Quality of Life Research (2019) 28:879–893
https://doi.org/10.1007/s11136-018-2046-6

REVIEW

Individualized health-related quality of life instrument Measure Yourself Medical Outcome Profile (MYMOP) and its adaptations: a critical appraisal

Sana Ishaque[1,2,3] · **Jeffrey A. Johnson**[1] · **Sunita Vohra**[2]

Accepted: 3 November 2018 / Published online: 13 November 2018

Abstract

Background Health-related quality of life (HRQL) is increasingly recognized for its importance in health research. As there is increasing recognition of the inter-individual difference in response to therapeutic interventions, it may be helpful to apply individualized measures of HRQL. The MYMOP is a commonly used example of such measures, although several adaptations have been developed.

Objective This review was conducted to identify adaptations of MYMOP, and evaluate the measurement properties of MYMOP and its adaptations.

Methods Adaptations were identified using MYMOP website and personal communication, supplemented by a SCOPUS search in April 2017. Bibliographies of included studies were hand-searched. COSMIN criteria were used to evaluate the measurement properties.

Results Sixteen studies were included in this review. Adaptations were developed to evaluate individualized therapies in cancer, psychiatry, and acupuncture. The included measures were MYMOP, measure yourself concern and wellbeing, psychological outcome profiles (PSYCHLOPS), and MYMOP-pictorial (MYMOP-P). The quality of the measurement properties varied; none of the included measures met all currently recommended quality criteria for measurement properties.

Conclusion Current literature provides evidence that MYMOP and its adaptations offer individualized assessment of patient-centered outcomes, and thereby provide a means to understand heterogeneity of treatment effects. However, current recommendations for psychometric testing suggest further validation of these measures would be beneficial.

Keywords Health-related quality of life · Quality of life · HRQL · HRQOL · QOL patient-generated · Individualized · Patient-centered · Domain specific

Abbreviations

CORE-OM	Clinical outcomes routine evaluation-outcome measure
COSMIN	Consensus-based standards for the selection of health status measurement instruments
EQ-5D	EuroQol Group health status index 5-dimensions
FACIT-SpEx	Functional assessment of chronic illness therapy questionnaire-spiritual subscale
HRQL	Health-related quality of life
ICC	Intraclass correlation coefficient
MOS-6A	Medical outcome study 6-item general health survey
MYCaW	Measure yourself concerns and wellbeing
MYMOP	Measure yourself medical outcome profile
MYMOP-P	MYMOP-pictorial
PSYCHLOPS	Psychological outcome profiles
SD	Standard deviation
SF-36	Medical outcomes study 36-item short-form health survey

✉ Sunita Vohra
svohra@ualberta.ca

[1] School of Public Health, University of Alberta, Edmonton, AB, Canada

[2] Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

[3] School of Public Health, University of Adelaide, Adelaide, SA, Australia

Introduction

Health-related quality of life (HRQL) has grown in its importance as an essential outcome for patient-centered research [1]. Advances in medical research have resulted in prolonged survival for those with chronic diseases, making the patient's experience vital to assessment of therapeutic effectiveness. Arguably, effective therapies not only alleviate the patient's signs or symptoms, but also make a significant difference in their HRQL.

According to the ISOQOL Dictionary of quality of life and health outcomes measurement "HRQL is a measure of the value assigned to duration of life as modified by impairments, functional states, perceptions and opportunities, as influenced by disease, injury, treatment and policy [2]."

Measurement approaches to assess HRQL can be broadly grouped into two: (a) traditional measures with predetermined domains, (b) individualized measures with real time patient-selected domains. Both methods have their pros and cons and therefore, the approach taken to assess HRQL may vary according to the aim of the measurement.

Traditional HRQL measures, with standardized set of questions, are convenient tools for group comparisons. These measures are also useful for economic evaluation of new or equally effective health care interventions. However, traditional measures may not represent all health domains valued by each individual patient [3–6]. Some researchers have expressed concern about the lack of patient-centeredness of traditional HRQL measures [7]. The predetermined questions on traditional measures may not be relevant to particular patients at different stages of their disease [6–8]. Moreover, what may be important for one patient may not have similar value for another patient [9]. Personalized approaches to treatment and variation in patient characteristics such as age, gender, disease severity and other environmental and genetic factors also enhance potential differences between treatment effects that a particular therapy may produce [9, 10]. To avoid the complexity of heterogeneous treatment effects [11], individual patient data are thought to be better captured by generic individualized outcome instruments that allow each patient to determine and measure what is important to him/her during a clinical consultation. Similar to any other measurement approach, individualized measures also have some limitations. Individualized measures cannot be used for economic evaluation as well as their scores cannot be used for comparability between individual patients or group of patients. This lack of comparability of scores has been a topic of tension and debate for a while and therefore it remains unclear whether the psychometric criteria that investigate the cross-sectional comparability of the scores of patient-reported outcome measures (such as structural validity and internal consistency) are applicable to individualized tools. Therefore, these psychometric criteria are not discussed in this review.

In the ISOQOL Dictionary of definitions, individualized measures are defined as "measures that allow patients to identify domains (or areas of life) that are important to them, and then to assign a weight on the relative importance of each one [2]." Measure yourself medical outcomes profile (MYMOP) is an individualized measure that allow patients to nominate and score two most important aspects of their lives (in the order of their importance) that contribute most to their overall quality of life but does not ask respondents to weigh their nominated domains. Presumably the wording on the questionnaire leads patients to name two top most important aspects of their life in the order of their importance and thereby weighting is implicit and not formalized.

Examples of commonly used individualized measures include: Schedule for the Evaluation of Individual Quality of Life-Direct Weighting [12–16] and Patient Generated Index [12, 17–19]. Critical analysis of the properties of SEIQoL and PGI has been reported in the literature as standalone measures [13, 17] and also in the context of a number of health conditions [12, 18]. Paterson et al.'s MYMOP [20] furthered the concept of individualized measures. The MYMOP has been invaluable since it is patient-centered and patient-completed; in this way, it is very different from clinical practice in which the clinician's treatment goals may drive the questions they choose to ask of patients. Despite MYMOP being in use since 1996 [20], there is no critical review of the properties of measure performed to date. The purpose of this paper is to critically appraise the measurement properties of MYMOP and its adaptations.

