

Choosing Your Partner for the PROM: A Review of Evidence on Patient-Reported Outcome Measures for Use in Primary and Community Care

Choisir son partenaire pour le PROM : examen des données sur l'évaluation de l'état de santé du point de vue du patient dans les établissements de soins primaires et communautaires



STIRLING BRYAN, PHD

*Director, Centre for Clinical Epidemiology & Evaluation, Vancouver Coastal Health Research Institute
Professor, School of Population & Public Health, The University of British Columbia, Vancouver, BC*

JENNIFER DAVIS, PHD

*Post-Doctoral Fellow, Centre for Clinical Epidemiology & Evaluation
Vancouver Coastal Health Research Institute, Vancouver, BC*

JAMES BROESCH, PHD

*Post-Doctoral Fellow, Centre for Clinical Epidemiology & Evaluation
Vancouver Coastal Health Research Institute, Vancouver, BC*

MARY M. DOYLE-WATERS, MA, MLIS

*Librarian, Centre for Clinical Epidemiology & Evaluation
Vancouver Coastal Health Research Institute, Vancouver, BC*

STEVEN LEWIS, PHD

*President, Access Consulting Ltd
Adjunct Professor, Faculty of Health Sciences, Simon Fraser University, Saskatoon, SK*

KIM MCGRAIL, PHD

*Assistant Professor, School of Population & Public Health, The University of British Columbia
Associate Director, Centre for Health Service & Policy Research, The University of British Columbia, Vancouver, BC*

On behalf of the authors (see Acknowledgements)

Abstract

Patient-reported outcome measures (PROMs) are assessments of health status from the patient's perspective. The systematic and routine collection and use of PROMs in healthcare settings adds value in several ways, including quality improvement and service evaluation. We address the issue of instrument selection for use in primary and/or community settings. Specifically, from the large number of available PROMs, which instrument delivers the highest level of performance and validity? For selected generic PROMs, we reviewed literature on psychometric properties and other instrument features (e.g., health domains captured). Briefly we summarize key strengths of the three PROMs that received the most favourable psychometric and overall evaluation. The Short-Form 36 has a number of strengths, chiefly, its strong psychometric properties such as responsiveness. The PROMIS/Global Health Scale scored highly on most criteria and warrants serious consideration, especially as it is free to use. The EQ-5D scored satisfactorily on many criteria and, beneficially, it has a low response burden.

Résumé

Les PROM (*patient-reported outcome measures*) sont des méthodes d'évaluation de l'état de santé du point de vue du patient. Le recours systématique et routinier aux PROM dans les établissements de santé apporte une valeur ajoutée de diverses façons, notamment en ce qui a trait à l'amélioration de la qualité et à l'évaluation des services. Nous nous penchons sur la question du choix d'instrument à utiliser dans les établissements de soins primaires et/ou de soins communautaires. Plus précisément, quel instrument parmi le vaste nombre de PROM disponibles obtient le meilleur rendement et offre la plus grande validité? Nous avons évalué la littérature sur les propriétés psychométriques et autres caractéristiques (par exemple, les domaines de santé saisis) de certains PROM génériques. Nous avons fait un bref sommaire des forces clés des trois PROM qui ont obtenu la meilleure évaluation psychométrique et générale. Le formulaire *Short-Form 36* présente plusieurs points forts, notamment ses solides propriétés psychométriques telles que la réactivité. Le *PROMIS/Global Health Scale* a obtenu un score élevé pour la plupart des critères et mérite une attention particulière, particulièrement en raison de sa gratuité. *LEQ-5D* a obtenu un score satisfaisant pour plusieurs critères et présente un faible fardeau de réponse, ce qui est un avantage.

Introduction

Patient-reported outcome measures (PROMs) are assessments of health status or health-related quality of life from the patient's perspective, a viewpoint not captured by clinical outcomes (Dawson et al. 2010). The process typically used to collect such data is for the patient-reported outcome (PROM) questionnaire to be completed by the patient at specific times in his/her clinical trajectory (e.g., at regular intervals following surgery or at defined follow-up points after diagnosis of a chronic condition).

The systematic and routine collection and use of PROMs in healthcare settings can add value in several ways, including direct patient management, quality improvement and service evaluation, technology assessment and research and the assessment of practitioner performance. In a previous paper, “Let’s All Go to the PROM” (McGrail et al. 2011), we set out the arguments for routine collection of PROMs data. In this paper we address the issue of instrument selection – there are many PROMs to choose from but which ones deliver the highest level of performance and validity?

The scope of the project reported here was to consider PROMs for use in primary and/or community care settings, as opposed to an elective surgery context which has received more attention in the research literature. We worked closely with stakeholders at a provincial Ministry of Health whose primary interest was in the use of PROMs to support evaluation of integrated primary and community care initiatives. This work thus provides information and a summary of evidence to inform PROM selection in a primary/community care context.

