CANADIAN CHIROPRACTIC PRACTICE-BASED RESEARCH NETWORK (PBRN) USAGE OF PATIENT-REPORTED OUTCOME MEASURES (PROMs)

Introduction

The purpose of this brief report is to provide recommendations for the use of Patient-Reported Outcome Measures (PROMs) for the Canadian chiropractic Practice-Based Research Networks (PBRNs) in future research projects. Employing a common set of PROMs could allow for pooling and comparison of data across PBRNs. PROMs offer numerous benefits for researchers, clinicians, and patients. They provide valuable clinical information for patients and clinicians to reflect upon and a means to track patient progress through care. For researchers valid, reliable, and responsive PROMs can serve as primary and/or secondary measures for outcome studies in particular. Patient-Reported Experience Measures (PREMs), particularly patient satisfaction with care, should also be recorded. Finally, a common set of additional clinical information such as demographic and other clinical data (such as medication usage, duration of complaint, etc) should be collected. Data collection and pooling may be simplified by use of a common online data collection tool, e.g. CareResponse (https://www.care-response.com).

To identify commonly used PROMs and PREMs, a rapid review of the literature was conducted. Several systematic reviews were identified and results summarized to inform this report and its recommendations. We also considered the Patient-Reported Outcome Measurement Information System (PROMIS), which provides clinicians and researchers with a valid, reliable, and useful set of health measures.

The final limited set of recommended PROMs were selected based on the following criteria:

   a) results of a survey of PBRN team leaders and determination of the PROMs that current PBRNs are using or planning to use;

   b) appropriate psychometric properties;

   c) reported frequency of use in research and in clinical practice; and

   d) ease by clinicians in practice including 'open source' or freely available instruments

Summary of Recommendations and Conclusions

The CCGI recommends using the following PROMs and PREMs for inclusion in future projects undertaken by Canadian chiropractic PBRNs:

1. *Neck Disability Index (NDI) for function in neck pain patients*
   - The NDI is the most commonly used neck pain specific outcome measure in practice and in research. It is also the only neck pain specific outcome measure being used by any of the PBRNs (4/7).
   - The NDI has well documented acceptable psychometric properties, although there are concerns about responsiveness and whether or not it is uni-dimensional.
   - There is another issue of familiarity, Canadian chiropractors should know the NDI well and in some jurisdictions using it is a requirement for 3rd party insurers.
   - Of the other measures, the Neck Bournemouth Questionnaire potentially shows the most promise from a psychometric standpoint, but is not as well known or as commonly used in practice or research.

2. *The Oswestry Disability Index (ODI) for function in low back pain patients*
   - The ODI is currently the most commonly used low back pain specific outcome measure in practice and in research. It is also the most common back pain specific outcome measure being used by any of the PBRNs (5/7).
   - The ODI has well documented acceptable psychometric properties.
   - There is again the issue of familiarity, Canadian chiropractors should know the ODI well and in some jurisdictions using it is a requirement for 3rd party insurers.
   - Of the other measures, the STarT Back tool appears to be interesting, but research has not yet confirmed its utility in chiropractic practice. It could be valuable to include the STarT Back tool in PBRN studies to assess its utility in these settings as long as there was a suitable rationale for its use.

3. *Numerical Rating Scale (NRS) for pain intensity*
   - Both the NRS and VAS appear to have adequate psychometric properties and studies suggest they may be interchangeable.
   - The VAS is more commonly used in research studies, although more studies in a systematic review recommended the NRS. The NRS is used more commonly in clinical practice. The NRS is also being used by all 4 of the PBRNs that are looking at or planning on looking at pain intensity.
   - The NRS is preferable due to ease of use and familiarity for clinicians and high compliance. An 11-point NRS with suitable anchors (no pain, worst pain imaginable) is recommended.
   - An electronic survey would likely work best with the NRS as well.

4. *PROMIS-Global Health Scale (GHS) for overall health.*
   - PROMIS-GHS is a relatively new measure of overall health, but it is psychometrically sound, user-friendly, and freely available.

- The Short-Form questionnaire, particularly SF-36, is most commonly used in practice and research. The SF-36 is easily available, however it is a 36-item questionnaire and is proprietary information.
- The EQ-5D has the most evidence among shorter length QOL questionnaires and sufficient psychometric strength. The EQ-5D also requires permission for use,and may involve licensing fees.
- More of the PBRNs are planning on using some form of the Short-Form questionnaire.
- If a quality of life questionnaire is to be used, the PROMIS-GHS is recommended, as it is a good combination of a psychometrically robust and easy-to-use shorter-length instrument that is free for use. A recently announced consensus statement also recommended the PROMIS-GHS.

