. This is the manner in which we tend to use it anyway" (21)

"You could perhaps solve this problem with the VAS and the NRS by marking the point that is like the zero in the SPARS" (08)

Participants viewed the SPARS as better suited than the NRS to capturing patients' clinical ratings that transitioned through a range of sensations from no sensation, through a non-painful range, into a painful range. This feedback aligns with the concept of face validity, and suggests that the SPARS was thought to have greater face validity than the NRS.

"This is where I think the SPARS is better than our other methods of assessing pain to be honest with you. The thing is that VAS and NRS and really anything is actually quite tricky because you don't have no sensation at all and then suddenly pain and I know some patients who talk about being sore but not painful for example." (46)

This greater face validity translated to a reminder that patients' treatment goals go beyond just reducing pain.

> *"We think patients ultimately prefer the SPARS. I am not sure exactly why this is but perhaps it is just a reminder that their treatment goals go beyond just getting pain down; almost aiming higher sort of this?" (32)*

> *"It is especially helpful when patients are getting better but still feel like they have a way to go to be 'just right'" (25)*

These two themes appeared to culminate in a notable enthusiasm for ongoing use of the SPARS. Several clinicians reported they now encourage the use of SPARS in their clinic, or have implemented it and even encouraged other clinicians to use it.

> *"Our clinic now uses the SPARS and we did an in-service on it to the GPs and other allied health professionals who attended. So, we think it is really good!" (08)*

However, the potential for across-the-board use of the SPARS was limited by the scales required by official reports, and a reluctance to have two pain scales in use.

> *"We are all using it and reserving the NRS for workcover and insurance requirements." (32)*

> "*We still have to use the NRS too for some reports and I don't think we will use both." (69)*

Taken together, this second theme captures that clinicians considered the SPARS to address known problems with the NRS, inferring that the SPARS was superior to the NRS. The superior utility and interpretational clarity of the SPARS was largely attributed to its anchors, and to its explicit labelling of the pain-free range.

Integration of results

There were four topics on which both quantitative and qualitative data were available. Table 4 provides a joint visual display of quantitative and qualitative data by topic.

Table 4 approximately here

Discussion

This prospective study used both quantitative and qualitative data to examine the utility and performance of the SPARS, and to compare the SPARS with the conventional 0-10 NRS for pain, in a clinical context. Overall, participant feedback on the SPARS was positive. Clinicians described the SPARS as "conceptually attractive" and reported that it was easy to use after initial familiarisation, supported more comprehensive reporting of recovery, and was easier to interpret than the conventional NRS. Particularly interesting was that clinicians described using several approaches – including anchor adaptations or extra anchors—to manage uncertainty in the interpretation of the

lower anchor of the conventional NRS. This uncertainty was not deemed a problem for the SPARS. These adaptations were requiring extra time and effort from clinicians, including to explain the anchor adaptations to external stakeholders. Considering that unplanned anchor adaptations are likely to compromise inter-rater reliability, the clarity of the SPARS anchors may support better reliability in repeated ratings than the NRS.

Although the SPARS anchors were clear, there was mixed feedback on how easy the SPARS was to explain to a patient, mostly because the new structure of the SPARS required an adjustment period. However, clinicians described the adjustment as rapid. Some gave suggestions for poster-style visual aids and electronic tools that could support explanation and familiarisation, and possibly even allow patients to independently provide ratings on the SPARS before consultations.

The SPARS was good at capturing change produced by an intervention that was expected to alter pain: the 'ES II' measure indicated excellent internal responsiveness at 0.9 (95% CI: 0.75 - 1.10). For the follow-up condition only—when patients were most likely to be feeling an improvement in symptoms—patient ratings on the SPARS covered more than double the range of the transformed NRS. This aligns with the qualitative feedback that the SPARS allowed reporting of "the feeling that there was still some improvement to go even though they weren't really having pain anymore", indicating the superior perceptual scope conferred by the opportunity to provide ratings within a non-painful range, particularly as a person recovers from a painful condition. The comparative 'floor effect' imposed by the 'no pain' anchor on the conventional NRS was also seen in our previously published laboratory testing of the SPARS, which used laser stimuli to elicit non-painful and painful percepts in healthy volunteers [27]. The current study expands on that finding to reveal the implications of the NRS's floor effect for clinical utility: the SPARS avoids the interpretational difficulties that result from idiosyncratic interpretation or adding of anchors to the NRS, and provides a valued opportunity for rating of percepts that are not painful but not yet 'right'.

After using the SPARS in this study, several clinicians reported a preference for the SPARS over the NRS, either individually or as a team. Reasons for this included ease of use and interpretation, and that the non-painful range of the SPARS reminded both patient and clinician that treatment goals go beyond pain reduction. Clinicians pointed out that many patients still wish to report stiffness or other discomfort that does not qualify as 'pain' but still warrants clinical attention. Another application for graded reporting of non-painful events would be for side-to-side comparisons of provocation tests, such as palpation, joint or soft tissue mobilisations, or end-of-joint-range assessments. For example, deep palpation of a ligament can be quite uncomfortable—approaching the pain threshold on the SPARS—in a non-painful or 'unaffected' joint, but frankly painful on the 'affected' side. Here, allowing the patient to report an 'almost-painful' experience in an unaffected joint simultaneously avoids misinterpretation of the report as painful (if rated NRS>0) and provides feedback to the patient that almost-painful discomfort to palpation of a ligament is normal. Although we did not specifically compare the SPARS to the NRS for provocation tests, we would anticipate that it would perform well for this purpose. Despite these advantages, enthusiastic uptake of the SPARS is likely to be tempered by the requirements of some standardised reporting (e.g. for disability or insurance assessments) that the conventional NRS be used to report pain.

Patient-reported outcome measures should perform well on the criteria of face validity, responsiveness to change, discriminating power, and test-retest reliability – although a trade-off between these criteria may be necessary. The current study addressed face validity and responsiveness to change. Participants expressed satisfaction with the opportunity for self-expression provided by the SPARS, although qualitative data collected directly from the (patient) raters themselves would be even more convincing in this regard. Responsiveness to change was excellent. It would now be helpful to clarify test-retest reliability and minimum clinically meaningful change on the SPARS. With respect to test-retest reliability, the range of the SPARS may be influential. When used as a discrete scale, the SPARS offers 101 response options, and even more if used as a continuous scale. Some studies have reported that scales with more than 7-9 points offer too many response options for users to meaningfully distinguish between isolated stimuli, compromising test-retest reliability [29]. It is true that our own previous data showed considerable variance in within-participant ratings of experimental stimuli on both the SPARS and on a 101-point NRS [18; 27; 28]. However, the 7-9-option threshold for meaningful differentiation may apply only to unidimensional stimuli: studies of multidimensional experiences suggest a capacity to meaningfully distinguish between more response options [11; 31]. Given this controversy and that clinical pain is indisputably a multidimensional experience, determining the test-retest reliability of the SPARS will be an important priority going forward.

With respect to the amount of change in SPARS rating that reflects a clinically meaningful change, we are cautiously optimistic. Although Rasch-based data on the conventional NRS suggest that minimal clinically important difference varies across the NRS range [37], our experimental data showing a curvilinear stimulus-response curve for the SPARS [27], in contrast to the established power stimulus-response curve for the NRS, suggest that this weakness may not be true of the SPARS. Formal testing of this is an important focus for further study.