For a number of widely used PROM instruments, this paper reports: (1) the psychometric properties of the instruments; (2) health domains captured; (3) implementation considerations (respondent burden, administration methods, translations available, readability and cost); and (4) availability of population norms and utility scoring algorithms. These areas were identified as important to instrument selection through consultation with a range of knowledge users (health researchers, practitioners, administrators and government).

Selection of PROMs as a Focus for This Study

This rapid review was undertaken for the BC Ministry of Health. As such, we were given six months for completion of the review. The initial weeks of the project were spent clarifying with the Ministry their specific goals and questions for the project so that our research team could provide the Ministry with an appropriate selection of instruments. Due to time constraints, a systematic review was not feasible; therefore, a rapid review was undertaken. The methods used made every effort to ensure reproducibility and rigour, and to avoid bias. Below we describe our approach to the selection of the shortlisted instruments.

According to the Mapi Research Trust’s Patient-Reported Outcome and Quality of Life Instruments Database (“PROQOLID. Patient-Reported Outcome and Quality of Life Instruments Database. [internet] 2013 [cited 2012 December]; available from: <http://www.proqolid.org/>”), there are now close to 800 PROM instruments, with more being developed all the time. Therefore, the first stage of this research was to select a shortlist of PROM instruments, and hence define a manageable task.

Given that our study focused on applications in primary and community care, the most relevant PROMs are comprised of generic measures that provide the opportunity to compare outcomes across many clinical conditions. Specifically, a measure applicable to a primary and community care setting was defined as one that is broad enough in content to capture important differences in health status resulting from a wide array of health conditions typically seen in primary/community care. As such, we focused on generic measures of health status that

cover domains used to assess overall health status. The benefit of a generic measure, with its ability to make comparisons across diverse clinical conditions, involves a tradeoff: by excluding condition-specific instruments, some sensitivity and responsiveness is almost inevitably sacrificed in subsets of the clinical population. On the other hand, one of the advantages of a generic PROM is that it provides information about a respondent's overall health status or health-related quality of life, which often includes multiple aspects or domains (i.e. physical, emotional, mental, social and general health).

The process for uncovering PROM measures involved re-examining existing reviews and performing structured searches of instrument databases and consulting measurement experts. We acquired a PROQOLID membership for the project. However, we did not rely exclusively on PROQOLID to identify relevant instruments. We used PROQOLID as a tool to screen all indexed generic instruments. Of note, several instruments that were not indexed in PROQOLID but were identified through expert and stakeholder consultations were also included. Internet and database searches were conducted to verify the completeness of our long- and short-lists of instruments. PROQOLID provided a comprehensive and up-to-date listing of 115 generic PROMs. From this, we developed a short-list of 25 potential generic PROMs applicable for use in a primary care setting. To be considered in this rapid review, the generic PROMs needed to include the following characteristics:

- + designed for use among adults;
- + able to be self-administered;
- + capable of generating a summary score to assess overall health status;
- + applicability to a primary care and community care setting; and
- + widespread in their current use (evaluated via a formal citation search).

Project experts also suggested including the measures How's Your Health, PROMIS and RAND-36, none of which was included in PROQOLID. The total number of eligible instruments was, therefore, 28. The "impact" of each instrument was considered by reviewing the number of cited references for the original papers (over the past six years) in the Web of Science, to ensure that the selection of instruments was targeted to those in widespread and current use (see Appendix 1 for the citation search details). Our final short-list of candidates included 8 PROM instruments:

- + Assessment of Quality of Life (AQoL-8D)
- + EuroQol EQ-5D-3L
- + Health Utilities Index (HUI3)
- + Nottingham Health Profile (NHP)
- + PROMIS-Global Health Scale (GHS)
- + Quality of Well-Being Scale (QWB)
- + Short-Form 36 (SF-36)
- + World Health Organization Quality of Life Instrument (WHOQoL-BREF)

For many of these instruments, there exist numerous versions. For example, there are three versions of the Health Utilities Index (HUI1, HUI2 and HUI3), and the SF “family” of instruments includes the SF-36 (versions 1 and 2), the SF-12 (versions 1 and 2) and the SF-8. For each instrument we made a pragmatic decision to focus primarily on the latest, most commonly used version that was most applicable to primary/community care settings. However, some of the review work, especially where studies researched multiple instruments, reports evidence from multiple-instrument versions.

Our review and evaluation efforts had a two-pronged focus: (1) ascertaining the psychometric properties of the instruments; and (2) identifying data and information regarding important instrument properties, such as domain coverage, implementation considerations and the availability of utility scores and population norms.