5. *Patient-Reported Experience Measures (PREMs)*
- Although beyond the scope of this report, PREMs including simple questions regarding patient expectations and/or satisfaction with care as well as quality of care are recommended.
- More robust or comprehensive assessments of patient care could also be undertaken using the Primary Care Assessment Survey (PCAS).

Supporting evidence from selected articles

1. Neck Pain Specific Outcome Measures

Four of the PBRNs (Passmore, French, UQTR, Quon) either plan on or are currently using the NDI.[1]

Most commonly used outcome measures in practice and research

MacDermid, *et al.* (2013)[2]
- 381 respondents, 44% were DCs, 44% Canadian.
- NDI most commonly used outcome measure for physical functioning among respondents, 49% indicated using it at least sometimes compared with the DASH (32%), Patient Specific Functional Scale (28%), SF-12/36 (9%), or EuroQol (3%)

Hinton, *et al*. (2010)[3]
- 62 Saskatchewan DCs responded.
- NDI was most commonly used neck pain related outcome measure. 31/62 respondents indicated using it at least occasionally. 8/62 used the DASH.

Khorsan, *et al*. (2008)[4]
- The NDI was the most commonly used neck related outcome measure in research studies at this point.

Systematic reviews

Nordin, *et al*. (2008)[5] (Neck Pain Task Force)
- The NDI shows moderate to good agreement with the SF-36 and is the most valid of the tools reported, it is responsive to change, it discriminates between those who improved or deteriorated, but did not detect change in score in those who remained stable. The Bournemouth Questionnaire shows high sensitivity and specificity in distinguishing neck patients who had clinically significant improvement compared with those who did not improve. The NDI has been cited in the literature as the gold standard for other questionnaires.

Schellingerhout, *et al*. (2012)[6]
- The Neck Disability Index is the most frequently evaluated questionnaire and its measurement properties seem adequate, showing positive results for internal consistency (Cronbach alpha = 0.87-0.92), content validity, structural validity, hypothesis testing, and responsiveness (AUC = 0.79), but a negative result for reliability (ICC = 0.50). The NDI has 1-factor structure although there is disagreement about that after Rasch analysis and there is concern it may not be one-dimensional for functional status. Other studies have indicated concerns with responsiveness.

667

Table 5 Quality of measurement properties per questionnaire

Questionnaire	Internal consistency	Measurement error	Reliability	Content validity	Structural validity	Hypothesis testing	Responsiveness
NDI	+++	?	−	+	++	+++	++
NPDS	?	na	?	?	+	+	+
NBQ	?	?	?	na	na	+	+
NPQ	?	?	?	?	na	+	++
WDQ	++	?	?	?	+	?	+
CNFDS	?	na	?	?	na	+	?
CNQ	na	na	+	?	na	+	na
CWOM	?	na	na	na	na	+	+

+++ or − − − strong evidence positive/negative result, ++ or − − moderate evidence positive/negative result, + or − limited evidence positive/negative result, ± conflicting evidence, ? unknown, due to poor methodological quality, *na* no information available

Pellicciari L, *et al*. (2016)[7]

- NDI time to administer is 5-10 minutes, internal consistency 0.72-0.99, reliability shows an ICC of 0.81-0.99. Again it is generally considered to be a one-factor measure of function, but some studies indicate the existence of others (pain and disability). There were again questions about responsiveness, particularly in high-functioning populations.

- Neck Bournemouth Questionnaire is 7 items, 5 minutes to complete, 2 factors. Studies tend to favor its psychometric properties. Cronbach's alpha 0.79-0.92, although there is no MDC.

2. Low Back Pain Specific Outcome Measures

Five of the PBRNs (Passmore, French, UQTR, Quon, Nova Scotia) either plan on or are currently using the Oswestry (ODI)[8], while 2 indicated the BDQ (Passmore, UQTR). None indicated using the Roland-Morris Disabilty Questionnaire (RMDQ) or STarT Back Tool (SBT) although we did not specifically ask about them.