The original rationale for developing the SPARS was to give respondents an opportunity to provide graded reporting on non-painful events. The current data support the importance of this for the clinical context. These findings are particularly timely, given the ongoing work to develop patient-reported outcome measures using a broadly consultative, rigorous process that incorporates perspectives from people with lived experience of pain [21]. It is worth noting that the work by Langford et al. [21] seems to be limited to adapting existing legacy measures (thus constraining ratings within a range that grades painfulness) and, in its first phase, used a cross-sectional design to obtain data from patients with actively painful conditions. That study design would be unlikely to reveal any potential need for graded reporting on non-painful but clinically relevant percepts. In contrast, the current study's experimental-and-longitudinal design allowed for the possibility that actively painful conditions resolved into non-painful percepts during the study, and, perhaps unsurprisingly, found considerable support for a scale range that allows reporting of non-painful percepts, particularly for the later stages of recovery. Indeed, one clinician emphasised that a lower scale anchor resembling "completely normal" would be ideal for the clinical context. It is possible that this non-painful range is important only in the context of rehabilitation, but it seems more likely that it would be similarly valuable in other contexts of recovery, such as the peri-operative hospital context, primary health care, and return-to-work programmes. The current study's mixed methods approach was particularly helpful in that the qualitative data could confirm, contradict, or extend the impressions provided by the quantitative data, and this added nuance was useful in interpreting

the data. Given these benefits, we would support the continued use of mixed methods for future clinical testing of the SPARS, ideally in varied clinical contexts in which resolution of pain is anticipated.

The current data are subject to two other important limitations. First, the clinicians for the current study were personally known to author GLM, introducing a possible response bias in the service of social support. Future work would benefit from recruiting participants without an existing inter-personal relationship with a study leader. Second, we tested the SPARS in a context where most participants would have been first-language English-speakers, although we did not collect language data from our participants. The original SPARS anchors were developed for first-language English-speaking users; adaptation of these anchors will be necessary for other groups and may influence the utility of the adapted scale. Any adaptations to the anchor wording or application of the SPARS with people with a first language other than English will require careful testing.

Overall, we found that the SPARS performed well and was deemed relatively easy to use and interpret. Clinicians found that it addressed important shortfalls of the conventional NRS. The strong endorsement of the scale's structure by participating clinicians provides good rationale for further testing of the SPARS in different clinical contexts, including quantification of test-retest reliability and minimal clinically important difference.

Figure legends

Figure 1: Overall structure of the study. NRS ratings were requested in Phase 1 only; SPARS ratings were requested in Phases 1 and 2. Dice indicate that the order of assessment was randomised between participants. Randomisation was performed by the study team. Provocation tests and interventions were selected by clinicians, as appropriate to each patient's presentation. NRS: 0-10 Numerical Rating Scale for pain. SPARS: Sensation and Pain Rating Scale. At the follow-up time point, the phrasing of '2 days' was used only if less than one week had passed since the preceding study time point. Icons from the Noun Project: 'Twist' by James Keuning; 'Dice' by kareemov; 'Exercise' by Jeevan Kumar (CC BY 3.0).

Figure 2: Clinician reports on explaining, using, and interpreting the SPARS and NRS; data coloured by response. Endorsements of 'Neither easy nor difficult' are not shown. Response option 'Very difficult' was never endorsed.

Figure 3: Change in SPARS ratings with intervention. a) Change in ratings by individual patients. Each line represents the change in ratings from one patient; each patient underwent both the active and sham interventions. Lines are darkened to improve visibility of data for patients who reported a SPARS rating < 0. b) Tukey boxplot shows median, two quartiles, and extremes after excluding outliers as per the ggdist package. Each rating is represented by one dot.

Figure 4: Ratings on SPARS vs ratings on NRS, faceted and coloured by the condition under which ratings were given.

Figure 5: Bland-Altman (left) and simple correlation (right) plots of events rated SPARS ≥ 0, faceted by condition. Data coloured by condition; solid coloured line in correlation plots shows actual line of best fit. Black solid (stippled) line: average (95% CI) difference in measurements. Grey line: line of theoretically perfect fit.

Figure 6: Distribution of ratings of events rated in 'non-painful' range of the SPARS (SPARS < 0), faceted by scale. Average over last week, at follow-up: 72 non-painful events of 97 (74% of total events in this condition); Average over 2 days, at baseline: 12 non-painful events of 97 (12%); After active intervention: 9 non-painful events of 97 (9%).

Figure 7: SPARS ratings by study phase. Each dot represents one rating. Boxplot shows median and interquartile range.

Figure 8: SPARS ratings by study phase, faceted and coloured by condition under which ratings were given. Each dot represents one rating. Boxplot shows median and interquartile range.

Acknowledgements

Financial support: VJM is supported by the US National Institutes of Health on grant K43 TW011442. GLM and HBL are supported by an Australian NHMRC Investigator Grant awarded to G. Lorimer Moseley (ID 1178444).

Disclosures: VJM receives payment for lectures on pain and rehabilitation. PK receives consultancy fees from Partners in Research, South Africa, and is the sole proprietor of Blueprint Analytics. HBL has received speakers fees for lectures on pain and rehabilitation. LCH has received consultancy fees from Blue Note Therapeutics. GLM has received support from: Reality Health, ConnectHealth UK, Institutes of Health California, AIA Australia, Workers' Compensation Boards and professional sporting organisations in Australia, Europe, South and North America. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research on pain and pain education at scientific conferences/symposia. He has received speaker fees for lectures on pain, pain education and rehabilitation. He receives royalties for books on pain and pain education. He is non-paid CEO of the non-profit Pain Revolution and an unpaid Director of Painaustralia. The authors declare no other conflicts of interest.

References

[1] Allaire J, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. rmarkdown: Dynamic Documents for r, 2022.

[2] Auguie B. gridExtra: Miscellaneous Functions for "Grid" Graphics, 2017.

[3] Bates D, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015;67(1):1-48.

[4] Bedwell GJ, Louw C, Parker R, van den Broeke E, Vlaeyen JW, Moseley GL, Madden VJ. The influence of a manipulation of threat on experimentally-induced secondary hyperalgesia. PeerJ 2022;10:e13512.

[5] Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021;18(3):328-352.

[6] Bryman A. Integrating quantitative and qualitative research: how is it done? Qualitative Research 2006;6(1):97-113.

[7] Canty A, Ripley BD. boot: Bootstrap r (s-Plus) Functions, 2021.

[8] Chaves TC, Stanton TR, Grant A, Pulling BW, Madden VJ, Newport R, Moseley GL. Imprecise Visual Feedback About Hand Location Increases a Classically Conditioned Pain Expectancy Effect. J Pain 2021;22(6):748-761.

[9] Dancho M, and Davis Vaughan. tidyquant: Tidy Quantitative Financial Analysis, 2022.

[10] Davison AC, and D. V. Hinkley. Bootstrap Methods and Their Applications. Cambridge: Cambridge University Press, 1997.

[11] Euasobhon P, Atisook R, Bumrungchatudom K, Zinboonyahgoon N, Saisavoey N, Jensen MP. Reliability and responsivity of pain intensity scales in individuals with chronic pain. Pain 2022;163(12):e1184-e1191.

[12] Firke S. janitor: Simple Tools for Examining and Cleaning Dirty Data, 2021.

[13] Gohel D. flextable: Functions for Tabular Reporting, 2022.

[14] Gordon M. Gmisc: Descriptive Statistics, Transition Plots, and More, 2022.

[15] Heinzen E, Jason Sinnwell, Elizabeth Atkinson, Tina Gunderson, and Gregory Dougherty. arsenal: An Arsenal of "R" Functions for Large-Scale Statistical Summaries, 2021.

[16] Ho M, Nguyen J, Talbot K, Heales L, Kean C, Kong PW, Stanton R. Immediate comfort perception of 3D-printed foot orthoses in individuals with unilateral heel pain. Prosthet Orthot Int 2022;46(1):31-36.

[17] Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53(5):459-468.

[18] Kamerman P, Madden V, Bellan V, Catley M, Russek L, Camfferman D, Moseley L. Analysis scripts and supplementary files: Was that painful or non-painful? The Sensation and Pain Rating Scale (SPARS) performs well in the experimental context: figshare, 2018.