Methods

Search strategies

A SCOPUS search for articles' titles, abstracts, and keywords was conducted up to April 2017 with the name of adaptations of MYMOP. The names were identified using the MYMOP website and personal communication with instruments developers. The search terms included MYMOP, MYCaW, PSYCHLOPS, 'MYMOP-pictorial,' and MYMOP-P; only English language adaptations were included in the review. Besides, to identify additional publications, the reference lists of the included articles and publication list on each instrument's primary websites was also scanned.

Finally, abstracts were screened to identify studies conducting formal psychometric evaluation, or qualitative evidence collection to validate the instruments of interest.

Quality assessment

We evaluated the results of measurement properties for each measure, identified, using the COSMIN checklist for systematic reviews of patient-reported outcome measures [21–23] and COSMIN taxonomy [24]. There are three domains of measurement properties: reliability, validity, and responsiveness [24–29]. Reliability is further subdivided into internal consistency, reliability, and measurement error. Validity is subdivided into content validity, construct validity, and criterion validity. The possible overall rating for each measurement property is "positive" (+), "indeterminate" (?), "negative" (−), or "no information available" (0) (Table 4).

Results

The Scopus search yielded 111 unique studies; an additional 28 studies were identified from questionnaires' websites. After screening the title, abstract, and keywords, we retrieved 34 articles in full text. We finally included 16 studies, evaluating four questionnaires (MYMOP and three adaptations: MYCaW, PSYCHLOPS, and MYMOP-P) [20, 30–43]. The new questionnaires were adapted for evaluation of therapies in cancer [42, 43], psychiatry [31–34, 37], and acupuncture [35, 36]. Table 1 presents the general characteristics of these studies. It is notable that 10 of 16 were applied to evaluate effectiveness of complementary therapies.

Measure yourself medical outcome profile (MYMOP)

MYMOP is a problem specific, individualized measure that was developed in a primary care setting (Table 2) [20]. Each patient is asked to report two symptoms that bother them the most over the previous week, one activity limited by the reported symptoms, and general wellbeing. After an initial pilot study, a brief medication questionnaire was added to the scale [41]. However, medication questions are not scored and thus do not contribute to the final MYMOP score [44]. The overall score is calculated by taking the average of item scores, and is interpreted in the presence of individual item scores. For meaningful comparison, the items chosen must remain unchanged between the first and the subsequent completion of the questionnaire.

Quality assessment of MYMOP

We did not identify any studies evaluating measurement error, floor, ceiling effect, and interpretability of the MYMOP. Three studies assessed content validity (Table 3). The first of these gives clear description of the measurement aim and information on the target population [20]. The second study, [39] gathered patients' views about MYMOP's ability to measure outcomes that are important to them. This study compared the qualitative interview data of 20 interviewees to their corresponding quantitative MYMOP score [39]. Incorporation of participants and practitioners' views resulted in the development of the current version called "MYMOP 2." The third study exploring content validity [40] involved interviewing 23 new patients of eight acupuncturists. They used two qualitative analytical techniques: focus groups, and cognitive interview. The issues identified about MYMOP2 were floor effect, inability to measure episodic symptoms, and inaccurate measurement of medication change. No revisions of MYMOP2 were performed based on the study results [40].

Construct validity was assessed in two studies by examining the correlation between "perceived change in condition" and MYMOP scores [20, 41]. Both studies confirmed the MYMOP scores correlated with the perceived change in condition. Similar results were observed for the correlation of clinical-outcome assessed by physicians and MYMOP scores [41]. Also, MYMOP scores of individuals with acute conditions and those with chronic conditions were compared; it was hypothesized that changes in MYMOP score would correlate well with changes in acute conditions (<4 weeks) rather than chronic conditions (>4 weeks). This correlation was confirmed [20]. In addition, expected correlations of MYMOP and SF-36 scores were also reported [20].

Responsiveness of MYMOP was determined by gradient change in score at repeat applications across perceived changes by clinicians [20] and by patients [20, 41]. Standardized response mean, and index of responsiveness were also reported [20, 41]. A *t* test was conducted to compare the scores of patients who described themselves as a "little better" to "about the same, [41]" and gradient changes in scores at two and four weeks were determined [20]. The authors applied the SF-36, MOS-6A, and EQ-5D, simultaneously to the study population, but did not report correlation coefficients for changes [20, 41].

Measure yourself concerns and wellbeing (MYCaW)

MYCaW [42, 43, 45] was adapted from MYMOP to evaluate cancer patients undergoing integrative treatments (Table 2). Like MYMOP, it allows patients to define and measure their two most important concerns and general wellbeing on a seven-point ordinal scale; higher score signifies poorer health [46]. MYCaW also has pictorial faces, and the wording added at the each end of the seven-point scale: "not bothering me at all $=0$," "bothers me greatly $=6$" [46]. There

Table 1 Study characteristics

Instrument	Study and country	Study design	Sample size and age (range)	Target population (diagnosis/underlying CONDITION)	Practice setting/context	Psychometric property (ies) involved/evaluated	"Gold Standard" or comparator	Application (time)
MYMOP	Paterson et al. (1996), the UK [20]	Longitudinal	$N=265$, 218 general practice patients, 47 CAM patients (2–84 years; mean and SD: 47 ± 17.6); 109 completed the f/u	Evaluation; patients receiving complementary med	Primary care (GPs and CAM providers)	Content validity, construct validity (hypothesis testing), responsiveness	SF- 36	0, 2, 4, 8, and 16 week
	Paterson et al. (2000), Scotland [41]	Longitudinal	$N=81$ (32.4–82.8 years)	Evaluation; chronic bronchitis with acute exacerbation	General practice (Glasgow, Scotland)	Responsiveness, construct validity (hypothesis testing)	MOS-6A, EQ-5D	0, and after 1 week within completion of treatment
	Paterson et al. (2000), the UK [39]	Longitudinal	176 patients completed MYMOP, 20 interviews; age: 16–86 years	Evaluation	New patients of 12 complimentary practitioners Somerset UK	Qualitative analysis to assess content validity	Semi-structured interviews	MYMOP was completed for a minimum of twice and maximum of nine times over 9 months
	Paterson et al. (2004), the UK [40]	Longitudinal	23 patients: 64 interviews; age: 26–83 years	Evaluation of primary care acupuncture patients	New patients of acupuncturists	Qualitative analysis to assess content validity	NA	Thrice to 18 patients, twice to 4 patients
MYCaW	Peace and Manasse (2002), the UK [45]	Longitudinal	$N=157$ (18–86 years; peak 50–59 years)	Evaluation of cancer patients and their caregivers	Multidisciplinary clinic (complementary and alternative med)	Content validity	None	Time period not defined (at first visit and time of completion of treatment)
	Cooke (2000), the UK [38]	Dissertation	$N=100$ (81 patient + 19 carers), 61% completed f/u at 3rd wk, 40% at 3rd month; 18–70+	Evaluation of cancer patients (or various primary and secondary cancers) taking CAM along with conventional medicine	Multidisciplinary cancer support	Content validity	None	0, 3 weeks, and 3 months
	Paterson et al. (2007) [42]	Publication based on findings of Peace and Manasse [45] and Cooke [38]	Cavendish Center: $N=254$ (21–84 years); Bristol center: $N=267$	NA	Cavendish Center Bristol center	NA	NA	NA
	Jolliffe et al. (2015) [47]	Longitudinal	$N=82$ (18 to >80 years)	Cancer patients and their carer/supporters attending "Living Well with cancer courses"	Penny Brohn Cancer Care UK	Construct validity (hypothesis testing), responsiveness	FACIT-SpEx	0 and after 6 week