Methods

Psychometric review

We synthesized evidence from existing reviews of the psychometric properties of our eight candidate instruments (Khangura et al. 2012).

Our searches were developed in MEDLINE (OvidSP) and Embase (OvidSP) using keywords because most instruments were not indexed as subject terms in the MeSH or Emtree thesauri. We used two search filters. The first, a filter specifically for measurement properties of health instruments reported by Terwee and colleagues (2009) and recommended by the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) group (Mokkink et al. 2010), was used in PubMed and then adapted for MEDLINE and Embase (OvidSP). The second filter, developed by the Scottish Intercollegiate Guidelines Network (Scottish Intercollegiate Guidelines Network; available from: <http://www.sign.ac.uk/methodology/filters.html>), is also used for systematic reviews and the publication-type “reviews.” Our search comprised three components: terms for the eight PROMs that were then combined with the measurement filter and then the review filter.

We also performed grey literature searches. The majority of the instruments have websites that were also examined for evidence relating to psychometric properties. Further details on our search strategy are given in Appendices 2 and 3, and are available in the main project report (Bryan et al. 2013).

Two researchers (J.B. and J.C.D.) independently evaluated the titles and abstracts. Articles were selected for further review that assessed an instrument’s psychometric performance in a general population applicable to primary and community care (i.e., not specific clinical populations) and articles that focused on at least one of the psychometric properties of the instrument as suggested by the COSMIN guidelines. Specifically, the extracted information included the following: reliability (internal consistency and test–retest reliability), validity (content, construct, cross-cultural and criterion validity), responsiveness, generalizability

and comparability with other candidate PROM instruments. Further, only manuscripts that focused on at least one of the eight selected candidate PROMs were included. Discrepancies between the two reviewers were discussed and resolved by a third reviewer (S.B.), yielding 21 articles selected for full-text review. An additional article on the PROMIS Global Health Scale was added to provide coverage for all candidate instruments, bringing the total number of articles to 22 (Anderson et al. 1993, 1996; Bouchet et al. 2000; Brazier et al. 1999; Butterworth and Crosier 2004; Coons et al. 2000; Doward et al. 2004; Ford et al. 2000; Furlong et al. 2001; Gandek et al. 2004; Hawthorne and Richardson 2001; Hays et al. 2009; Haywood et al. 2005; Horsman et al. 2003; Kopec and Willison 2003; McHorney and Tarlov 1995a; Revicki and Kaplan 1993; Sintonen 2001; Skevington et al. 2004; Ware 2000; Wiklund 1990).

The data extraction process was guided by the COSMIN criteria and a corresponding data extraction template was created covering reliability (internal consistency and test–retest reliability), validity (content, construct, cross-cultural and criterion) and responsiveness. Reviewers (J.B. and J.C.D.) provided a score by strictly adhering to guidelines established by COSMIN for each of the data extraction items listed above (Mokkink et al. 2010). Scores ranged from very strong positive evidence (+++) to very strong negative evidence (---). Conflicting evidence was also noted (+/-), as was an absence of evidence in that particular category (?). Further details on the scoring approach are given in Appendix 4. We also evaluated, but did not score, additional information not included in the COSMIN guidelines, including generalizability and comparability with other candidate PROM instruments, which we considered important for decision-making.

Two reviewers (J.B. and J.C.D.) also extracted all text from these articles pertaining to each instrument's psychometric properties. This information was entered into a central database and subsequently scored independently by the two reviewers (J.B. and J.C.D.), in line with the COSMIN guidelines (Consensus-Based Standards for the Selection of Health Measurement Instruments [COSMIN] 2013; Mokkink et al. 2010).

Review of other performance attributes

The second focus of this rapid review was on other performance attributes of the selected candidate instruments. The instruments were reviewed in terms of their domain coverage through detailed inspection of the questions asked. The domain framework of the PROMIS instrument, given its comprehensive and inclusive nature, was used as a reference point (www.nihpromis.org/Documents/PROMIS_Full_Framework.pdf).

In terms of implementation considerations for policy makers and patients, we also sought information on respondent burden, readability, cost of using the instrument and the availability of official translations. Data sources for this information were primarily materials published by the instrument's developers and third-party websites focused on evaluating PROM tools. We independently assessed readability using the Flesch–Kincaid grade-level

test, which assesses an individual’s education grade level equivalent (Microsoft 2013).

Finally, information relating to selected decision-making criteria was also obtained. These criteria included: norm reference sets (to allow comparison of sample data to the general population), and utility/preference scoring algorithms (to allow the calculation of quality-adjusted life years and so facilitate cost-utility analyses). Information on these criteria was obtained for each instrument through examination of the appropriate website or scoring guides.