[19] Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots, 2020.

[20] Kay M. ggdist: Visualizations of Distributions and Uncertainty, 2022.

[21] Langford DJ, Gewandter JS, Amtmann D, Reeve BB, Corneli A, McKenna K, Swezey T, McFatrich M, Jensen MP, Turk DC, Dworkin RH. Initial Content Validation and Roadmap for a New Patient-Reported Outcome Measure of Pain Intensity. J Pain 2022;23(11):1945-1957.

[22] Lawrence IKL. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 1989;45(1):255-268.

[23] Lawrence IKL. Correction: A Note on the Concordance Correlation Coefficient. Biometrics 2000;56(1):324-325.

[24] Littman GS, Walker BR, Schneider BE. Reassessment of verbal and visual analog ratings in analgesic studies. Clin Pharmacol Ther 1985;38(1):16-23.

[25] Madden VJ, Bellan V, Russek LN, Camfferman D, Vlaeyen JWS, Moseley GL. Pain by Association? Experimental Modulation of Human Pain Thresholds Using Classical Conditioning. J Pain 2016;17(10):1105-1115.

[26] Madden VJ, Kamerman P, Moseley GL. The Sensation and Pain Rating Scale project: clinical testing phase 1: OSF, 2023.

[27] Madden VJ, Kamerman PR, Bellan V, Catley MJ, Russek LN, Camfferman D, Moseley GL. Was That Painful or Nonpainful? The Sensation and Pain Rating Scale Performs Well in the Experimental Context. J Pain 2019;20(4):472 e471-472 e412.

[28] Madden VJ, Kamerman PR, Catley MJ, Bellan V, Russek LN, Camfferman D, Lorimer Moseley G. Variability in experimental pain studies: nuisance or opportunity? Br J Anaesth 2021;126(2):e61-e64.

[29] Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 1994;101(2):343-352.

[30] Pedersen TL. patchwork: The Composer of Plots, 2020.

[31] 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):1-15.

[32] Rodríguez-Sánchez F, Jackson CP, Hutchins SD. grateful: Facilitate citation of R packages, 2023.

[33] Stevenson M, Sergeant E. epiR: Tools for the Analysis of Epidemiological Data, 2022.

[34] Team R. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Inc., 2016.

[35] Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2017.

[36] Traxler J, Madden VJ, Moseley GL, Vlaeyen JWS. Modulating pain thresholds through classical conditioning. PeerJ 2019;7:e6486.

[37] Walton DM, Elliott JM, Salim S, Al-Nasri I. A reconceptualization of the pain numeric rating scale: Anchors and clinically important differences. J Hand Ther 2018;31(2):179-183.

[38] Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. Theriogenology 2010;73(9):1167-1179.

[39] Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. J Open Source Softw 2019;4(43).

[40] Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2", 2020.

[41] Xie Y. knitr: A Comprehensive Tool for Reproducible Research in R. In: FL Victoria Stodden, and Roger D. Peng, editor. Implementing Reproducible Computational Research: Chapman; Hall/CRC, 2014.

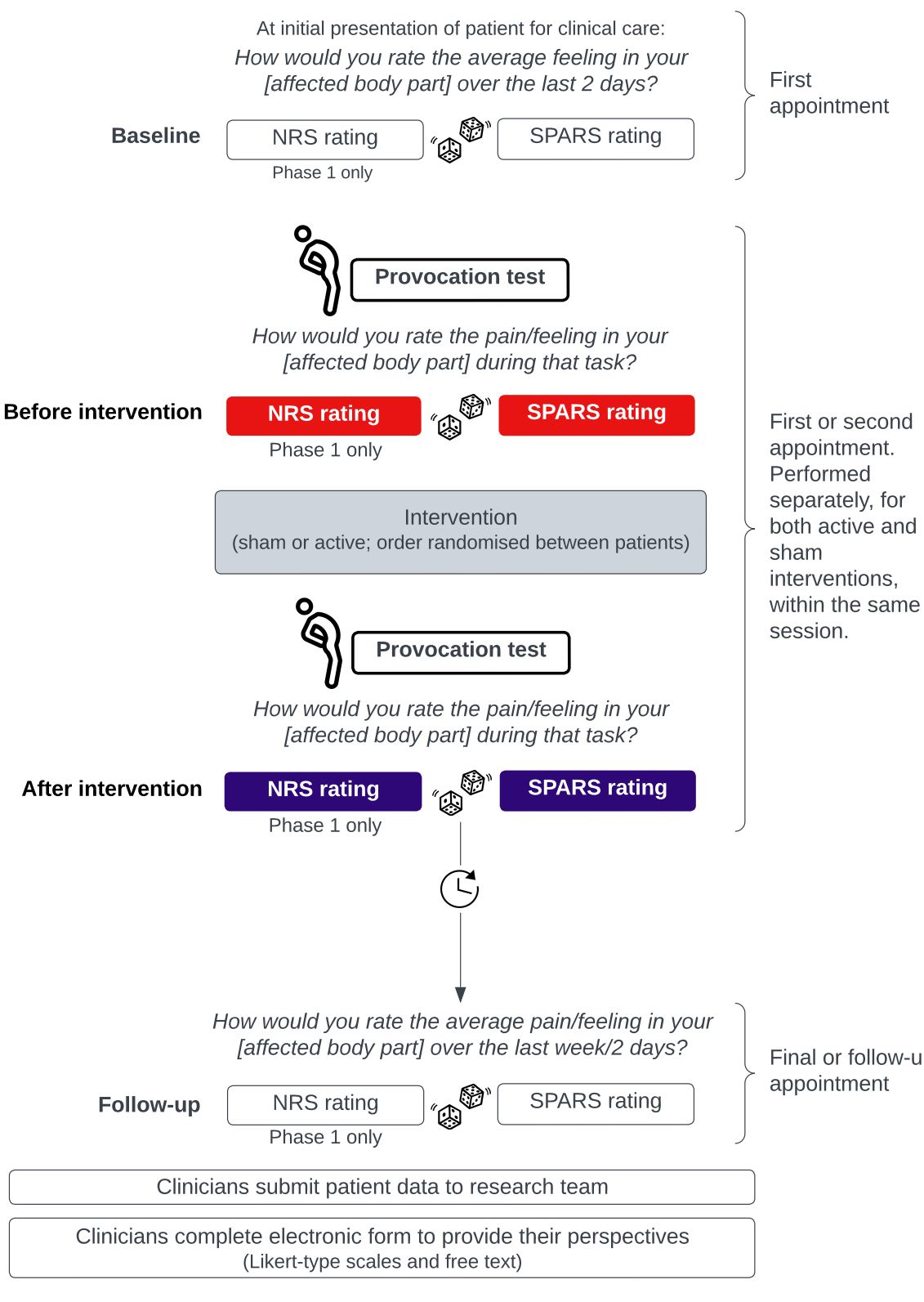
[42] Xie Y. Dynamic Documents with R and Knitr. In: B Raton, editor. Florida: Chapman; Hall/CRC, 2015.