Table 1 (continued)

Instrument	Study and country	Study design	Sample size and age (range)	Target population (diagnosis/underlying CONDITION)	Practice setting/context	Psychometric property (ies) involved/evaluated	"Gold Standard" or comparator	Application (time)
PSYCHLOPS	Ashworth et al. (2004), the UK [30]	Group consultations re adaption	Not reported	Primary care based psychotherapy patients	NA	Content validity	NA	NA
	Ashworth (2005), the UK [33]	Survey	Four primary care mental health practitioners; age: not reported	Primary care patients undergoing talk/psychotherapy	Primary care	Content validity	CORE-OM	NA
	Ashworth (2005), the UK [34]	Longitudinal	$N=235$ completed pre-therapy questionnaires; $N=110$ post-therapy; age for whom complete set available for analysis: 15–64 years	Patients entering psychotherapy in primary care	Primary care	Internal consistency (Cronbach's alpha), construct validity (hypothesis testing), interpretability, responsiveness [32]	CORE-OM	Pre- and post-therapy (no time interval specified)
	Ashworth (2007), the UK [32]	Cross-sectional	$N=215$; 16–64 year (<18 included if not full time students)	Patients entering to primary care mental health	Primary care	Qualitative analysis performing comparison of PSYCHLOPS responses to CORE-OM (content validity)	CORE-OM	NA
	Ashworth (2009), the UK [31]	Longitudinal	$N=336$, complete data available for $N=114$; 17–75 years	Primary care psychological therapy patients	Primary care clinical psychologists performing talking therapy based on CBT model	Internal consistency (Cronbach's alpha), construct validity (hypothesis testing), interpretability, responsiveness [29]	HADS	Pre- and post-therapy (no time interval specified)
	Evans (2010), the UK [37]	Longitudinal	$N=73$ (1st time responders), $N=56$ (completed both 1st and 2nd rounds); age not reported	Students from three institutes from London	Non clinical sample	Internal consistency, reliability	None	0, 1–2 weeks later
MYMOP-pictorial (MYMOP-P)	Day (2004), the UK [36]	Audit of acupuncture	$N=62$ initial form, 55 f/u completed (23–80 years)	Patients undergoing acupuncture	Varying complaints	Content validity	None	Not specified
	Day (2004), the UK [35]	Article explaining MYMOP-P	n/a	n/a	n/a	n/a	n/a	n/a

are two versions, self-administrated and face-to-face interview scale. Each version has initial and follow-up forms. The questionnaire consists of three scored domains, two of which are individualized. The followup form includes two open-ended questions: "other things affecting your health" and "reflecting on your time with (service name) what were the most important aspects for you? [42]." MYCaW provides quantitative (mean change in score and SD), and qualitative data.

Quality assessment of MYCaW

Adaptation and validation of MYCaW started in 2002 [45] (Table 3). Initial draft, for content validation, was discussed with experts and patient-representatives resulting in subsequent revision to the layout and wording of the instrument [42, 45]. A later study defined minimal important change for the interpretation of scores as 0.5, 1, and 1.5 as minimal, moderate, and large, respectively [38].

Construct validity of MYCaW was evaluated by testing a priori hypothesized negative correlation of $r > 0.3$ with functional assessment of chronic illness therapy questionnaire-spiritual subscale (FACIT-SpEx) [47]. The FACIT-SpEx is an expanded version of the FACIT questionnaire. In addition to physical, social/family, emotional, and functional wellbeing, it also includes questions on spiritual wellbeing relating to cancer therapy. The reported results confirmed a correlation of $r = -0.57$ [47].

Responsiveness indices reported were standardized response mean and effect size of baseline and 6-week MYCaW and FACT-SpEx patient scores [47]. The Guyatt's responsiveness index for MYCaW concern 1, 2, wellbeing, and overall profile were grouped according to five predefined categories on FACIT-SpEx scale. The categories were as follows: 'substantial improvement,' 'clinically relevant improvement,' 'stable,' 'clinically relevant deterioration,' and 'substantial deterioration.' Scores on MYCaW were consistent with the categories except for the 'stable' group. The category of 'clinically relevant deterioration' did not have enough participants to analyze.

One of the advantages of MYCaW is its ability to capture range of qualitative information at individual level [42]. There have been substantial efforts to provide a frame of analysis for the rich qualitative information gathered by the questionnaire [43, 48]. Three questions of MYCaW were qualitatively analyzed: (i) "concerns and problems" question on the first form; (ii) "other things affecting your health," and (iii) "what has been most important for you?" of the follow-up form. Sample of 782, 407, and 588 patients reported on "concerns and problems," "other things affecting your health," and "what has been important for you?" respectively. Their responses were organized into categories and a qualitative analysis guideline for MYCaW was developed; a focus group of five women validated the categories for appropriateness and acceptability. Four of the women who participated in the focus group had cancer, and one of them was a caregiver of a cancer patient. Later, for generalizability of the coding framework it was reviewed by mapping data from Penny Brohn Cancer Care UK and Ottawa Integrative Cancer Clinic Canada. As a result, some new categories under 'physical concern,' 'hospital cancer treatment concerns,' 'concerns about wellbeing,' and 'practical concerns' were identified.