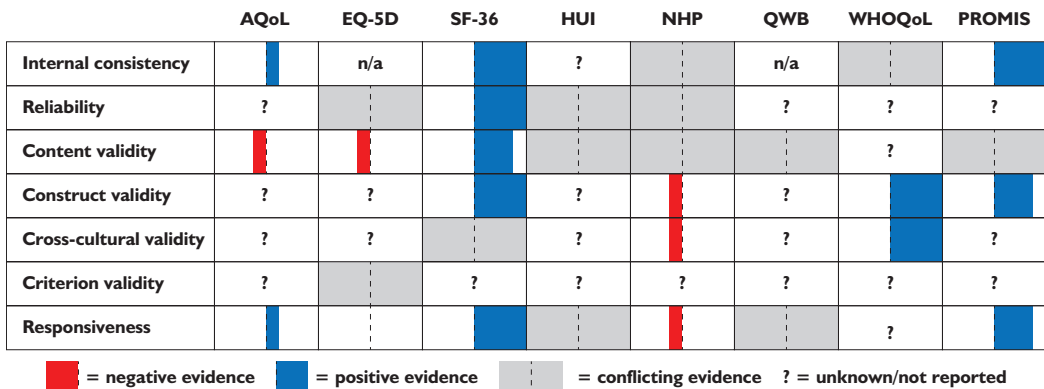
Results

Psychometric review

The details of the 22 articles providing information about the psychometric properties of the candidate PROM instruments (Anderson et al. 1993, 1996; Bouchet et al. 2000; Brazier et al. 1999; Butterworth and Crosier 2004; Coons et al. 2000; Doward et al. 2004; Ford et al. 2000; Furlong et al. 2001; Gandek et al. 2004; Hawthorne and Richardson 2001; Hays et al. 2009; Haywood et al. 2005; Horsman et al. 2003; Kopec and Willison 2003; McHorney and Tarlov 1995a; Revicki and Kaplan 1993; Sintonen 2001; Skevington et al. 2004; Ware 2000; Wiklund 1990) are reported in Table 1 (see Table 1. at www.longwoods.com/content/XXXXXX). Ten of the articles also contained direct comparisons between two or more candidate instruments. All of the articles provided a good rationale and key objectives for the review, and most also provided a summary of their results in relation to the key objectives. Each of the 10 psychometric categories we considered was evaluated for most of our candidate instruments. However, none of the reviews explicitly discussed measurement error, an important psychometric category related to reliability.

A brief summary of results from our rapid review of the candidate instruments’ psychometric properties is presented in Figure 1 and Table 2, and an overview of the main sources for evidence on each criterion is given by Table 1.

FIGURE 1. Overview of results from psychometric review



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)

Choosing Your Partner for the PROM: A Review of Evidence on Patient-Reported Outcome Measures for Use in Primary and Community Care

The SF-36 had a very large evidence base and performed as well as, or better than, the other instruments in most of the psychometric domains we considered. It was noted to be reliable (Gandek et al. 2004; Haywood et al. 2005), to be comprehensive in its coverage of various aspects of health (Haywood et al. 2005; Ware 2000) and for having strong content validity, which indicates that it taps into the domains it proposes to examine (Butterworth and Crosier 2004; Gandek et al. 2004; Hays et al. 2009; Skevington et al. 2004; Ware 2000). Its strong score in the responsiveness category, indicating the ability to detect change when it was known to have occurred or differences between groups known to vary in health status, suggested that it would be better suited for evaluation studies than some of the other instruments (Anderson et al. 1996; Bouchet et al. 2000; Hays et al. 2009; Kopec and Willison 2003; McHorney and Tarlov 1995; Ware 2000). PROMIS-GHS also performed well in most categories, demonstrating good internal consistency, construct validity and responsiveness. However, the evidence base is smaller, in part due to its more recent development (Hays et al. 2009). While the WHOQoL-BREF did not perform as well as the SF-36 in some domains, and had a smaller evidence base, it did have stronger evidence for cross-cultural validity than any of the other instruments (Skevington et al. 2004).

Several limitations were noted for the remaining instruments. Both the HUI3 and QWB-SA were reported to be lacking in their coverage of mental health (Brazier et al. 1999; Coons et al. 2000). While the NHP does perform well in populations with major burdens of disease, and may be more responsive than the SF-36 in those groups, it may be less relevant for general population samples with lower burdens of disease (Haywood et al. 2005; Wiklund 1990). The response format of the QWB-SA is problematic for detecting changes in severity of illness, and may overweight minor health conditions such as wearing eyeglasses (Brazier et al. 1999). Finally, we found the evidence base for the AQoL-8D to be smaller than some of the other instruments we considered, despite it being in use for over 10 years, which has also been noted by other authors (Hawthorne and Richardson 2001). The limited psychometric information uncovered relating to AQoL-8D may have been due in part to the rapid review methodology.