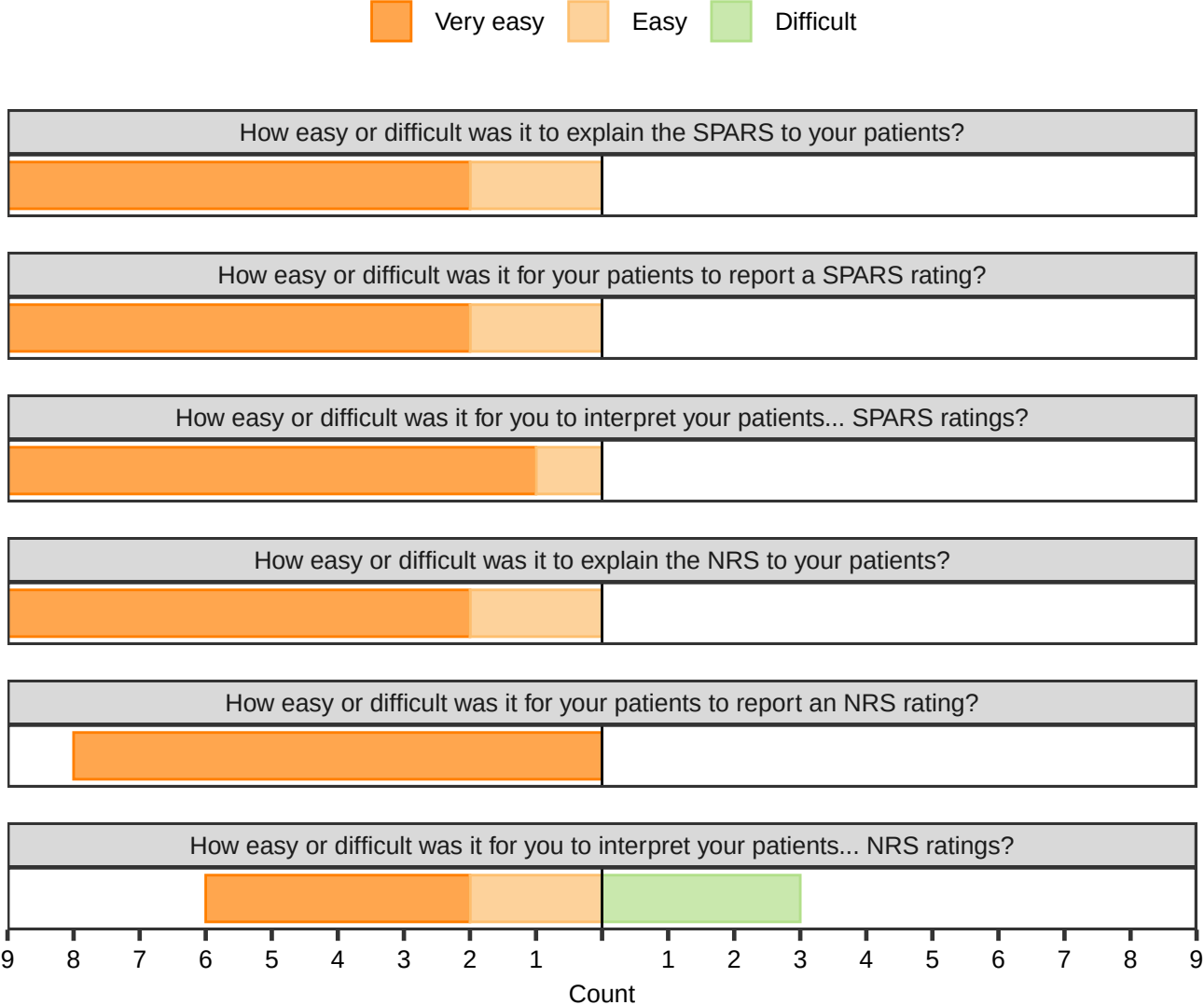
[43] Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in r, 2022.

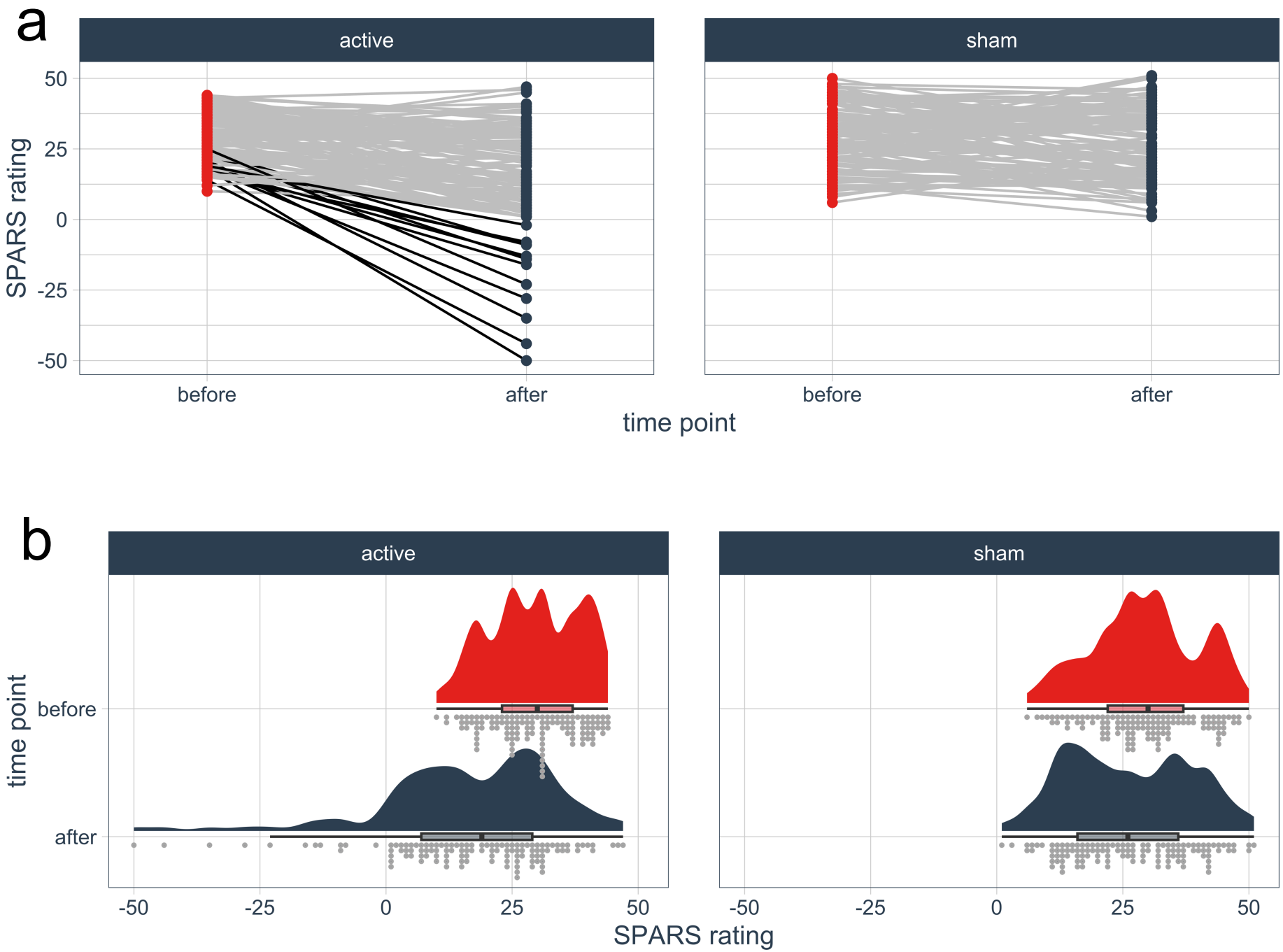
[44] Xie Y, Christophe Dervieux, and Emily Riederer. R Markdown Cookbook. In: B Raton, editor. Florida: Chapman; Hall/CRC, 2020.