Psychological outcome profiles (PSYCHLOPS)

PSYCHLOPS is an individualized mental health outcome measure [30]. Similar to MYMOP, PSYCHLOPS measures the score of unique issue(s) for an individual (Table 2). PSYCHLOPS is a one-page questionnaire [49] that consists of three domains: problems, function, and wellbeing. The questionnaire has three versions: pre-therapy, during-therapy, and post-therapy. Four questions are common to each version. The initial two questions ask patients to identify and measure their most bothersome problems, the third identifies and measures one function limited due to the identified problem(s), and fourth is about general wellbeing over the last week. A fifth question in the during-therapy version identifies any new problem that arises amidst therapy. A sixth question on the post-therapy version asks the patients to score how they feel compared to the start of therapy. PSYCHLOPS does not assign a score to every question. The questions related to Problems, Functioning and Wellbeing have six-point (0–5) scales, where higher score signify worse outcomes. The "individually identified" items from the initial form are transferred to the subsequent versions for patient to re-score them. This process provides changes in score from pre- to post-therapy [49].

Quality assessment of PSYCHLOPS

A group of clinical psychologists, counseling psychologists, psychotherapists, counselors, general practitioners, and academic mental health researchers interested in mental health started adaptation of PSYCHLOPS in 2004 (Table 3) [30].

Content validity was assessed by consulting patient representatives, and three expert groups. The initial draft was piloted to 30 patients [30], and it was revised as required [30]. In 2005 (Table 2), Ashworth et al. gathered information about the feasibility, validity, and usefulness of PSYCHLOPS from experts [33]. Internal consistency was determined via Cronbach's alpha, and the values were within acceptable range [31, 34, 37].

In terms of construct validity, PSYCHLOPS has moderate to strong correlation with clinical outcomes routine evaluation-outcome measure (CORE-OM) [34] and Hospital

Table 2 Description of included measures

Included measure(s)	Definitions	MYMOP [20, 39–41, 44]	MYCaW [38, 42, 45–47]	PSYCHLOPS [30–34, 37]	MYMOP-pictorial [35, 36]
Version(s) available		Two: initial form and follow-up form	Two: face to face, and self completion version; each version has initial and f/u forms	Three: pre-therapy, during therapy, post-therapy	One
Construct	Description of what the questionnaire intends to measure	Physical symptoms, activity, wellbeing	Concern/problems, wellbeing	Mental health outcomes	Same as MYMOP
Domain	A domain or dimension refers to the area of behavior that we are trying to measure	Symptom(s), and activity as chosen by patient; wellbeing	Concern(s) chosen by patient; wellbeing	Problems, and function chosen by patient; wellbeing	Same as MYMOP
Setting (e.g. Clinical, general population, epidemiological study)	In what setting the measurement was made?	Clinical trial, primary care [20] Clinical trial, general practice [39]	Observational, multidisciplinary clinic CAM [45]; questionnaire development, multidisciplinary clinic CAM	Clinical and general population	Clinical (acupuncture patients)
Recall Period	What is the recall period to which the questionnaire refers	Last week (last 7 days)	Current concerns are asked	Range from under one month to over five year	Same as MYMOP
Purpose (evaluative, discriminative, or both)	Purpose of questionnaire	Evaluative	Evaluation	Evaluation of primary care mental health patients undergoing psychotherapy	Evaluation
Target population diagnosis, age	For the kind of people questionnaire was originally developed	$N=265$ (2–84 years; mean and SD: 47 ± 17.6); 109 completed the f/u patients receiving CAM [20] $N=81$ (32.4–82.8 years) patients of chronic bronchitis with acute exacerbation [41]	$N=157$ (18–86 years; peak 50–59 years), cancer patients [45] Cavendish Center: $N=254$ (21–84 years), Bristol center: $N=267$, cancer patients	Primary care psychotherapy sessions' patient(s)	Acupuncture patients suffering from various conditions
Mode of administration (e.g. self, interview, proxy administered if proxy administered: name of proxy e.g. parent or health care provider)		Self administered	Self and interview administered	Self administered	Self administered
# of items	Number of questions in a questionnaire	Eight on the initial form, four scored four un-scored; six on the f/u form four of which are scored	Three scored items (of these two are patient generated) and two open-ended questions	Six items on pre-therapy and post-therapy questionnaires: three preset and three patient/individual specific; five items on during therapy: two preset and four individualized	Same as MYMOP

Table 2 (continued)

Included measure(s)	Definitions	MYMOP [20, 39–41, 44]	MYCaW [38, 42, 45–47]	PSYCHLOPS [30–34, 37]	MYMOP-pictorial [35, 36]
# and type of response options	Scale type	0–6 point scale (seven points)	0–6 point scale with smiley face adjacent to "0," and sad face adjacent to "6"	0–5 point scale	Six point faces scale
Time to complete	Average time to complete the questionnaire	Not specified	Not specified	Not specified	Not specified
Full copy available for free		Yes	Yes	Yes	Yes
Instructions (not described, clearly described, unclear)		Clear description on filling and scoring are available	Completion and scoring methods clearly described	Clearly described	Not reported
Country (related to cross-cultural validity)		The UK	The UK	The UK	The UK
Translation/cultural adaptations available		Yes (refer to MYMOP website)	None	None	None
Generic or specific (disease of population specific)		Problem specific (individualized)	Problem specific	Specific—condition specific	Problem specific

Anxiety Depression Scale [22]. Responsiveness was defined as "sensitivity to change" and was measured by effect size [31, 34]. Interpretability was assessed by mean and SD of pre- and post-therapy scores [31, 34]. Test–retest reliability was reported as intraclass correlation coefficients (ICC) between baseline and retest as 0.70, 0.68, 0.69, and 0.79 for problems domain, activity that was hard-to-do, wellbeing, and overall score respectively [37]. The study participants for reliability assessment were healthy individuals and remained stable during the interim period.

In 2007, Ashworth analyzed if the preset items on CORE-OM identify the individualized PSYCHLOPS responses [32]. There were 611 individual responses on PSYCHLOPS and the responses were categorized into 8 themes and 61 sub-themes. Of 61 sub-themes, 27 (44%) were not mapped to preset questions of CORE-OM. Of 215 clients, 128 (60%) reported at least one response that could not be mapped to CORE-OM.