Most commonly used outcome measures in practice and research

Hinton, *et al*. (2010)[3]
- Oswestry (34/62) was the most commonly used OM specifically for low back pain, then Roland-Morris (13/62)

Khorsan, *et al*. (2008)[4]
- The ODI was the most commonly used back pain related outcome measure in research studies at this point, followed by the RMDQ at a rate of about 3:1.

Deyo, *et al*. (2014)[9]
- Recommended using the short-form PROMIS, but indicated that researchers could also substitute the ODI or RMDQ for the PROMIS physical function items if wanting to have more information on physical function .

Scale	Purpose/content	Method of administration	Respondent burden	Administrative burden	Score interpretation	Reliability evidence	Validity evidence	Ability to detect change	Strengths	Cautions
PILE	Capacity to tolerate strenuous lifting throughout a day and to evaluate lifting capacity	Observer-led task	5–15 min, temporary increase of pain	5–15 min, increase of lifting weight, register time, HR, and quality of lifting	Weight lifted adjusted for sex/weight or number of completed lifting cycles	Good ICCs; however, LOA large (46% of baseline score)	Good and proof for construct validity	Poor to moderate	Safe, inexpensive, easy to administer psychophysical lifting end point, and unconstrained lifting reflects "real-world" lifting	N/A in patients taking HR-limiting medication. Unable to discriminate the "weak link" of the biomechanical lifting chain 10% of patients are not able to perform task
ODI	Measuring pain-related disability in people with acute, subacute, or chronic low back pain	Self-completed questionnaire by patient on paper and/or phone	<5 min	<1 min	Total score ranges from 0 (no disability) to 100 (maximum disability)	Good ICCs	Adequate content and construct validity; however, lacks generic activities such as work, leisure, recreational, or sporting activities	Cutoff point for minimum important change is 10 points or a 30% score improvement	Simple to use and score, and has minimal respondent and administrative burden	Face-to-face or computer administration would be the preferred method over telephone interview
LBPRS	Measuring 3 clinical components of low back pain (back pain [leg], disability, and physical impairment	Self-completed questionnaire by patient on paper or by interview	~15 min	~15 min	Score ranges: 0–60 for pain, 0–40 points for disability, 0–40 points for impairments. Recommended not to use the total sum score. Higher scores are indicative of more problems	High interrater reliability (97.7%)	Correlates highly with RDQ	MCID for the disability scale is 17 and for pain scale is 10 points	Simple and contains a well-balanced distribution of items across the ICF components pain, activity limitation, and physical impairment	Responsiveness is lower compared to RDQ and ODI Lacks information on MDC and SEM
RDQ	Measuring daily physical activities and functions that may be affected by low back pain	Self-completed questionnaire by patient on paper and electronic version	<5 min	<1 min	Scores range from 0 (no disability) to 24 (maximal disability)	Internal consistency and ICC are good. MDC and SEM are known, but are influenced by several factors (time intervals, methods used, etc.)	Acceptable; contains a small number of items that are not functional limitations. Correlates well with other disability measures	MCID ranges from 2–5 points. A 30% change from baseline was proposed as a clinically meaningful improvement (normally equivalent to an absolute change of 5 points)	Short, simple to complete, and readily understood by patients and clinicians. Psychometric properties are acceptable to good and the RDQ is available in many language versions. It can be used in acute, subacute, and chronic low back pain patients	Less suitable for patients with low levels of disability. Can be improved through the removal of items with poor fit statistics and the addition of items toward the extremes of the scale hierarchy
QBPDS	Measuring elementary daily activities that patients with back pain might perceive difficult to perform. Items can be classified into 6 domains of activity affected by back pain	Self-completed questionnaire by patient on paper, mail, and/or phone	<5 min	<1 min	Score ranges from 0 (no disability) to 100 (maximal disability)	Internal consistency and ICC are good	Good, contains various domains of activity that were selected by patients and health care providers; correlates well with other disability measures	MCID ranges from 8.5–32.9 mainly due to the heterogeneity of the study populations. A 30% change from baseline was proposed as a clinically meaningful improvement	Short, easy to use, and acceptable. Measures functional disability in daily life that is essential in patients with low back pain	Due to changes regarding the scale's format and the wording of some of the items, one cannot be sure that all clinimetric properties reported in studies are identical for the newly proposed version

> **Summary Table for Measures of Function in Low Back Pain/Disorders***

* PILE = Progressive Isoinertial Lifting Evaluation; HR = heart rate; ICC = intraclass correlation coefficient; LOA = limits of agreement; N/A = not applicable; ODI = Oswestry Disability Index; LBPRS = Low Back Pain Rating Scale; RDQ = Roland-Morris Disability Questionnaire; MCID = minimum clinically important difference; ICF = International Classification of Functioning, Disability and Health; MDC = minimum detectable change; QBPDS = Quebec Back Pain Disability Scale.