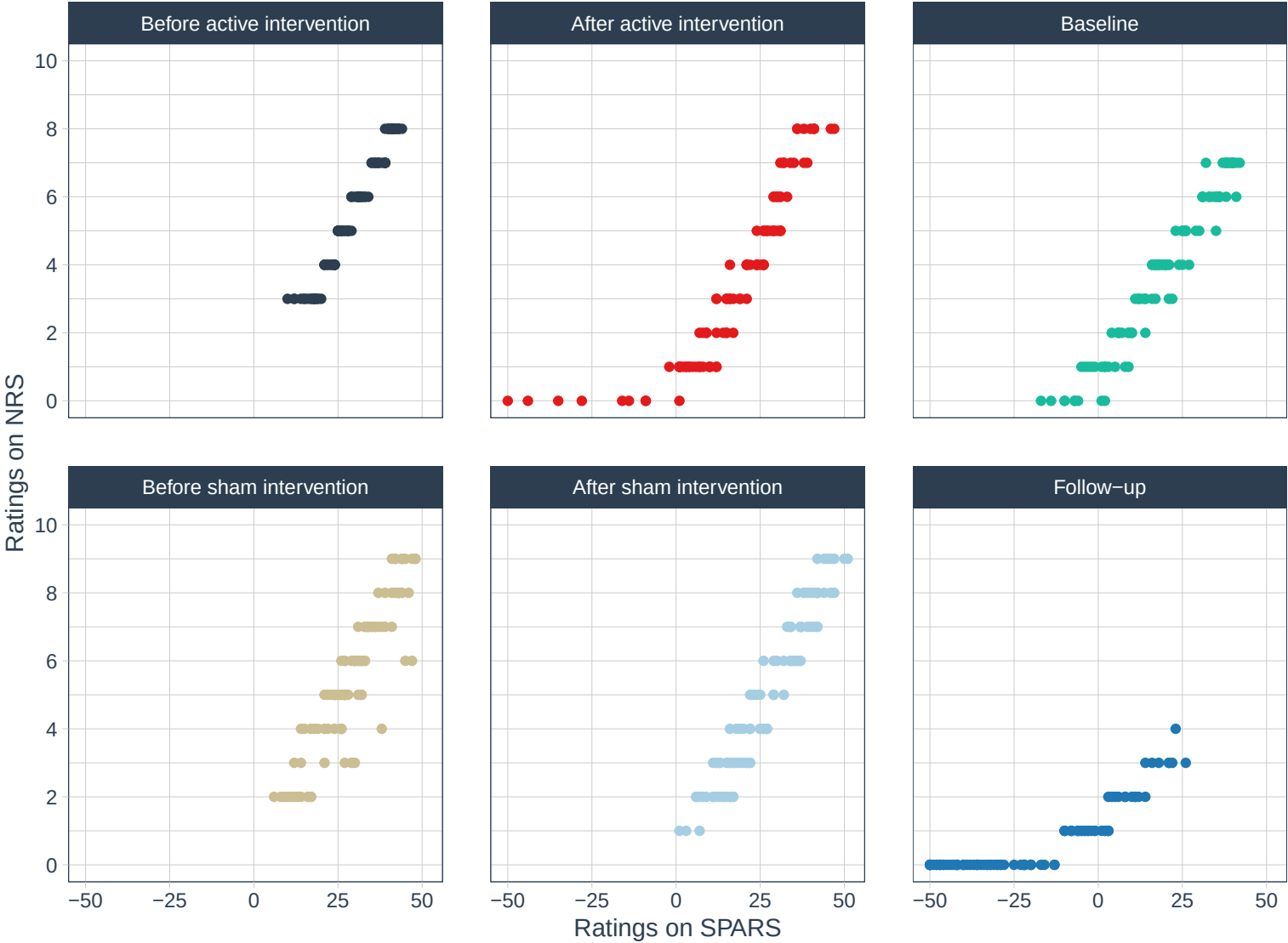
[45] Xie Y, J. J. Allaire, and Garrett Grolemund. R Markdown: The Definitive Guide. In: B Raton, editor. Florida: Chapman; Hall/CRC, 2018.

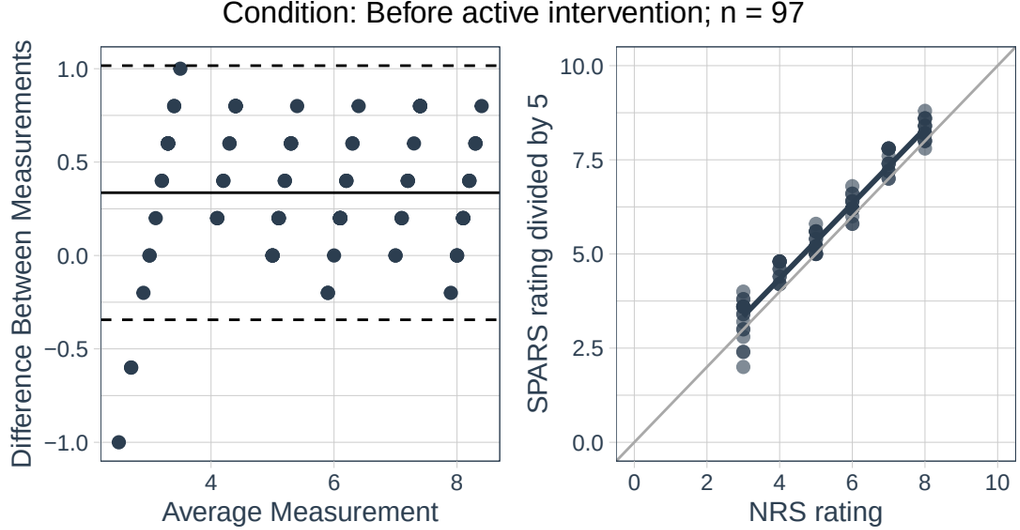
[46] Zhu H. kableExtra: Construct Complex Table with "kable" and Pipe Syntax, 2021.

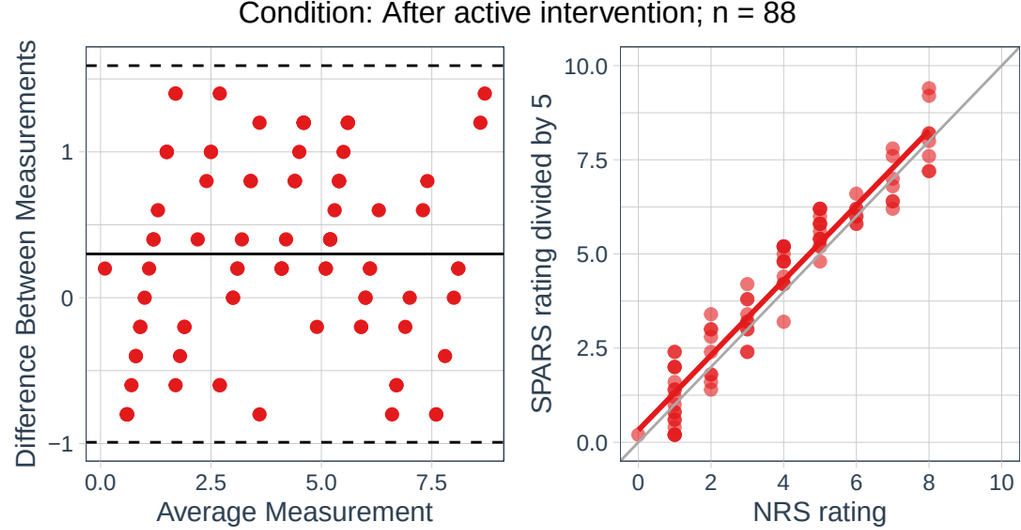

At initial presentation of patient for clinical care:
How would you rate the average feeling in your [affected body part] over the last 2 days?
Baseline
NRS rating
SPARS rating
Phase 1 only
First appointment
Provocation test
How would you rate the pain/feeling in your [affected body part] during that task?
Before intervention
NRS rating
SPARS rating
Phase 1 only
Intervention
(sham or active; order randomised between patients)
Provocation test
How would you rate the pain/feeling in your [affected body part] during that task?
After intervention
NRS rating
SPARS rating
Phase 1 only
First or second appointment. Performed separately, for both active and sham interventions, within the same session.
How would you rate the average pain/feeling in your [affected body part] over the last week/2 days?
Follow-up
NRS rating
SPARS rating
Phase 1 only
Final or follow-up appointment
Clinicians submit patient data to research team
Clinicians complete electronic form to provide their perspectives
(Likert-type scales and free text)