MYMOP-pictorial (MYMOP-P)

MYMOP-P was developed to assess elderly patients' outcomes (Table 2) [35, 36]. During the study [36], the author found that patients who were "elderly," "having low confidence in completing forms," "low literacy," or "mother tongue not English" were not able to fill MYMOP2 properly. To solve this issue MYMOP-P was developed. The measure has six points scale (0–5) that range from "as good as it could be" to "as bad as it could be." Each response option has a "face" that corresponds to the current state of patient, and patients are asked to choose one face in order to score their reported issue. The author did not explain the method of questionnaire adaptation any further, it is not clear if any patient representatives were involved in the development process. To our knowledge, no formal evaluations of the instrument's measurement properties are reported yet.

Discussion

In this article, we reviewed the format, content and evidence of measurement properties for MYMOP [20, 37–39], and its three adaptations [30–43, 45–49]. Of these measures, PSYCHLOPS was the most thoroughly evaluated [30–34, 37, 49], and therefore had the greatest evidence of its measurement properties, including test–retest and internal consistency reliability. To our knowledge, MYMOP-P [35, 36] has had the least formal evaluation regarding its measurement properties; only reported evidence on content validity was identified in this review.

Content validity was the most widely reported measurement property [20, 30, 33, 35, 36, 38–40, 45]. Of four

Table 3 Summary of the assessment of measurement properties (based on COSMIN Criteria [21, 22])

Questionnaires	Validity			Reliability			Responsiveness	Floor or ceiling effect	Interpretability
	Content Validity	Construct Validity:, HT, CC	Criterion validity	Internal consistency	Measurement error	Reliability			
MYMOP [20, 39–41]	+	+	0	0	0	0	+	0	0
MYCaW [38, 42, 45, 47]	+	+	0	0	0	0	?	0	?
PSYCHLOPS [30–34, 37]	+	?	0	? **	0	+	?	0	?
MYMOP-P [35, 36]	?	0	0	0	0	0	0	0	0

HT hypothesis testing, *CC* cross-cultural adaptation

**Not discussed in this review

Rating: + = positive, ? = indeterminate, − = poor (negative), 0 = no information available

measures, three had positive [20, 30, 33, 35, 45], and one (MYMOP-P) had indeterminate rating [35, 36] for content validity. The reason MYMOP-P had indeterminate rating for content validity was the lack of information on what and how target population was involved in ascertainment of the relevance of the questionnaire content. The author has been contacted for unpublished data on validity more than three times, but was unreachable. Construct validity was the second commonly tested measurement property [20, 31, 34, 47]; it was reported for all measures except MYMOP-P. Evidence on construct validation was limited in terms of reporting a priori hypotheses regarding expected correlations. Modern day reporting standards for assessment of construct validity [23, 24, 50] suggests that a priori hypotheses regarding the strength and direction of the correlation also be specified. Given our results, future validation studies should consider developing and reporting a priori hypotheses for construct validity evaluation.

Criterion validity was reported for three measures in five studies [20, 31, 34, 41, 47]; however, we find that all claims of criterion validity were actually supportive of construct validity under the current definitions [24]. We find it difficult to see an instrument as a "gold standard," unless a short version of a questionnaire was tested against its long version [23, 24, 50]. Similar challenges in the evaluation of criterion and construct validity have also been highlighted in the review of PGI's measurement properties [17]. We therefore evaluated these claims as we would evaluate construct validity. Our approach did not affect the grading of the evidence. For future researchers we recommend to avoid reporting such evaluations as criterion validity, unless it involves testing a short version of a questionnaire against a long version (gold standard); when a "gold standard" does not exist, criterion validity cannot be assessed. Further, assessment against SF-36 may be considered assessment of construct validity, not criterion validity, since some would argue that SF-36 is not a universally accepted "gold standard."

Evidence internal consistency reliability is not relevant to the included measures. Internal consistency reliability is applicable for questionnaires with predetermined multidimensional domains and therefore is not calculated for individualized measures [17].

Of five studies reporting on responsiveness [20, 31, 34, 41, 47], two [31, 34] assessed responsiveness by effect sizes. We were unable to evaluate this evidence because the reported statistic did not meet the COSMIN and modified Terwee criteria for evaluation of responsiveness; both studies [31, 34] were published before these criteria were developed. Given these more recent criteria for measurement properties, we would recommend further evaluation of responsiveness of included measures. Lack of external anchor, a priori hypothesis and change in patients' priorities/concerns are the common challenges that there also identified in the evaluation of responsiveness in SEIQoL-DW [13] and PGI [17].

Another limitation of the included studies is the imprecise use of terminology to define measurement properties. This finding is not unique to these studies; Mokkink et al. [50, 51] reported similar finding in a study of quality assessment of systematic reviews of measurement properties. Of note, international consensus on taxonomy of measurement

Table 4 Quality criteria for measurement properties

Property (definitions are based on COSMIN taxonomy)	Rating	Quality criteria
Reliability: the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions		
Internal consistency: the degree of the interrelatedness among the items	+	(Sub)scale uni-dimensional AND Cronbach's alpha(s) ≥ 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	−	(Sub)scale not uni-dimensional OR Cronbach's alpha(s) < 0.70
Measurement error: the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined
	−	MIC ≤ SDC OR MIC equals or inside LOA
Reliability: the proportion of the total variance in the measurements which is due to 'true' differences between patients	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's $r \geq 0.80$
	?	Neither ICC/weighted Kappa, nor Pearson's r determined
	−	ICC/weighted Kappa < 0.70 OR Pearson's $r < 0.80$
Validity: the degree to which an HR-PRO instrument measures the construct(s) it purports to measure		
Content validity: the degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
	?	No target population involvement
	−	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Construct validity: the degree to which the scores of an HR-PRO instrument are consistent with hypotheses		
Cross-cultural: the degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument	+	Original factor structure confirmed OR no important DIF
	?	Confirmation original factor structure AND DIF not mentioned
	−	Original factor structure not confirmed OR important DIF
Structural : the degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	−	Factors explain < 50% of the variance
Hypothesis testing : idem construct validity	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	−	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Criterion validity: the degree to which the scores of an HR-PRO instrument are an adequate reflection of a 'gold standard'	+	Convincing arguments that gold standard is ''gold'' AND correlation with gold standard > 0.70
	?	No convincing arguments that gold standard is ''gold'' OR doubtful design or method
	−	Correlation with gold standard < 0.70, despite adequate design and method
	0	No information found on criterion validity