Systematic reviews

Smeets, *et al.* (2011)[10]
- Provides a detailed summary of the properties (including psychometric properties) of several back pain specific instruments. See table below for an overview.

Chiarotto, *et al.* (2016)[11]
- Identified 11 studies of fair or poor methodological quality, performing head-to-head comparisons of the RMDQ and ODI.
- The ODI showed better reliability and measurement error, whereas the RMDQ showed better construct validity as a measure of physical functioning.
- Concluded there are no strong reasons to prefer one instrument over the other to measure physical functioning in patients with LBP.

Table 4.
Best Evidence Synthesis of Measurement Properties of the RMDQ and ODI in Head-to-Head Comparison Studies Conducted in Patients With Nonspecific Low Back Pain[a]

Measurement Properties	RMDQ Level of Evidence (Rating)	ODI Level of Evidence (Rating)	Is One Instrument Better Than the Other?
Internal consistency	?	?	?
Reliability	Conflicting (+/–)	Moderate (+)	Yes, ODI
Measurement error	Moderate (–)	Moderate (+)	Yes, ODI
Face validity	?	?	?
Content validity	?	?	?
Structural validity	?	?	?
Construct validity	Moderate (+)	Conflicting (+/–)	Yes, RMDQ
Cross-cultural validity	?	?	?
Criterion validity	?	?	?
Responsiveness	Conflicting (+/–)	Conflicting (+/–)	No

[a] RMDQ=Roland-Morris Disability Questionnaire, ODI=Oswestry Disability Index, ?=unknown due to only studies of poor methodological quality or no studies on that measurement property, +/–=conflicting findings, +=consistent positive findings, –=consistent negative findings.

Froud, e*t al.* (2016)[12]
- RM was the most commonly used OM but only comprised 28% of trials in 2012. ODI use has steadily increased to the point of now being greater than the RM with about 39% of studies in 2012 employing it.
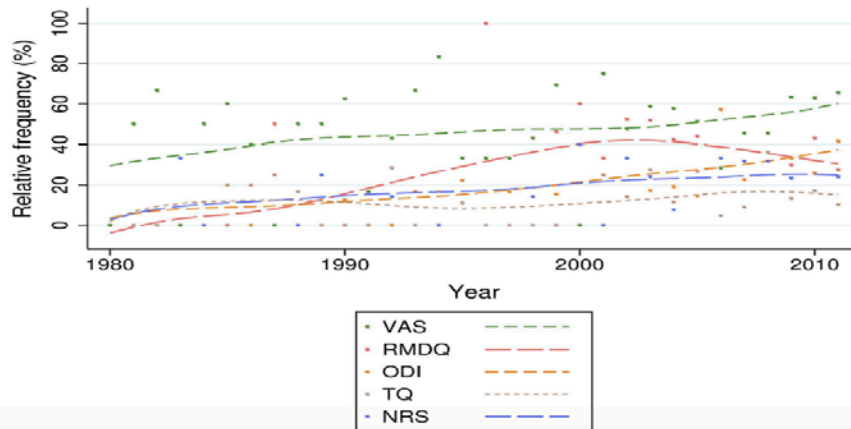
Bournemouth Disability Questionnaire

LarsenK, Leboeuf-Yde C. (2005)[13]
- Concluded that the BDQ was not useful at baseline, in monitoring, or predicting 1-year status.
- Considerable disagreement between the ODI and BDQ.

STaRT Back Screening Tool

Field J, Newell D. (2012)[14]
- Concluded that STaRT Back screening tool may not be useful in chiropractic practice as it may not have prognostic utility.
- Study in 2013 by Irgens *et al*. indicated that the STaRT Back tool and BDQ scores seem to be in agreement (correlation of 0.59).
- A review of the STaRT Back tool by Khan in 2016 indicates that the tool is potentially useful due to correlations with the BDQ, but as chiropractic patients often have shorter durations of LBP its prognostic utility may be limited or at least may be more appropriate in chronic LBP populations.
- All studies conducted thus far using the STaRT Back tool have been in Europe.