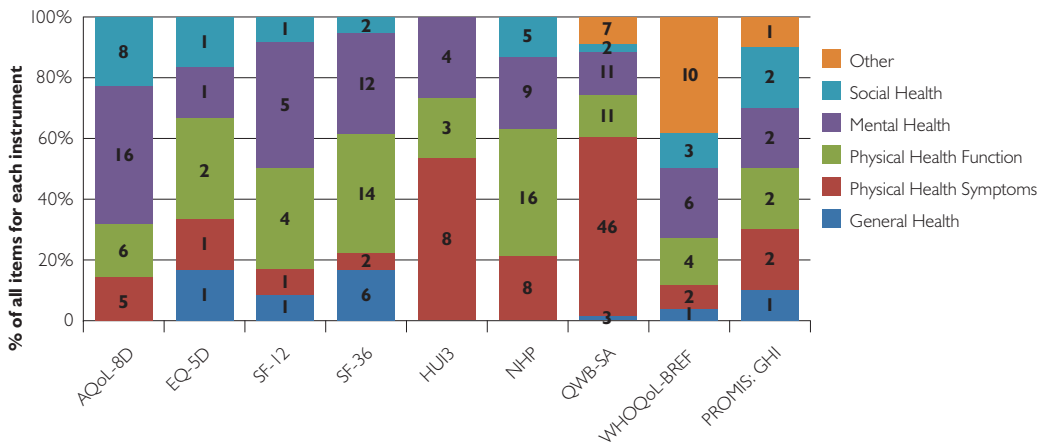
In several cases, the evidence reported in the papers reviewed may have provided useful information that was not part of the COSMIN scoring framework. An example of this is cross-cultural validity. The WHOQoL-BREF, which prioritized cross-cultural validity during its initial development, performed best in this category. This instrument has been systematically evaluated in many different contexts, and only a few meaningful differences have been noted (Skevington et al. 2004). Both the EQ-5D and HUI are used in different cultural and linguistic contexts. The SF-36 is also widely used in many contexts, and some evidence of validation across different groups (using differential item functioning analysis) has been reported for both the SF-36 and SF-12 (Ford et al. 2001; Gandek et al. 2004). The NHP had positive evidence in terms of how it was developed and its widespread use (Anderson et al. 1996;

Coons et al. 2000), but one review noted potential differences between how the instrument functions in French versus English (Anderson et al. 1996). There was no information on the cross-cultural validity of the QWB-SA, PROMIS and the AQoL.

Review of other performance attributes

In terms of domain coverage, Figure 2 provides a summary of our data. While the number of questions in each domain varied across instruments, all of our candidate instruments included questions that assessed physical and mental health. Questions regarding social health were absent from the HUI3, and questions on general health were absent from the HUI3, AQoL-8D and NHP.

FIGURE 2. Domain coverage for selected PROMs



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

All of our candidate instruments are available in the two official languages of Canada (English and French). The EQ-5D and SF-36 are available in more languages than the other instruments, including Mandarin Chinese, Punjabi, Korean and Tagalog, which are frequently spoken at home in British Columbia.

The mode of PROM administration may also affect the decision-making process. Our initial inclusion criteria required that all of our candidate PROMs could be self-administered, and all are able to be administered on paper. Alternative administration modes, such as by telephone and online, have been validated for some of the instruments, particularly for SF-36, EQ-5D and PROMIS/GHS.

Other key considerations are an instrument’s readability and the amount of time required to complete it, presented in Table 3. Readability scores reflect the estimated grade level required to understand the instrument. The EQ-5D required the highest reading level, grade 11, due to its long questions and use of multi-syllabic words. All other instruments are within the recommended ranges for working with adult populations. All of our candidate instruments are relatively short, taking between 2 and 15 minutes to complete.

Choosing Your Partner for the PROM: A Review of Evidence on Patient-Reported Outcome Measures for Use in Primary and Community Care

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

A final practical consideration is the financial costs associated with using the instrument. While many of the candidate instruments are free to use (AQoL-8D, NHP, WHOQoL-BREF and the PROMIS/GHS), others require payment to the developers when used outside of a research context (EQ-5D, SF-36, HUI3 and QWB-SA). The SF-36 requires licensing fees.¹

Depending on the context under which the PROM will be used, the availability of population norms for comparison to a general population, and utility scores for use in cost-effectiveness evaluation, may be important decision-making criteria. Population norms are available for all of the instruments, but only the EQ-5D, SF-36 and SF-12 and HUI3 have norms from Canadian samples. Utility scores are also available for most instruments, with the exception of the NHP and WHOQoL-BREF, but value weights from Canadian samples are only currently available for the EQ-5D and HUI3.