Table 4 (continued)

Property (definitions are based on COSMIN taxonomy)	Rating	Quality criteria
Responsiveness: the ability of an HR-PRO instrument to detect change over time in the construct to be measured	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70) AND correlation with related constructs is higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	−	Correlation with an instrument measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlation with related constructs is lower than with unrelated constructs
Floor and ceiling effect: the number of respondents who achieved the lowest or highest possible score	+	< 15% of the respondents achieved the highest or lowest possible scores
	?	Doubtful design or method
	−	> 15% of the respondents achieved the highest or lowest possible scores, despite adequate design and methods
	0	No information found on interpretation
Interpretability: the degree to which one can assign qualitative meaning to quantitative scores	+	+ Mean and SD scores presented of at least four relevant subgroups of patients and MIC or MID defined
	?	? Doubtful design or method OR less than four subgroups OR no MIC or MID defined
	0	0 No information found on interpretation

Based on Terwee et al.[22]

MIC minimal important change, *MID* minimal important difference, *SDC* smallest detectable change, *LOA* limits of agreement, *ICC* intraclass correlation coefficient, *DIF* differential item functioning, *AUC* area under the curve

+ = positive rating, ? = indeterminate rating, − = negative rating, 0 = no information

properties is a recent development in the field of psychometrics [24].

Strength and weaknesses of our approach

Critical appraisal is essential to evaluate medical research; it helps identify methodological strengths and limitations. Critical appraisal can be done using checklist or score-based scales. For our review, we considered appraisal tools such as Criteria by the Scientific Advisory Committee of the Medical Outcomes Trust (MOT) [52], evaluating the measurement of patient-reported outcomes (EMPRO) [53], and Terwee [22] and COSMIN criteria [21, 23]. The MOT criteria provide a list of items that instrument developers should have considered ascertaining optimal properties of their tool. However, MOT does not provide guidance on how the reported evidence should be classified if any of the listed items are absent. Evaluating the measurement of patient-reported outcomes criteria has an integral scoring system, the weighting of which is not clearly described nor explicitly justified with empiric data [53]. We used the COSMIN criteria because the COSMIN checklist was developed through a consensus-based Delphi study and has empirical evidence supporting its measurement properties [50, 51]. We preferred to use a checklist rather than a summary score because a summary score does not provide specific details on methodological strengths or limitations. A checklist approach is also preferred by the Cochrane Collaboration, based on empirical evidence that the summary scores of quality assessment tools can be problematic [54–56]. As such, Cochrane has moved from the popular use of a score-based quality assessment tool [57], to the new descriptive checklist assessment, the Risk of Bias tool [54].

Unlike a systematic review, study inclusion, data abstraction, and quality assessment were not independently duplicated in this paper. We acknowledge that lack of independent duplication can be a source of error to a review; however, single data extraction does not result in any difference in the effect estimates for many outcomes [58]. Moreover, to strengthen our critical appraisal, we chose objective checklist criteria to evaluate the quality of measurement properties, enhancing the reproducibility of our results. Although we only included studies published in English, a Chinese, and German translation of the tools were identified in the database search, demonstrating the sensitivity of our search method to identify all relevant studies. Also, the MYMOP and PSYCHLOPS websites provided contact information of 12 and 10 language translations, respectively. However, translations into other languages were not included in this review as non-English questionnaires would not be applicable to English speaking populations, which was our primary interest. Future research should evaluate the cross-cultural validity of other language translations before application of these tools to target population.

Assessing HRQL offers the opportunity to improve physician-patient communication and achieve better outcomes [59, 60]. Given the multiple demands put on the health care system and the time constraints faced by health care providers, individualized measures that are short, straightforward and quick to administer may help integrate routine HRQL assessment in clinical settings. MYMOP and its adaptations offer a set of brief and easy-to-complete questionnaires that can be used to measure variation in patient-concerns regardless of their diagnosis. MYMOP has been criticized for being symptom specific [61, 62]; however, the recent development by patient-reported outcomes information system (PROMIS) encourages the use of domain-specific rather than disease-specific measures [63]. Researchers at PROMIS state that the experience of fatigue, headache, nausea, sleep problems, and etc. are less likely to be influenced by the mere presence or absence of a disease. MYMOP was developed primarily to overcome the diagnostic differences in different disciplines of health care in a primary care setting. MYMOP (and its adaptations) being generic domain (patient selected)-specific measure can be used to overcome issue of variability in outcome measurement in clinical trials.

As seen in this review, MYMOP and its adaptations have been widely used in the evaluation of complementary therapies because of their excellent fit with individualized patient-centered approach. Given the global initiatives advocating patient-centered research and outcomes [64–67], and a better understanding of limited application of evidence from group data of clinical trials to individual patients [11]; MYMOP and its adaptations can help provide rigorous data from patient perspective. While there are sophisticated methods to deal with heterogeneity of treatment effect [10], because they are often unavoidable and may not be necessarily seen as ‘undesirable’ there is a need to have robust generic individualized outcomes measures such as MYMOP and its adaptations. Therefore individualized outcome assessment tools such as MYMOP is the way forward to personalized medicine approaches to tailor conventional therapies from patient perspective.

Conclusion

MYMOP and its adaptations can be a starting point for *domain-specific* measurement of symptoms like pain, nausea, anxiety, etc. Given that validation is an iterative/ongoing process and considerable efforts have been put to develop and achieve sound psychometrics of these measures, we would recommend researchers to further the

validation of MYMOP and its adaptations before considering developing a new measure. We recommend future studies on construct validity and responsiveness include well defined a priori hypotheses with direction and magnitude of expected correlations [23, 24, 50], and thoughtful consideration of external anchors against which the MYMOP measures are validated. Also to improve consistency, modern day recommended taxonomy should be used to define instrument measurement properties [24].

Acknowledgements The authors are grateful to Charlotte Paterson for her helpful comments. During the time of this work, Jeffrey Johnson was a Senior Scholar with Alberta Innovates- Health Solutions and a Centennial Professor at the University of Alberta. During the time of this work, Sunita Vohra was a Health Scholar with Alberta Innovates-Health Solutions and a Killam Professor at the University of Alberta.