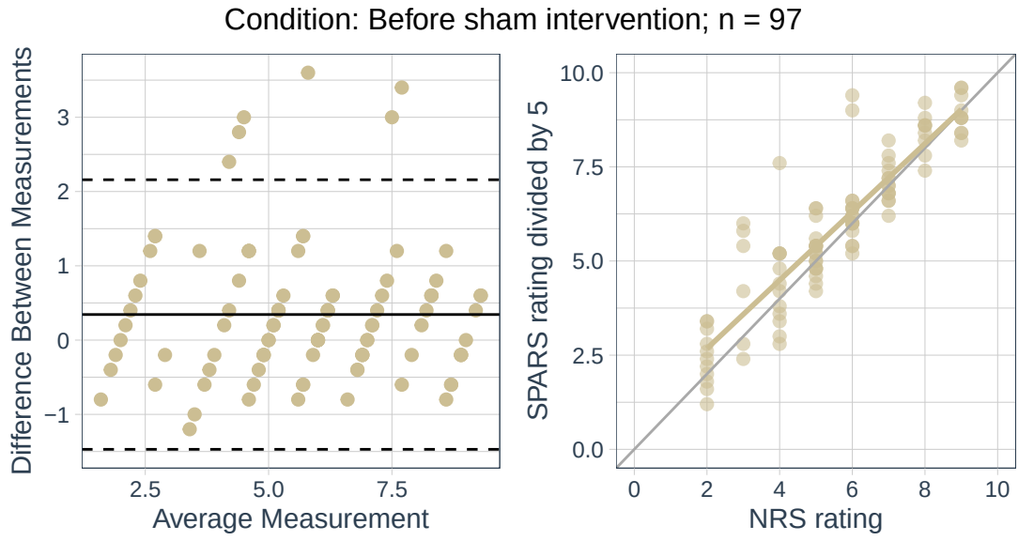

Very easy
Easy
Difficult
How easy or difficult was it to explain the SPARS to your patients?
How easy or difficult was it for your patients to report a SPARS rating?
How easy or difficult was it for you to interpret your patients... SPARS ratings?
How easy or difficult was it to explain the NRS to your patients?
How easy or difficult was it for your patients to report an NRS rating?
How easy or difficult was it for you to interpret your patients... NRS ratings?
9 8 7 6 5 4 3 2 1 1 2 3 4 5 6 7 8 9
Count

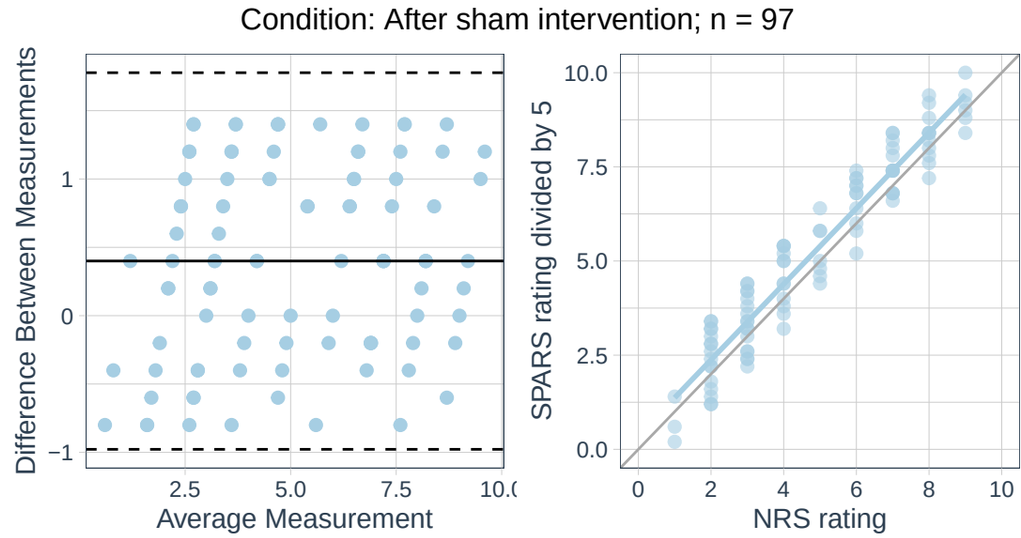

a
active
sham
SPARS rating
50
25
0
-25
-50
before
after
time point
b
active
sham
time point
before
after
-50
-25
0
25
50
SPARS rating

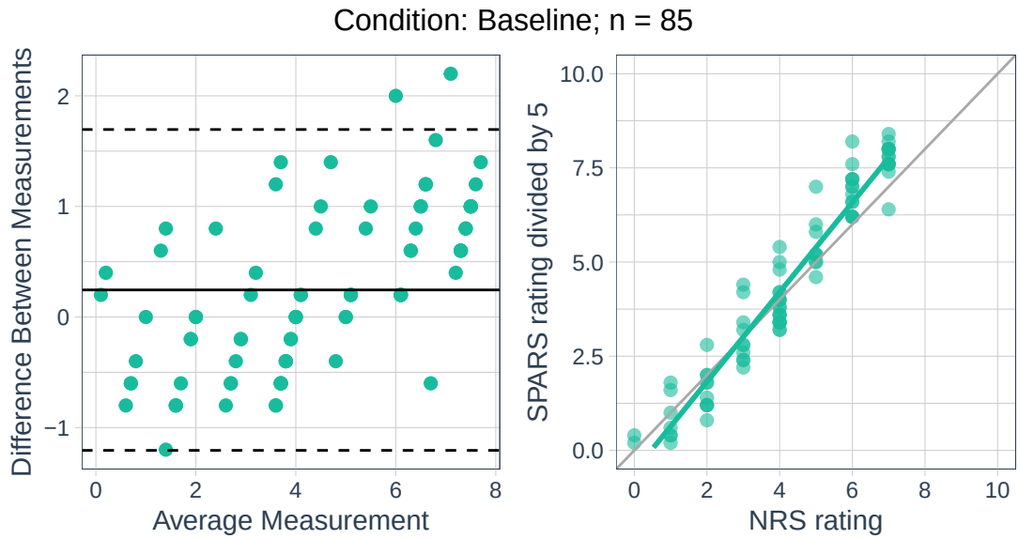

Before active intervention
After active intervention
Baseline
Before sham intervention
After sham intervention
Follow−up
Ratings on NRS
Ratings on SPARS
0
2
4
6
8
10
−50
−25
0
25
50