Discussion

Summary and key findings

Our research team collaborated closely with colleagues working in primary care and those in health policy positions at regional and provincial levels. Further, we consulted widely with academic partners with knowledge and experience of PROM use in primary and community care. The culmination of the project was an end-of-project knowledge translation workshop, hosted by the BC Ministry of Health. In sharing our findings with health sector evaluators, researchers and administrators, we asked for views on the top instruments for use in a primary/community care setting. A “dotmocracy” process was used whereby all workshop participants were able to express their overall instrument preference using dot stickers. The highest-ranking instruments were SF-36, EQ-5D and PROMIS/GHS.

Based on the review of evidence presented in this paper, and when thinking specifically about PROMs in the context of primary and community care, the SF-36 has a number of strengths, chiefly its strong psychometric properties (particularly its responsiveness) and widespread usage facilitating comparisons, as well as the availability of population norms and utility scores. The EQ-5D scored satisfactorily on many of the criteria we considered, and has a lower response burden than the SF-36. However, for evaluation research and quality improvement purposes, there is a concern relating to responsiveness and hence its ability to detect change. It may be better suited than the SF-36 where brevity outweighs the importance of responsiveness. The PROMIS/GHS also scored highly and should be given serious consideration. Despite its smaller evidence base, its relatively strong psychometric properties, absence of licensing fees, increasing utilization and integration into a broader information system are all strong arguments in its favour (Cella et al. 2010).

The selection of a PROM instrument is a complex task that will inevitably involve trade-offs. The time for completion of the instrument is, not surprisingly, correlated with the number of items included in the instrument, which also influences its ability to detect changes when they occur (see psychometric results above). For example, the EQ-5D takes less time to complete than the SF-36, but is also less responsive to detecting change. Decision-makers must, therefore, consider the relative importance of being able to detect changes, which is crucial for use in evaluation and quality improvement, with the burden placed on respondents.

Despite general acknowledgement that the responsiveness of the SF-36 was a strong point compared to other instruments, there was concern that the instrument took too long to complete. Practical considerations pertaining both to how it would be administered, and by whom, featured prominently in the discussion. Some workshop attendees expressed a preference for the EQ-5D, despite its lower responsiveness, because it would be quicker to administer and therefore a lesser burden to patients. Cost was another important consideration and was considered a particular strength of the PROMIS/GHS instrument.

Limitations

We acknowledge the following limitations of this rapid review. A systematic review may have

produced a greater breadth and depth of evidence relating to generic PROMs applicable to a primary care setting. Specifically, a search of the primary literature may have uncovered other PROMs not identified in our search. Further, we may have missed some papers providing more depth compared with the review papers selected that addressed the psychometric properties of the eight instruments. A quality assessment of the selected papers was not undertaken and thus the conclusions need to be interpreted with caution. We made a pragmatic decision to focus on versions of instruments that were the latest or most commonly used within a primary/community care setting. Ideally, our review would have included all available versions to first establish which is most scientifically robust within a primary care setting to allow all instruments equal opportunity to emerge into practice. Although we report that there were no psychometric weaknesses for the SF-36, we highlight the following feasibility considerations. The SF-36 has a longer completion time, which may predispose this instrument to reduced response rates and greater cumulative missing data, but we did not review such evidence in this project. Finally, the quality of the available evidence detailing population norms was not evaluated in this rapid review due to time constraints. As such it is possible that for some of the instruments, there may be issues relating to the external validity of these published norms. The final limitation pertains to the observed data extraction discrepancies between J.C.D. and J.B. These discrepancies were largely due to the heterogeneity of the included studies. As such, a third reviewer (S.B.) was added to resolve all differences.

Conclusions

The results provide a summary of the characteristics of and evidence base for some of the most commonly used generic PROM instruments. Identifying the instrument best suited for an evaluation in healthcare, or for routine data collection, depends on the context (i.e., population of interest and research question) in which it would be used. By providing information pertaining to psychometric and practical considerations, we hope to encourage the wider utilization of PROMs in Canadian healthcare, and make the process of selecting a suitable PROM instrument easier.

Note

1. Some alternative instruments that share many similarities with the SF-36 (RAND-36, RAND-12, Veteran's RAND-36 and Veteran's RAND-12) are available without cost.

