**APPENDIX 1. Numeric Rating Scale recommended as core outcome measurement instruments to measure pain intensity in non-specific low back pain clinical trials**

**How would you rate your average low back pain intensity over the last week?**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

No pain

Worst imaginable pain

**APPENDIX 2. Evidence synthesis from systematic reviews on measurement properties of potential core outcome measurement instruments in patients with low back pain**

**A. Physical functioning patient-reported outcome measures**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PROMs | Content validity | | | | | | Structural validity | |
| **Relevance** | | **Comprehensiveness** | | **Comprehensibility** | |
| **Rating** | **Quality of evidence** | **Rating** | **Quality of evidence** | **Rating** | **Quality of evidence** | **Rating** | **Quality of evidence** |
| RMDQ-24 | + | Very Low | – | High | + | High | – | High |
| RMDQ-23 | + | Very Low | – | Very Low | + | Very Low | ± | Moderate |
| RMDQ-18 | + | Very Low | – | Very Low | + | Very Low | – | High |
| ODI 1.0 | ± | Very Low | – | Very Low | + | Very Low | – | High |
| ODI 2.1a | ± | Very Low | – | Very Low | + | Very Low | ± | Moderate |
| CLBPDQ | ± | Very Low | – | Very Low | + | Very Low |  |  |
| MLBPDQ | ± | Very Low | – | Very Low | + | Very Low |  |  |
| BPI-PI | ± | Very Low | – | Very Low | + | Very Low | + | Moderate |
| MPI-PI | ± | Very Low | – | Very Low | + | Very Low | ± | Moderate |
| SF36-PF | + | Very Low | – | Very Low | + | Very Low | – | Moderate |
| LBPRS-DI | ± | Very Low | – | Very Low | + | Very Low |  |  |
| QBPDS | + | Low | + | Very Low | + | Very Low | ± | Moderate |
| PROMIS-PF-4 | + | Low | – | Very Low | + | Low |  |  |
| PROMIS-PF-6 | + | Low | – | Very Low | + | Low |  |  |
| PROMIS-PF-8 | + | Low | – | Very Low | + | Low |  |  |
| PROMIS-PF-10 | + | Low | + | Very Low | + | Low |  |  |
| PROMIS-PF-20 | + | Low | + | Very Low | + | Low |  |  |
| Empty cells represent measurement properties not assessed in any study. + = sufficient results; – = insufficient results; ± = inconsistent results; ? = indeterminate.  The other measurement properties are not displayed because they were not assessed in this systematic review. | | | | | | | | |

**B. Pain intensity patient-reported outcome measures**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measurement properties | | | VAS | NRS | BPI-PS |
| Content validity | **Relevance** | **Rating** | ± | ± | ± |
| **Quality of evidence** | Low | Low | Low |
| **Comprehensiveness** | **Rating** | ± | ± | ± |
| **Quality of evidence** | Low | Low | Low |
| **Comprehensibility** | **Rating** | + | + | + |
| **Quality of evidence** | Very Low | Very Low | Very Low |
| Structural validity | | **Rating** | na | na | + |
| **Quality of evidence** |  |  | Moderate |
| Internal consistency | | **Rating** | na | na | + |
| **Quality of evidence** |  |  | Moderate |
| Test-retest reliability | | **Rating** | + | ± |  |
| **Quality of evidence** | Very Low | Low |  |
| Measurement error | | **Rating** | ± | – |  |
| **Quality of evidence** | Very Low | High |  |
| Construct validity | | **Rating** | ± | ± | ± |
| **Quality of evidence** | Low | Very Low | Moderate |
| Responsiveness | | **Rating** | ± | ± | ± |
| **Quality of evidence** | Low | Moderate | Very Low |
| Empty cells represent measurement properties not assessed in any study.  “+” = sufficient results; “‒” = insufficient results; “±” = inconsistent results; na = not applicable.  The cross-cultural validity row is not displayed because it was not assessed in any study. | | | | | |