3. Pain Intensity Outcome Measures

4 of the PBRNs (Passmore, UQTR, French, Quon) either plan on or are currently using the Numerical Pain Rating Scale (NPRS/NRS), 3 for the VAS (Quon, UQTR, French), 2 for the Verbal Rating Scale (VRS) (Passmore, Quon), and 1 the McGill Pain Questionnaire (UQTR).

Most commonly used in practice and research

MacDermid, *et al*. (2013)[2]
- NRS was most commonly used pain scale (75% at least sometimes), followed by VAS (49% at least sometimes)

Hinton, *et al*. (2010)[3]
- For pain most frequently used were pain diagrams (50/62 at least occasionally), then NRS (50/62), then VAS (40/62)

Khorsan, *et al*. (2008)[4]
- VAS was most commonly used in chiropractic/SMT studies, followed by NRS (at a ration of about 2.5:1), then McGill Pain Questionnaire

Systematic Reviews

Hjermsted, *et al*. (2011)[15]

- Pain intensity should be assessed by uni-dimensional scales based on self-report.
- Well-validated instruments, such as the Brief Pain Inventory or the short-form McGill Pain Questionnaire are recommended for more comprehensive pain assessment.
- VAS is more frequently used in research studies. The 11-point NRS is the most frequently used version.
- Less educated and elderly people tend to prefer the VRS, although the NRS is generally preferred in mixed-age and chronic pain populations.
- There tends to be better compliance with the NRS and VRS
- The most common anchors were 'no pain/worst pain imaginable'
- 11 papers recommended NRS due to its ease of use, responsiveness, and high compliance, 7 recommended the VRS, 4 recommended the VAS, and 29 had no recommendation.
- The majority of the reviewed papers showed relatively consistent findings with respect to the correlation between scales, and when assessed, most coefficients between changes in scores over time were high, indicating that the scales tended to measure variations in the same direction.
- The NRS, VAS, and VAS all work quite well. Most important are the conditions related to its use, which include: a standardized choice of anchor descriptors,

methods of administration, time frames, information related to the use of scales, interpretation of cut-offs and clinical significance, and the use of appropriate outcome measures and statistics in clinical trials.

Nordin, *et al*. (2008)[5] (Neck Pain Task Force)

- The VAS is best at detecting change in patients who improve. The VAS has been used to show a weak association between pain and disability and a negative correlation between neck strength output and pain. Responsiveness to change was high for the VAS (in patients who improve).
- The VAS has been cited in the literature as the gold standard for other questionnaires.

Froud, e*t al*. (2016)[12]
- Most commonly used is VAS – 60% of LBP trials in 2012, the NRS was used in just over 20% of LBP trials in 2012.

Other studies
- The VAS and VRS should not be used interchangeably in chronic pain, as there is systematic disagreement and a low probability of agreement (Lund 2005, Kliger 2014).
- The VAS and NRS could be used interchangeably for acute pain as strong correlations have been noted (=0.94) (Bahreini 2015)
- The VRS and NRS measures of current pain exhibited at least small responsiveness in chronic pain patients. Among patients with improved pain, however, the current pain NRS demonstrated superior responsiveness to the VRS (Chien 2013).

4. Health Status Measures

Three of the PBRNs (Passmore, UQTR, French) either plan on or are currently using the SF-36 or some version of the Short-Form, compared with 2 planning or using the EQ-5D (Passmore, UQTR).

Most commonly used outcome measures in practice and research

MacDermid, *et al*. (2013)[2]
- SF-12/36 was used by 9% at least sometimes, compared with 3% at least sometimes using the EuroQol.

Hinton, *et al*. (2010)[3]
- 5/62 respondents indicated using the SF-36 at least occasionally.

Khorsan, *et al*. (2008)[4]
- SF-36 was the only QOL-related outcome measure mentioned in chiropractic/SMT studies, although not often.