Acknowledgements

Other authors include Margaret J. McGregor, MD, Associate, Centre for Clinical Epidemiology & Evaluation, Vancouver Coastal Health Research Institute, Clinical Assistant Professor, Department of Family Practice, The University of British Columbia, Associate, Centre for Health Service & Policy Research, The University of British Columbia; Janice

M. Murphy, PhD, Research Consultant, Balfour, BC; and Rick Sawatzky, PhD, Associate Professor, School of Nursing, Trinity Western University, Scientist, Centre for Health Evaluation and Outcome Sciences, Providence Healthcare, Langley, BC.

This study was supported by a grant from the Canadian Institutes of Health Research. The authors would also like to thank their other research team colleagues, Kacey Dalzell and Martin Dawes, and colleagues at the BC Ministry of Health and the Michael Smith Foundation for Health Research.

Correspondence may be directed to: Stirling Bryan, PhD, Centre for Clinical Epidemiology & Evaluation, Vancouver Coastal Health Research Institute, 701 – 828 West 10th Avenue, Vancouver, BC V5Z 1M9; e-mail: stirling.bryan@ubc.ca.

References

- Anderson, R.T., N.K. Aaronson, M. Bullinger and W.L. McBe. 1996. "A Review of the Progress Towards Developing Health-Related Quality-of-life Instruments for International Clinical Studies and Outcomes Research." *Pharmacoeconomics* 10(4): 336–55.
- Anderson, R.T., N.K. Aaronson and D. Wilkin. 1993. "Critical Review of the International Assessments of Health-Related Quality of Life." *Quality of Life Research* 2(6): 369–95.
- Bouchet, C., F. Guillemin, A. Paul-Dauphin and S. Briancon. 2000. "Selection of Quality-of-Life Measures for a Prevention Trial: A Psychometric Analysis." *Controlled Clinical Trials* 21(1): 30–43. doi: 10.1016/S0197-2456%2899%2900038-0.
- Brazier, J., M. Deverill, C. Green, R. Harper and A. Booth. 1999. "A Review of the Use of Health Status Measures in Economic Evaluation." [Comparative Study Research Support, Non-US Gov't Review]. *Health Technology Assessment* 3(9): i-iv, 1–164.
- Bryan, S., J. Broesch, K. Dalzell, J. Davis, M. Dawes, M.M. Doyle-Watters et al. 2013. "What Are the Most Effective Ways to Measure Patient Health Outcomes of Primary Health Care Integration through PROM (Patient Reported Outcome Measurement) Instruments?" Retrieved November 5, 2014. <<http://www.c2e2.ca>>.
- Butterworth, P. and T. Crosier. 2004. "The Validity of the SF-36 in an Australian National Household Survey: Demonstrating the Applicability of the Household Income and Labour Dynamics in Australia (HILDA) Survey to Examination of Health Inequalities." [Research Support, Non-U.S. Gov't Validation Studies]. *BMC Public Health* 4: 44. doi: 10.1186/1471-2458-4-44.
- Cella, D., W. Riley, A. Stone, N. Rothrock, B. Reeve, S. Yount et al. 2010. "The Patient-Reported Outcomes Measurement Information System (PROMIS) Developed and Tested its First Wave of Adult Self-Reported Health Outcome Item Banks: 2005–2008." *Journal of Clinical Epidemiology* 63(11): 1179–94.
- Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN). 2013. Retrieved April 1, 2013. <<http://www.cosmin.nl>>.
- Coons, S.J., S. Rao, D.L. Keininger and R.D. Hays. 2000. "A Comparative Review of Generic Quality-of-Life Instruments." [Comparative Study Research Support, Non-US Gov't Review]. *Pharmacoeconomics* 17(1): 13–35.
- Dawson, J., H. Doll, R. Fitzpatrick, C. Jenkinson and A.J. Carr. 2010. "The Routine Use of Patient Reported Outcome Measures in Healthcare Settings." *BMJ* 340: c186.
- Doward, L.C., D.M. Meads and H. Thorsen. 2004. "Requirements for Quality of Life Instruments in Clinical Research." *Value Health* 7(1 Suppl): S13–16. doi: 10.1111/j.1524-4733.2004.7s104.x.
- Ford, M.E., S.L. Havstad, D.D. Hill and C.S. Kart. 2000. "Assessing the Reliability of Four Standard Health Measures in a Sample of Older, Urban Adults." *Research on Aging* 22(6): 774–96.