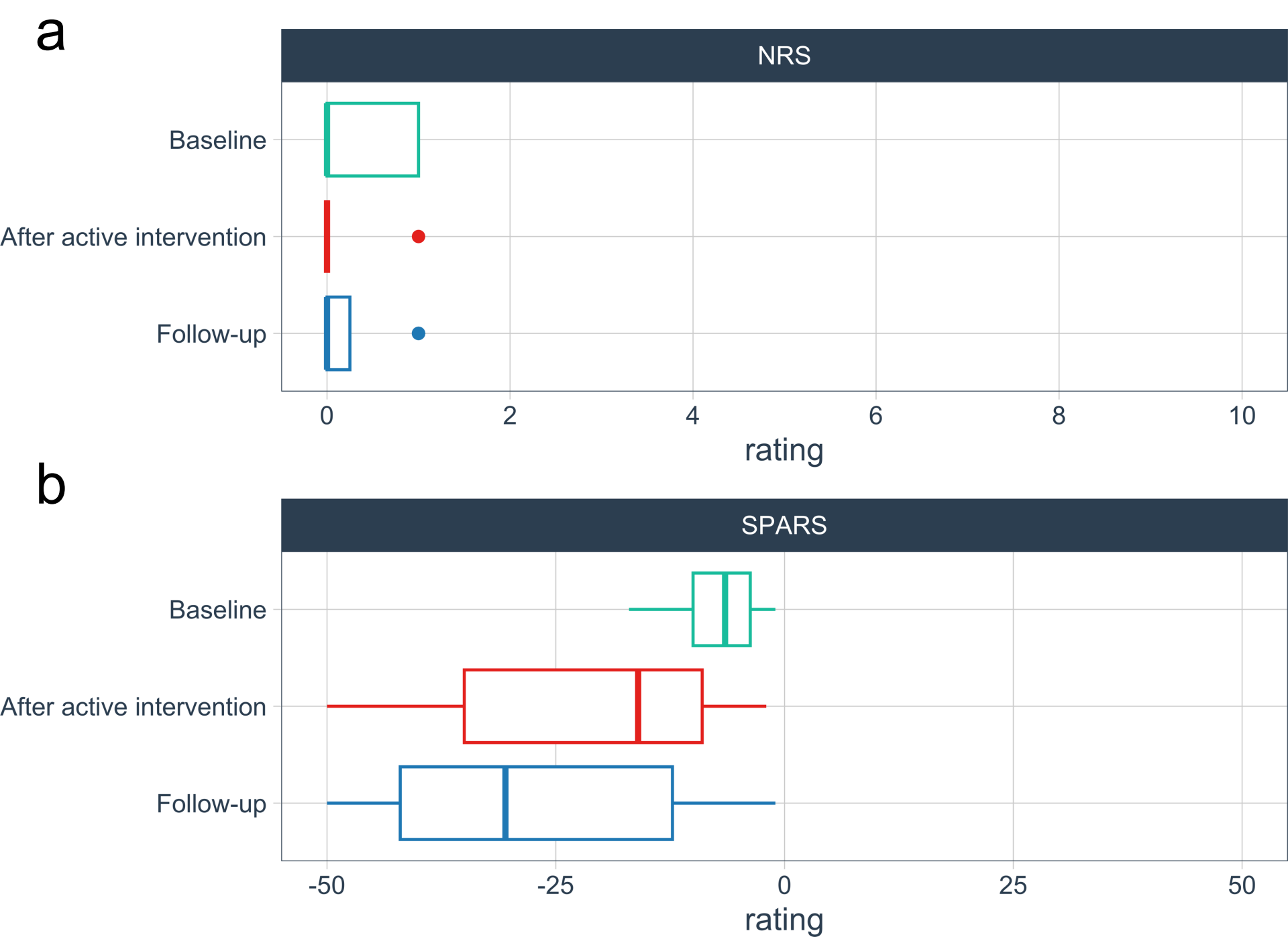

Condition: Before active intervention; n = 97
Difference Between Measurements
Average Measurement
SPARS rating divided by 5
NRS rating

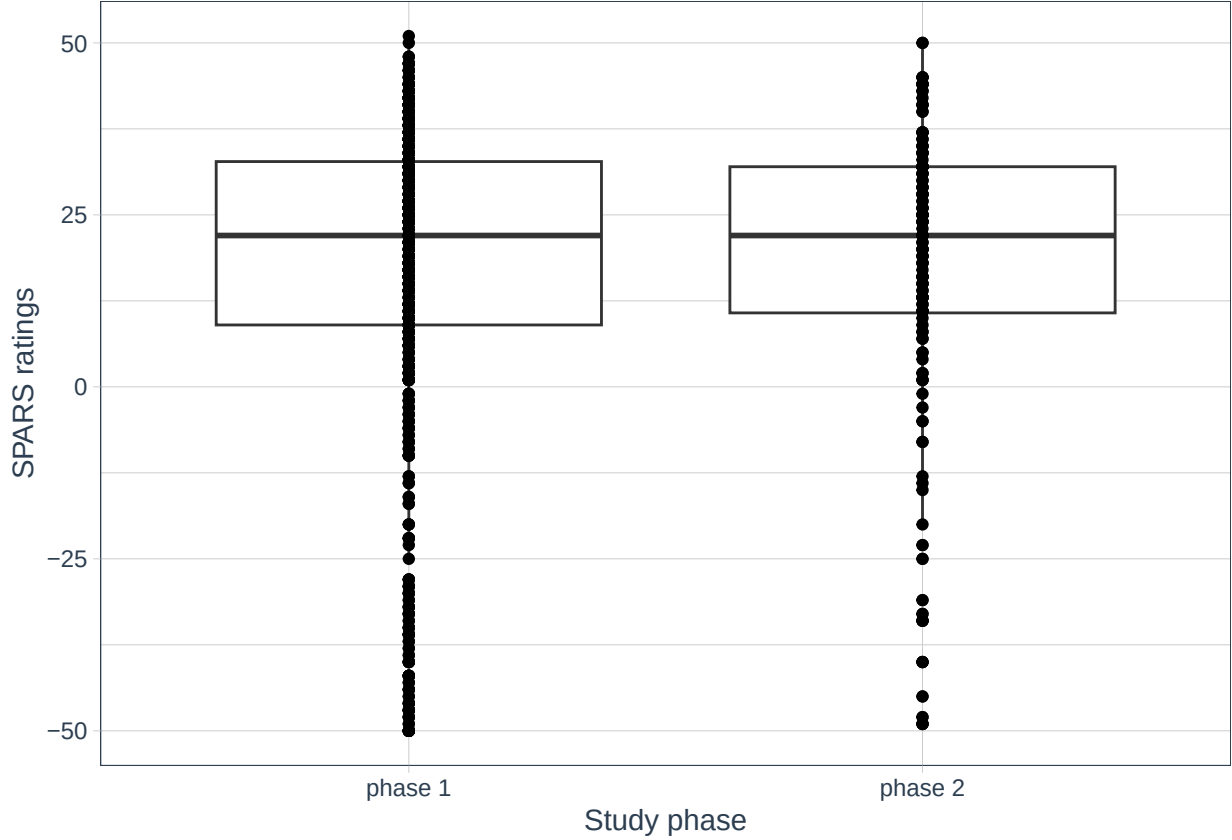

Condition: After active intervention; n = 88
Difference Between Measurements
Average Measurement
SPARS rating divided by 5
NRS rating


Condition: Before sham intervention; n = 97
Difference Between Measurements
Average Measurement
SPARS rating divided by 5
NRS rating


Condition: After sham intervention; n = 97
Difference Between Measurements
Average Measurement
SPARS rating divided by 5
NRS rating


Condition: Baseline; n = 85
Difference Between Measurements
Average Measurement
SPARS rating divided by 5
NRS rating


a
NRS
Baseline
After active intervention
Follow-up
0
2
4
6
8
10
rating
b
SPARS
Baseline
After active intervention
Follow-up
-50
-25
0
25
50
rating