Systematic Reviews

Finch, Dritsaki, Jommi. (2016)[16]
- 37 papers on 35 studies
- The EQ5D correlates fairly well with ODI (0.21-0.74) and RM (0.42-0.82) and there is strong correlation between EQ5D and VAS (0.67). There is moderate correlation between the EQ5D and SF6D (0.55) as well as the EQ5D and SF36 (0.49).
- The EQ-5D is generally able to detect improvements and deteriorations in health states because of health interventions or disease progression. It is valid and responsive for LBP but not as responsive as disease specific OMs.
- The EQ-5D performs well in the LBP population and its scores are suitable for economic evaluation of LBP interventions, the use of EQ-5D in combination with disease-specific instruments is recommended for clinical evaluation, given its lack of sensitivity to change in health state compared with them.
- Results for SF-6D and HUI III are too scarce to draw any conclusion

Bryan, *et al.* (2014)[17]
- Rapid review, conducted at the University of British Columbia
- Strengths and weaknesses are noted for each of the instruments reviewed (see Tables and Figures below).
- SF-36, EQ-5D, and PROMIS-GHS each merit consideration based on their psychometric properties.

- PROMIS-GHS is available free of use, whereas there are licensing or cost considerations with both the SF-36 and EQ-5D. However, PROMIS-GHS has the smallest supporting evidence base.
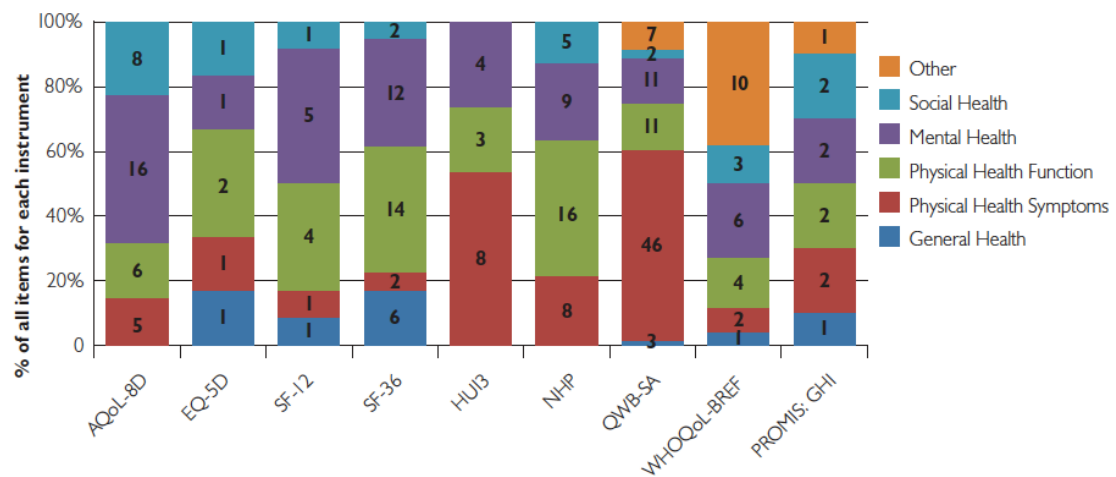
FIGURE 1. Overview of results from psychometric review

	AQoL	EQ-5D	SF-36	HUI	NHP	QWB	WHOQoL	PROMIS
Internal consistency	positive	n/a	positive	?	conflicting	n/a	conflicting	positive
Reliability	?	conflicting		conflicting	conflicting	?	?	?
Content validity	negative	negative	positive	conflicting	conflicting		?	conflicting
Construct validity	?	?	positive	?	negative	?	positive	positive
Cross-cultural validity	?	?	conflicting	?	negative	?	positive	?
Criterion validity	?	conflicting	?	?	?	?	?	?
Responsiveness	positive	conflicting	positive	conflicting	negative	conflicting	?	positive

■ = negative evidence   ■ = positive evidence   ┊ = conflicting evidence   ? = unknown/not reported

Note: The width of the bars indicates the volume of available evidence demonstrating the observations (i.e., narrow bars are indicative of fewer studies and wide bars are indicative of a larger number of studies)

FIGURE 2. Domain coverage for selected PROMs

Legend: Other; Social Health; Mental Health; Physical Health Function; Physical Health Symptoms; General Health

Y-axis: % of all items for each instrument (0 to 100%)

X-axis instruments: AQoL-8D, EQ-5D, SF-12, SF-36, HUI3, NHP, QWB-SA, WHOQoL-BREF, PROMIS: GHI

Item counts within bars:
- AQoL-8D: 8, 16, 6, 5
- EQ-5D: 1, 1, 2, 1, 1
- SF-12: 1, 5, 4, 1, 1
- SF-36: 2, 12, 14, 2, 6
- HUI3: 4, 3, 8
- NHP: 5, 9, 16, 8
- QWB-SA: 7, 2, 11, 11, 46, 3
- WHOQoL-BREF: 10, 3, 6, 4, 2, 1
- PROMIS: GHI: 1, 2, 2, 2, 1

Note: The number within the bars represents the number of items for each instrument.