Compliance with ethical standards

Conflict of interest The authors do not have any conflict of interest.

Research involving human participants or animals This review does not contain any studies with human participants performed by any of the authors.

Informed consent Informed consent was not applicable to this review as no primary data were collected.

Appendix

See Table 4.

References

1. Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*, 129–136.
2. Mayo, E. N., et al. ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement. 2015; First Edition.
3. Sales, C. M. D. (2017). Seeing the person in the patient: Making the case for individualized proms in mental healthcare. *Current Psychiatry Reviews, 13*(3), 184–187.
4. Sales, C. M. D., & Alves, P. C. G. (2016). Patient-centred assessment in psychotherapy: A Review of individualized tools. *Clinical Psychology: Science and Practice, 23*(3), 265–283.
5. Wiering, B., de Boer, D., & Delnoij, D. (2017). Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expectations, 20*(1), 11–23.
6. Bowling, A. (1995). What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Social Science & Medicine, 41*, 1447–1462.
7. Carr, A. J., Gibson, B., & Robinson, P. G. (2001). Measuring quality of life is quality of life determined by expectations or experience? *British Medical Journal, 322*, 1240–1243.
8. Ostenfeld-Rosenthal, A., & Johannessen, H. (2014). How to capture patients' concerns and related changes: Comparing the MYCaW questionnaire, semi-structured interview and a priority list of outcome areas. *Complementary Therapies in Medicine, 22*, 690–700.
9. Ernst, E., & Hung, S. K. (2011). Great expectations: What do patients using complementary and alternative medicine hope for? *Patient, 4*, 89–101.
10. Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W., & Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology, 66*, 818–825.
11. Dahabreh, I. J., Hayward, R., & Kent, D. M. (2016). Using group data to treat individuals: Understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *International Journal of Clinical Epidemiology, 45*, 2184–2193.
12. Aburub, A. S., & Mayo, N. E. (2017). A review of the application, feasibiliyt, and the psychometric properties of the individualized measures in cancer. *Quality of Life Research, 26*, 1091–1104.
13. Wettergren, L., Kettis-Lindblad, Å, Sprangers, M., et al. (2009). The use, feasibility and psychometric properties of an individualized quality-of-life instrument: a systematic review of the SEIQoL-DW. *Quality of Life Research, 18*, 737–746.
14. Hickey, A. M., Bury, G., O'Boyle, C. A., Bradley, F., O'Kelly, F. D., & Shannon, W. (1996). A new short form individual quality of life measure (SEIQoL-DW): Application in a cohort of individuals with HIV/AIDS. *BMJ, 313*, 29–33.
15. Moons, P., Marquet, K., Budts, W., & De Geest, S. (2004). Validity, reliability and responsiveness of the "schedule for the evaluation of individual quality of life-direct weighting" (SEIQoL-DW) in congenital heart disease. *Health and Quality of Life Outcomes, 2*, 27.
16. Lhussier, M., Watson, B., Reed, J., & Clarke, C. L. (2005). The SEIQoL and functional status: How do they relate? *Scandinavian Journal of Caring Sciences, 19*, 403–409.
17. Martin, F., Camfield, L., Rodham, K., et al. (2007). Twelve uears-experience with the Patient Generated Index (PGI) of quality of life: A graded structured review. *Quality of Life Research, 16*, 705–715.
18. Mayo, N. E., Aburub, A., Brouillette, M. J., et al. (2017). In support of an individualized approach to assessing quality of life: comparison between Patient Generated Index and standardized measures across four health conditions. *Quality of Life Research, 26*, 601–609.
19. Ruta, D. A., Garratt, A. M., Leng, M., Russell, I. T., & MacDonald, L. M. (1994). A new approach to the measurement of quality of life. The patient-generated index. *Medical Care, 32*, 1109–1126.
20. Paterson, C. (1996). Measuring outcomes in primary care: A patient generated measure, MYMOP, compared with the SF-36 health survey. *BMJ, 312*, 1016–1020.
21. Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., et al. (2018). COSMIN guideline for systematic reviews of patient reported outcome measures. *Quality of Life Research., 27*, 1147.
22. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34–42.
23. Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., et al. (2018). COSMIN Risk of Bias checklist for systematic review of patient-reported outcome measures. *Quality Life Research., 27*, 1171. https://doi.org/10.1007/s11136-017-1765-4.
24. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (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*, 737–745.
25. Quality of life and pharmaco-economics in clinical trials (Book Review) R. Launois. Quality of Life Research, 1 January 1997, Vol. 6(1), pp. 97-98 [Peer Reviewed Journal]
26. Chen, A. Y., Frankowski, R., Bishop-Leone, J., Hebert, T., Leyk, S., Lewin, J., & Goepfert, H. (2001). The development and validation of a dysphagia-specific quality-of-life questionnaire for patients with head and neck cancer: The M. D. anderson dysphagia inventory. *Archives of Otolaryngology–Head & Neck Surgery, 127*, 870–876.
27. Guyatt, G. H., Kirshner, B., & Jaeschke, R. (1992). Measuring health status: What are the necessary measurement properties? *Journal of Clinical Epidemiology, 45*, 1341–1345.
28. Ganz, P. A., Lee, J. J., & Siau, J. (1991). Quality of life assessment. an independent prognostic variable for survival in lung cancer. *Cancer, 67*, 3131–3135.
29. Idler, E. L., & Angel, R. J. (1990). Self-rated health and mortality in the NHANES-I epidemiologic follow-up study. *American Journal of Public Health, 80*, 446–452.
30. Ashworth, M., Shepherd, M., Christey, J., Matthews, V., Wright, K., Parmentier, H., Robinson, S., & Godfrey, E. (2004). A client-centred psychometric instrument: The development of "PSYCHLOPS"("psychological outcome profile"). *Counselling and Psychotherapy Research, 4*, 27.
31. Ashworth, M., Evans, C., & Clement, S. (2009). Measuring psychological outcomes after cognitive behaviour therapy in primary care: A comparison between a new patient-generated measure "pSYCHLOPS" psychological outcome profiles and "hADS" hospital anxiety and depression scale. *Journal of Mental Health, 18*, 169–177.
32. Ashworth, M., Robinson, S., Evans, C., Shepherd, M., Conolly, A., & Rowlands, G. (2007). What does an idiographic measure (PSYCHLOPS) tell us about the spectrum of psychological issues and scores on a nomothetic measure (CORE-OM)? *Primary Care and Community Psychiatry, 12*, 7–12.
33. Ashworth, M., Robinson, S., Godfrey, E., Parmentier, H., Shepherd, M., Christey, J., Wright, K., & Matthews, V. (2005). The experiences of therapists using a new client-centred psychometric instrument, "PSYCHLOPS" ("psychological outcome profile"). *Counselling and Psychotherapy Research, 5*, 37–41.
34. Ashworth, M., Robinson, S. I., Godfrey, E., Shepherd, M., Evans, C., Seed, P., Parmentier, H., & Tylee, A. (2005). Measuring mental health outcomes in primary care: The psychometric properties of a new patient-generated outcome measure, 'PSYCHLOPS' ('psychological outcome profiles'). *Primary Care Mental Health, 3*, 261–270.
35. Day, A. (2004). The development of the MYMOP pictorial version. *Acupuncture in Medicine, 22*, 68–71.
36. Day, A., & Kingsbury-Smith, R. (2004). An audit of acupuncture in general practice. *Acupuncture in Medicine, 22*, 87–92.
37. Evans, C., Ashworth, M., & Peters, M. (2010). Are problems prevalent and stable in non-clinical populations? problems and test-retest stability of a patient-generated measure, PSYCHLOPS (psychological outcome profiles), in a non-clinical student sample. *British Journal of Guidance & Counselling, 38*, 431–439.
38. Cooke, H. J. (2000). Is the adapted "measure yourself medical outcome profile questionnaire an appropriate tool to evaluate the bristol cancer help centre's supportive programme? Dissertation. University of Exeter
39. Paterson, C., & Britten, N. (2000). In pursuit of patient-centred outcomes: A qualitative evaluation of the 'measure yourself medical outcome profile'. *Journal of Health Services Research and Policy, 5*, 27–36.
40. Paterson, C. (2004). Seeking the patient's perspective: A qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Quality of Life Research, 13*, 871–881.
41. Paterson, C., Langan, C. E., McKaig, G. A., Anderson, P. M., Maclaine, G. D. H., Rose, L. B., Walker, S. J., & Campbell, M. J. (2000). Assessing patient outcomes in acute exacerbations of chronic bronchitis: The measure your medical outcome profile (MYMOP), medical outcomes study 6-item general health survey (MOS-6A) and EuroQol (EQ-5D). *Quality of Life Research, 9*, 521–527.
42. Paterson, C., Thomas, K., Manasse, A., Cooke, H., & Peace, G. (2007). Measure yourself concerns and wellbeing (MYCaW): An individualised questionnaire for evaluating outcome in cancer support care that includes complementary therapies. *Complementary Therapies in Medicine, 15*, 38–45.
43. Polley, M. J., Seers, H. E., Cooke, H. J., Hoffman, C., & Paterson, C. (2007). How to summarise and report written qualitative data from patients: A method for use in cancer support care. *Supportive Care Cancer, 15*, 963–971.
44. MYMOP – welcome. Retrieved February 12, 2011, from http://sites.pcmd.ac.uk/mymop/index.php?c=welcome .
45. Peace, G., & Manasse, A. (2002). The cavendish centre for integrated cancer care: Assessment of patients' needs and responses. *Complementary Therapies in Medicine, 10*, 33–41.
46. MYMOP—Sister. (2011). Retrieved February 12, 2011, from http://sites.pcmd.ac.uk/mymop/index.php?c=sister&s=mycaw .
47. Jolliffe, R., Seers, H., Jackson, S., Caro, E., Weeks, L., & Polley, M. J. (2015). The responsiveness, content validity, and convergent validity of the measure yourself concerns and wellbeing (MYCaW) patient-reported outcome measure. *Integrative Cancer Therapies, 14*, 26–34.
48. Seers, H. E., Gale, N., Paterson, C., Cooke, H. J., Tuffrey, V., & Polley, M. J. (2009). Individualised and complex experiences of integrative cancer support care: Combining qualitative and quantitative data. *Supportive Care Cancer, 17*, 1159–1167.
49. Psychlops|Home. (2011). Retrieved February 12, 2011, from http://www.psychlops.org.uk/.
50. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international delphi study. *Quality of Life Research, 19*, 539–549.
51. Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., & De Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research, 21*, 651–657.
52. Lohr, K.N. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research, 11*(3), 193–205.
53. Valderas, J. M.. Ferrer, M.. Mendivil, J.. Garin, O.. Rajmil, L., Herdman, M.. Alonso, J., Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network. (2008). Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value in Health, 11*(4), 700–708.
54. Higgins, J. P., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, 5928.
55. Juni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA, 282*(11), 1054–1060.
56. Greenland, S., & O'Rourke, K. (2001). On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics, 2*(4), 463–471.

57. Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials, 17*(1), 1–12.
58. Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2006). Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology, 59*(7), 697–703.
59. Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., & Selby, P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology, 22*, 714–724.
60. de Wit, M., Delemarre-van de Waal, H. A., Bokma, J. A., Haasnoot, K., Houdijk, M. C., Gemke, R. J., & Snoek, F. J. (2008). Monitoring and discussing health-related quality of life in adolescents with type 1 diabetes improve psychosocial well-being: A randomized controlled trial. *Diabetes Care, 31*, 1521–1526.
61. Brewin, T., & Garrow, J. (1995). Commissioning complementary medicine. Evaluations of efficacy of treatments should be consistent. *BMJ, 311*, 809.
62. Fisher, P., & Eden, A. (1995). Commissioning complementary medicine. Homoeopathic hospitals have unique skill. *BMJ, 311*, 809.
63. PROMIS. Retrieved July 19, 2012, from http://www.nihpromis.org/?AspxAutoDetectCookieSupport=1.
64. National Health Services. Retrieved September 4, 2017 http://www.nhs.uk/Search/Pages/Results.aspx?___JSSniffer=true&q=patient+reported+outcomes.
65. Strategy for Patient-Oriented Research – CIHR. Retrieved September 4, 2017, from http://www.cihr-irsc.gc.ca/e/41204.html.
66. Patient-Centered Outcomes Research Institute. Retrieved September 4, 2017, from https://www.pcori.org/.
67. International Society for Pharmacoeconomics and Outcomes Research. Retrieved September 4, 2017, from https://www.ispor.org/about-ispor.asp.