Choosing Your Partner for the PROM: A Review of Evidence on Patient-Reported Outcome Measures for Use in Primary and Community Care

- Ford, M. E., S.L. Havstad and C.S. Kart. 2001. "Assessing the Reliability of the EORTC QLQ-C30 in a Sample of Older African American and Caucasian Adults." [Research Support, US Gov't, P.H.S. Validation Studies]. *Quality of Life Research* 10(6): 533–41.
- Furlong, W.J., D.H. Feeny, G.W. Torrance and R.D. Barr. 2001. "The Health Utilities Index (HUI) System for Assessing Health-Related Quality of Life in Clinical Studies." [Review]. *Annals of Medicine* 33(5): 375–84.
- Gandek, B., S.J. Sinclair, M. Kosinski and J.E. Ware Jr. 2004. "Psychometric Evaluation of the SF-36 Health Survey in Medicare Managed Care." *Health Care Financing Review* 25(4): 5–25.
- Hawthorne, G., S. Korn and J. Richardson. 2013. "Population Norms for the AQoL Derived from the 2007 Australian National Survey of Mental Health and Wellbeing." [Research Support, Non-US Gov't]. *Australian and New Zealand Journal of Public Health* 37(1): 7–16. doi: 10.1111/1753-6405.12004.
- Hawthorne, G. and J. Richardson. 2001. "Measuring the Value of Program Outcomes: A Review of Multiattribute Utility Measures." *Expert Review of Pharmacoeconomics and Outcomes Research* 1(2): 215–28. doi: <http://dx.doi.org/10.1586/14737167.1.2.215>.
- Hays, R.D., J.B. Bjorner, D.A. Revicki, K.L. Spritzer and D. Cella. 2009. "Development of Physical and Mental Health Summary Scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) Global Items." *Quality of Life Research* 18(7): 873–80.
- Haywood, K.L., A.M. Garratt and R. Fitzpatrick. 2005. "Quality of Life in Older People: A Structured Review of Generic Self-Assessed Health Instruments." *Quality of Life Research* 14(7): 1651–68.
- Horsman, J., W. Furlong, D. Feeny and G. Torrance. 2003. "The Health Utilities Index (HUI): Concepts, Measurement Properties and Applications." *Health and Quality of Life Outcomes* 1: 54. doi: 10.1186/1477-7525-1-54.
- Kopec, J.A. and K.D. Willison. 2003. "A Comparative Review of Four Preference-Weighted Measures of Health-Related Quality of Life." *Journal of Clinical Epidemiology* 56(4): 317–25.
- McGrail, K., S. Bryan and J. Davis. 2011. "Let's All Go to the PROM: The Case for Routine Patient-Reported Outcome Measurement in Canadian Healthcare." *HealthcarePapers* 11(4): 8–18; discussion 55–18.
- McHorney, C.A. and A.R. Tarlov. 1995. "Individual-Patient Monitoring in Clinical Practice: Are Available Health Status Surveys Adequate?" *Quality of Life Research* 4(4): 293–307. doi: 10.1007/BF01593882.
- Microsoft. 2013. "Test Your Document's Readability." Retrieved January 6, 2013. <<http://office.microsoft.com/en-ca/word-help/test-your-document-s-readability-HP010148506.aspx>>.
- Mokkink, L.B., C.B. Terwee, D.L. Patrick, J. Alonso, P.W. Stratford, D.L. Knol et al. 2010. "The COSMIN Study Reached International Consensus on Taxonomy, Terminology, and Definitions of Measurement Properties for Health-Related Patient-Reported Outcomes." *Journal of Clinical Epidemiology* 63(20494804): 737–45.
- PROQOLID (Patient-Reported Outcome and Quality of Life Instruments Database). 2013. Retrieved September 14, 2014. <<http://www.proqolid.org/>>.
- Revicki, D.A. and R.M. Kaplan. 1993. "Relationship between Psychometric and Utility-Based Approaches to the Measurement of Health-Related Quality of Life." *Quality of Life Research* 2(6): 477–87.
- Sintonen, H. 2001. "The 15D Instrument of Health-Related Quality of Life: Properties and Applications." *Annals of Medicine* 33(5): 328–36.
- Skevington, S.M., M. Lotfy and K.A. O'Connell. 2004. "The World Health Organization's WHOQoL-BREF Quality of Life Assessment: Psychometric Properties and Results of the International Field Trial, a Report from the WHOQoL Group." *Quality of Life Research* 13(2): 299–310. doi:10.1023/B:QURE.0000018486.91360.00.
- Terwee, C.B., E.P. Jansma, II. Riphagen and H.C. de Vet. 2009. "Development of a Methodological PubMed Search Filter for Finding Studies on Measurement Properties of Measurement Instruments." *Quality of Life Research* 18(8): 1115–23. doi: 10.1007/s11136-009-9528-5.
- Ware, J.E. 2000. "SF-36 Health Survey Update." *Spine* 25(24): 3130–39. doi:10.1097/00007632-200012150-00008.
- Wiklund, I. 1990. "The Nottingham Health Profile – A Measure of Health-Related Quality of Life." [Review]. *Scandinavian Journal of Primary Health Care* 1: 15–18.