TABLE 2. Summary of strengths and weaknesses of selected PROMs

Instrument	Strengths	Weaknesses
AQoL	Discriminates between groups with clinical variations in health.	Smaller evidence base.
EQ-5D	Discriminates between groups with clinical variations in health.	Not as comprehensive. Not sensitive to small changes, limited responsiveness in healthy populations.
SF-36	Top instrument in most psychometric categories. Widely used, multiple cultural contexts and many versions available.	
HUI	Can distinguish between groups with clinical variations in health, and widespread use in a variety of cultural contexts.	Lacking in mental health. Less reliability. Less responsive in populations of fairly good health.
NHP	More responsive than SF-36 in populations with poor health. Widespread use in a variety of cultures.	Not ideal for use in general population, or outside of populations with major health issues.
QWB	Good for capturing change in primarily healthy populations.	Lacking on mental health, may overweight minor conditions.
WHOQoL	Very strong cross-cultural validity. Correlated with groups with clinical variations in health.	Smaller evidence base.
PROMIS GHS	Good internal consistency, responsiveness and correlation with other instruments.	Smaller evidence base.

TABLE 3. Respondent burden and readability

Instrument	Number of items	Word count	Time for completion (min)	Flesch–kincaide grade level
AQoL-8D	35	1,188	5	5.3
EQ-5D	6	239	"few minutes"	10.6
SF-36®	36	692	10	5.9
HUI3®	15	1,173	8–10	7.4
NHP	38	353	5–15	2
QWB-SA	80	1,934	15	5.6
WHOQoL-BREF	26	607	5	6.7
PROMIS/GHS	10	217	2	7.6

References

1. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther. 1991; 14(7): 409-415.
2. MacDermid JC, Walton DM, Cote P, *et al*. Use of outcome measures in managing neck pain: an international multidisciplinary survey. Open Orthop J. 2013; 7(Suppl 4): 506-520.
3. Hinton PM, *et al*. Outcome measures and their everyday use in chiropractic practice. JCCA. 2010; 54(2): 118-131.
4. Khorsan R, Coulter ID, Hawk C, Goertz C. Measures in chiropractic research: choosing patient-based outcome assessments. JMPT. 2008; 31: 355-375.
5. Nordin M, *et al*. Neck Pain Task Force. Assessment of neck pain and its associated disorders. Spine. 2008. 33; 4S: S101-S122.
6. Schellingerhout JM, *et al*. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res. 2012; 21: 659-570.
7. Pellicciari L, *et al*. Patient-reported outcome measures for non-specific neck pain validated in the Italian-language: a systematic review. Archives Physiother. 2016; 6:9.
8. Fairbank JC, Couper J, Davies JB. The Oswestry Low Back Pain Questionnaire. Physiother. 1980; 66: 271-273
9. Deyo RA, *et al*. Report of the NIH Task Force on research standards for chronic low back pain. J Pain. 2014; 15(6): 569-585.
10. Smeets R, *et al*. Measures of function in low back pain/disorders. Arthr Care Res. 2011; 63 (S11): S158-S173.
11. Chiarotto A, *et al*. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. Phys Ther. 2016; 96(10): 1620-1637.
12. Froud R, e*t al*. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: respice, adspice, et prospice. PLOS One. 2016. 11(10): e0164573.
13. LarsenK, Leboeuf-Yde C. The Bournemouth Questionnaire: Can it be used to monitor and predict treatment outcome in chiropractic patients with persistent low back pain? JMPT. 2005; 28: 219-22.
14. Field J, Newell D. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. Chiropr Man Ther. 2012; 20:17.
15. Hjermsted MJ, *et al*. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Symptom Pain Manage. 2011; 41(6): 1073-1093.
16. Finch APP, Dritsaki M, Jommi C. Generic preference-based measures for low back pain. Which of them should be used? Spine. 2016; 41(6): 364-374.

17. Bryan S, *et al*. Choosing your partner for the PROM: a review of evidence and patient-reported outcome measures for use in primary and community care. Healthcare Policy. 2014; 10(2): 38-51.