**C. Health-related quality of life patient-reported outcome measures**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measurement properties | | | SF36 | | | | SF12 | | EQ-5D | | | NHP | PROMIS-GH-10 | |
| **Sub-scales** | **PCS** | **MCS** | **TOT** | **PCS** | **MCS** | **Utility** | **VAS** | **Domains** | **Sub-scales** | **PCS** | **MCS** |
| Content validity | **Relevance** | **Rating** | + | | | | + | | + | | | + | + | |
| **Quality of Evidence** | Very Low | | | | Very Low | | Very Low | | | Very Low | Very Low | |
| **Comprehensiveness** | **Rating** | + | | | | + | | − | | | + | + | |
| **Quality of Evidence** | Very Low | | | | Very Low | | Very Low | | | Very Low | Very Low | |
| **Comprehensibility** | **Rating** | + | | | | + | | + | | | + | + | |
| **Quality of Evidence** | Very Low | | | | Very Low | | Very Low | | | Very Low | Very Low | |
| Construct validity | | **Rating** | ± | ± | ± | + | ± | ± | − | − | ± | + |  |  |
| **Quality of Evidence** | Low | Moderate | Moderate | Low | Low | Low | High | High | Low | Low |  |  |
| Responsiveness | | **Rating** | ± | ± | ± | + | ± | – | ± | ± |  |  |  |  |
| **Quality of Evidence** | Very Low | Low | Low | Low | Very Low | Low | Moderate | Low |  |  |  |  |
| Empty cells represent measurement properties not assessed in any study.  PCS = Physical Component Summary score; MCS = Mental Component Summary score; TOT = Total score; VAS = Visual Analogue Scale score.  “+” = sufficient results; “‒” = insufficient results; “±” = conflicting results.  The following measurement properties were assessed only for one instrument and their evidence synthesis is not reported in the table: structural validity of SF-36 PCS and MCS was inconsistent (low quality evidence), test-retest reliability of SF-36 subscales was sufficient (low quality evidence), measurement error of SF-12 PCS and MCS was sufficient (low quality evidence). Internal consistency was assessed for the SF-36 subscales, SF-36 TOT, and for SF-12 PCS and MCS; the results of these studies are presented in the text but they do not represent for this measurement property, as the unidimensionality of the assessed tools has not been tested in patients with low back pain. Cross-cultural validity was not assessed for any instrument. | | | | | | | | | | | | | | |

**APPENDIX 3. Anonymous comments made by participants of the second Delphi round against recommending the Oswestry Disability Index version 2.1a as the only core outcome measurement instrument for physical functioning in low back pain**

|  |  |
| --- | --- |
| **n** | **comment** |
| 1 | *I think any fee is a barrier.* |
| 2 | *I am still worried about the fees in a core outcome measurement instrument.* |
| 3 | *I do have a problem with fees being charged for academic use, where no financial gain is expected from the research. We are essentially expecting funders (such as charities) to pay money to the questionnaire developers (i.e. into their own pockets).* |
| 4 | *Core set instruments should definitely be free to use if they are supposed to be endorsed and then possibly used in a large majority of studies.* |
| 5 | *Obviously, this is the state-of-affairs as-is, meaning that if whatever circumstances change, either higher fees may apply, or accessibility may seize for whatever reasons. So, although the absolute cost should not be restrictive, I am much more cautious about any other future implication, because of copyright protectiveness of the instrument.* |
| 6 | *This costing framework remains unacceptable. Whilst the cost for academic research is modest and would not be a problem for those of us working with well-funded research, it will remain a problem for others. I think the point of principal must stand that we should not recommend any measure attracting a fee for use. There is no guarantee that this fee structure will remain if ODI becomes part of core outcome set when we will all be forced to use it. In my view is not in The Pythia, then at the very least we need to approach the copyright holders and ensure there are no charges for researchers in LMICs. I make a general point here that if we are considering recommending measures that might attract a fee for there needs to be a conflict of interest policy. We cannot be in a position that one of the developers of measure who might stand to gain financially from use of a measure is contributing to this exercise. I further suggest that best practice would be to ask outcome developers without a financial stake to recuse themselves from discussion on that domain.* |
| 7 | *I will not pay for the use of any measurement which is developed with public money (most of them, or all are…). I am really strongly against any form of payment for the use of any questionnaire whatsoever.* |
| 8 | *What about in developing countries? Might still pose a problem if there were fees. Generally I think core sets should be freely available so this is still a little tick against for me.* |
| 9 | *Costs which seem affordable in Europe or the US may not be abroad. This is not a reason for not endorsing ODI, but a rational to ensure that at least one instrument which can be used for free is (also endorsed).* |
| 10 | *Must be free of charge. Content validity needs to be demonstrated (qualitative studies with patients).* |
| 11 | *I am against using any outcomes which are not free but maybe I can accept ODI 2.1a despite of this substantial drawback.* |