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.

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 Disability 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.

### Summary Table for Measures of Function in Low Back Pain/Disc Disease

<table>
<thead>
<tr>
<th>Scale</th>
<th>Purpose/Domain</th>
<th>Method of Administration</th>
<th>Response Format</th>
<th>Administration</th>
<th>Scoring Interpretation</th>
<th>Reliability Evidence</th>
<th>Validity Evidence</th>
<th>Ability to Measure Change</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLE</td>
<td>Quantify time lost associated with specific activity and to assess lifting capacity</td>
<td>Uncompleted task</td>
<td>Shortest time to complete</td>
<td>Number of correct responses</td>
<td>Percentage of correct responses</td>
<td>Good ICCs</td>
<td>Good and fair to excellent for short-term reliability</td>
<td>Moderate to high</td>
<td>Good, information about patient's daily activities, can be used as a baseline, can be used to monitor improvement</td>
</tr>
<tr>
<td>ODI</td>
<td>Measuring pain-related disability in people with acute or chronic low back pain</td>
<td>Self-assessed, no examiner present</td>
<td>Numeric scale (0-50)</td>
<td>Total score range</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Good ICCs</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Low to moderate</td>
<td>Good, information about patient's daily activities, can be used as a baseline, can be used to monitor improvement</td>
</tr>
<tr>
<td>RMDQ</td>
<td>Measuring self-assessed function and disability related to low back pain</td>
<td>Self-assessed, no examiner present</td>
<td>Numeric scale (0-100)</td>
<td>Total score range</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Good ICCs</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Low to moderate</td>
<td>Good, information about patient's daily activities, can be used as a baseline, can be used to monitor improvement</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Measuring health status in people with low back pain</td>
<td>Self-assessed, no examiner present</td>
<td>Numeric scale (0-100)</td>
<td>Total score range</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Good ICCs</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Low to moderate</td>
<td>Good, information about patient's daily activities, can be used as a baseline, can be used to monitor improvement</td>
</tr>
<tr>
<td>EQ-VAS</td>
<td>Measuring health status in people with low back pain</td>
<td>Self-assessed, no examiner present</td>
<td>Numeric scale (0-100)</td>
<td>Total score range</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Good ICCs</td>
<td>Adequate content validity and internal consistency, high test–retest reliability</td>
<td>Low to moderate</td>
<td>Good, information about patient's daily activities, can be used as a baseline, can be used to monitor improvement</td>
</tr>
</tbody>
</table>

*P.L.E = Progression Level Evaluation; HR = Health-related; ICC = Intraclass correlation coefficient; LUM = limits of agreement; N/A = not applicable; ODI = Oswestry Disability Index; DRS = Low Back Pain Rating Scale; EQ-5D = EuroQol-5D; MDC = minimal detectable change; EQ-VAS = EuroQol Visual Analog Scale; MDC = minimal detectable change*
Systematic reviews

Smeets, et al. (2011)\textsuperscript{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)\textsuperscript{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>
<thead>
<tr>
<th>Measurement Properties</th>
<th>RMDQ Level of Evidence (Rating)</th>
<th>ODI Level of Evidence (Rating)</th>
<th>Is One Instrument Better Than the Other?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Reliability</td>
<td>Conflicting (+/-)</td>
<td>Moderate (+)</td>
<td>Yes, ODI</td>
</tr>
<tr>
<td>Measurement error</td>
<td>Moderate (-)</td>
<td>Moderate (+)</td>
<td>Yes, ODI</td>
</tr>
<tr>
<td>Face validity</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Content validity</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Structural validity</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Construct validity</td>
<td>Moderate (-)</td>
<td>Conflicting (+/-)</td>
<td>Yes, RMDQ</td>
</tr>
<tr>
<td>Cross-cultural validity</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Conflicting (+/-)</td>
<td>Conflicting (+/-)</td>
<td>No</td>
</tr>
</tbody>
</table>

\textsuperscript{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, et al. (2016)\textsuperscript{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)\textsuperscript{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)\textsuperscript{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)\textsuperscript{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, et al. (2016)\textsuperscript{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.
### TABLE 2. Summary of strengths and weaknesses of selected PROMs

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

### TABLE 3. Respondent burden and readability

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Number of items</th>
<th>Word count</th>
<th>Time for completion (min)</th>
<th>Flesch–Kincaide grade level</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQoL-8D</td>
<td>35</td>
<td>1,188</td>
<td>5</td>
<td>5.3</td>
</tr>
<tr>
<td>EQ-SD</td>
<td>6</td>
<td>239</td>
<td>“few minutes”</td>
<td>10.6</td>
</tr>
<tr>
<td>SF-36</td>
<td>36</td>
<td>692</td>
<td>10</td>
<td>5.9</td>
</tr>
<tr>
<td>HUI</td>
<td>15</td>
<td>1,173</td>
<td>8–10</td>
<td>7.4</td>
</tr>
<tr>
<td>NHP</td>
<td>38</td>
<td>353</td>
<td>5–15</td>
<td>2</td>
</tr>
<tr>
<td>QWB-SF</td>
<td>80</td>
<td>1,934</td>
<td>15</td>
<td>5.6</td>
</tr>
<tr>
<td>WHOQoL-BREF</td>
<td>26</td>
<td>607</td>
<td>5</td>
<td>6.7</td>
</tr>
<tr>
<td>PROMIS GHS</td>
<td>10</td>
<td>217</td>
<td>2</td>
<td>7.6</td>
</tr>
</tbody>
</table>