50
25
0
−25
−50
SPARS ratings
phase 1
phase 2
Study phase

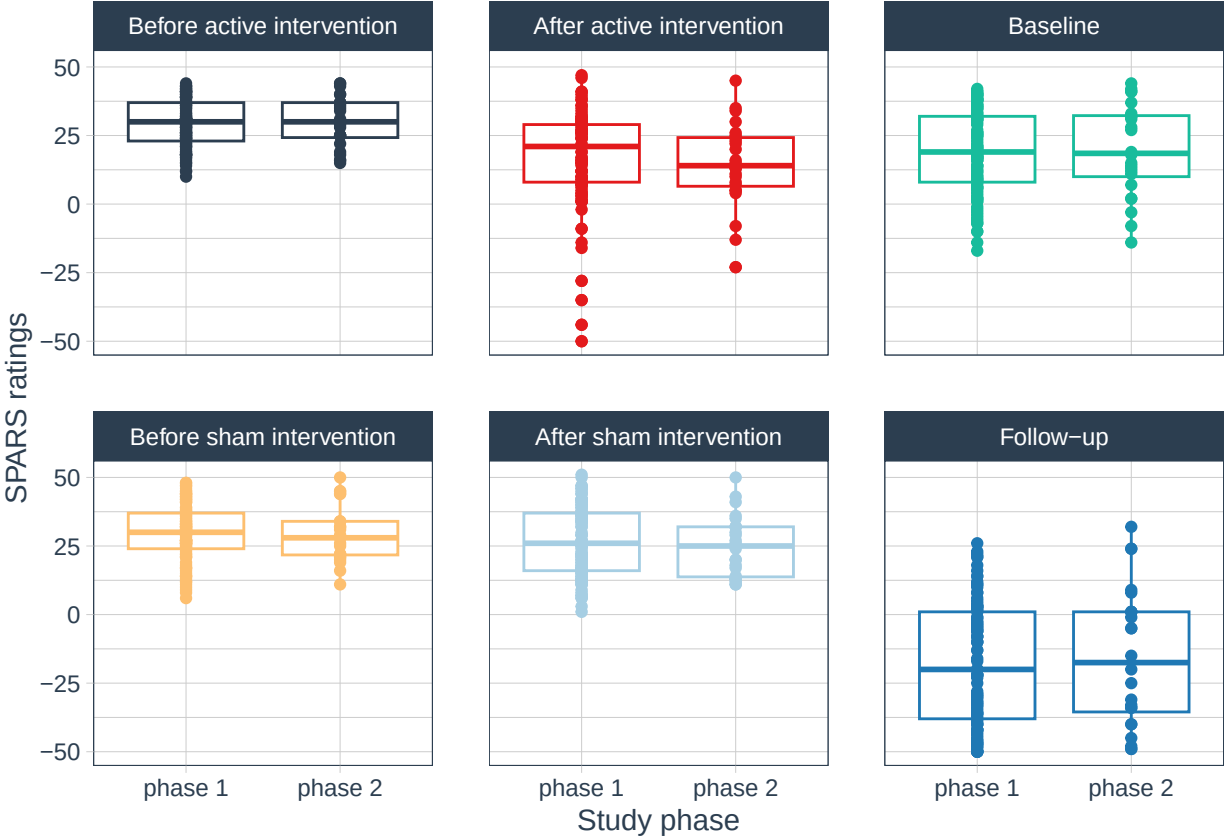

Before active intervention
After active intervention
Baseline
Before sham intervention
After sham intervention
Follow−up
SPARS ratings
50
25
0
−25
−50
phase 1
phase 2
Study phase

Table 1: Patient information by study phase. CRPS: Complex regional pain syndrome.

	phase 1 (N=97)	phase 2 (N=24)	Total (N=121)
Gender			
Female	45 (46.4%)	17 (70.8%)	62 (51.2%)
Male	48 (49.5%)	7 (29.2%)	55 (45.5%)
Other	4 (4.1%)	0 (0.0%)	4 (3.3%)
Age (yrs)			
Mean (SD)	41 (10)	40 (9)	41 (10)
Range	18 - 66	22 - 64	18 - 66
Diagnosis			
CRPS	7 (7.2%)	2 (8.3%)	9 (7.4%)
Limb	26 (26.8%)	4 (16.7%)	30 (24.8%)
Pelvic	7 (7.2%)	2 (8.3%)	9 (7.4%)
Spinal	38 (39.2%)	12 (50.0%)	50 (41.3%)
Widespread	19 (19.6%)	4 (16.7%)	23 (19.0%)

Table 2: Interventions used and need for ongoing treatment.

	Overall (N=121)
Order in which scales were used	
NRS then SPARS	61 (50.4%)
SPARS then NRS	60 (49.6%)
Provocation test	
functional	71 (58.7%)
manual therapy	11 (9.1%)
movement	25 (20.7%)
pressure	14 (11.6%)
Active intervention	
N missing	17
active exercise	28 (26.9%)
education/advice	50 (48.1%)
manual therapy	26 (25.0%)
Sham intervention	
N missing	25
active exercise	26 (27.1%)
education/advice	25 (26.0%)
manual therapy	45 (46.9%)
Days from baseline to follow-up time point	
Mean (SD)	32 (14)
Median	30
25th, 75th percentile	22, 42
Range	5 - 62
Do you think you still need treatment?	
no	44 (36.4%)
yes	77 (63.6%)
Do you think you have completed your recovery?	
no	107 (88.4%)
yes	14 (11.6%)

Table 3: Concordance correlation coefficients (CCC) for each condition when SPARS data were divided by 5.

Condition	CCC (95% CI)
Before active intervention	0.96 (0.95 - 0.97)
After active intervention	0.95 (0.93 - 0.97)
Before sham intervention	0.89 (0.84 - 0.92)
After sham intervention	0.95 (0.92 - 0.96)
Baseline	0.94 (0.92 - 0.96)

Table 4: Joint visual display of quantitative, qualitative, and mixed-methods meta-inferences by topic.

Topic	**Quantitative result**	**Qualitative excerpts**	**Meta-inference about complementarity**
Ease of explanation	Both the SPARS and NRS were rated as easy or very easy to explain	"[The SPARS] is very easy to explain" (46) "I found it difficult at first to adjust to the SPARS because I am so used to using the zero to ten scale" (42) "I think they are equally easy to use" (08) "Really [the SPARS] is no harder to use than the usual NRS, just less familiar." (29)	*Expansion* Clinicians described an initial period of familiarisation before use of the SPARS was easy; this was not identified in the Likert scale ratings.
Ease of reporting	All clinicians but one reported that it was very easy or easy for patients to report on the SPARS or the NRS; one clinician provided an ambiguous response for the NRS.	[On the SPARS:] "I think the explanation is easy and patients get it" (46) "All my patients got the hang of [the SPARS] quickly but initially it was not as intuitive as the 0 - 10 scale (32) "Once they understand that it is only positive numbers that are about pain then [the SPARS] is easy as pie" (29)	*Expansion* Patients needed familiarisation before reporting on the SPARS was easy; this was not identified in the Likert scale ratings.
Ease of interpreting	The SPARS was easy or very easy to interpret. In contrast, the NRS was noted as difficult to interpret by three of the nine clinicians (25, 42, 69), whereas the remaining clinicians rated it as easy or very easy to interpret.	"[The SPARS] is way more informative to us" (42) "[Interpretation] is where I think the SPARS is better than our other methods of assessing pain" (46) "The SPARS is much easier… because it makes very clear where the pain threshold is" (42) "[On the SPARS] zero really was zero and it was particularly interesting when patients were about zero and were really evaluating things" (29)	*Confirmation* Interpreting the SPARS was viewed as easier than interpreting the NRS and other conventional assessments of pain.
Does the SPARS provide quantification over a greater perceptual range than a 0–10 Pain NRS?	NRS scores transformed to a 100-point scale had greater spread than SPARS scores in 4 of the 6 conditions. However, the largest difference in score range was for average pain rating over the week preceding the follow-up consultation, where SPARS ratings covered more than twice the range of transformed NRS ratings.	Allows patients to express sensation that is 'not quite right', although not painful: *"This is where I think the SPARS is better than our other methods of assessing pain to be honest with you… you don't have no sensation at all and then suddenly pain, and I know some patients who talk about being sore but not painful, for example."* (46) *"…[the SPARS] allowed them to communicate the feeling that there was still some improvement to go even though they weren't really having pain anymore."* (29) *"I think [the SPARS] is more useful as people get better" (29)*	*Expansion* The non-painful range of the SPARS may be most valuable during the later stages of recovery from a painful episode, when its ability to capture percepts that are non-painful but also not 'normal', may be clinically informative.