Title: Systematic review of psychometric properties of instruments measuring patient preferences for engagement in health care

Authors: Teresa Jerofke-Owen PhD, RN, Mauricio Garnier-Villarreal PhD, Alissa Fial BA, MA, MLIS, Georgia Tobiano PhD, BN

Aim: To identify, critically appraise, and summarize instruments that measure patients' preferences for engagement in healthcare.

Design: Psychometric systematic literature review.

Data sources: PubMed, Embase, CINAHL, and PsycINFO were searched from inception to March 2019.

Review methods: Three reviewers independently evaluated the 'methodological quality' and the 'measurement properties' of the included instruments using The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist and Terwee's quality criteria. Each instrument was given a Grading of Recommendations Assessment, Development, and Evaluation (GRADE) score. The review was registered at PROSPERO (registry number CRD42018109253).

Results: A total of 16 studies evaluating 8 instruments measuring patients' preferences for engagement in healthcare were included. All instruments were downgraded for their 'methodological quality' or 'measurement properties', or a combination of both. Common concerns were lack of theoretical basis, absence of patient input during development, incorrect usage and reporting of validity measures, and absence of a priori hypotheses to test validity.

Conclusions: There were no identified instruments that demonstrated adequate evidence for all measurement properties. The Patient Preferences for Patient Participation Scale (4Ps) and 10-item Decisional Engagement Scale (DES-10) had the highest overall GRADE scores; however, each had some underlying developmental or methodological issues.

Impact: Assessing how patients prefer to engage in their care is a critical first step to truly individualize engagement interventions to meet patient expectations. Systematic reviews of measures of patient experience with engagement in healthcare have been undertaken but none are available on measures of patient preferences for engagement. The results highlight the need to further develop and test instruments that measure patients' preferences for engagement in healthcare within a framework for consumerism. Involving the consumer in the instrument development process will ensure that engagement strategies utilized by healthcare providers are relevant and individualized to consumer preferences.

Keywords: adult population, instruments, literature review, nursing, patient participation, patient preference, psychometric properties, systematic review

Introduction

Recent patient-focused models of healthcare emphasize the importance of patient engagement for the delivery of safe and effective healthcare. There are multiple definitions of patient engagement, with most focusing on involving patients and/or their families in individualized care, while working as a team, to obtain the greatest benefits from the healthcare system (Carman et al., 2013; Gruman et al., 2010; Higgins, Larson, & Schnall, 2017). Patients and their family may engage in their care by gathering information on topics such as available providers, treatments, and diagnoses, participating in decision-making, planning and setting goals, collaborating and communicating with the care team, providing feedback about the care received, participating in the development and conduct of research or quality improvement efforts, and utilizing various aspects of one's electronic health record (Agency for Healthcare Research and Quality, 2018; Drenkard, Swartwout, Deyo, & O'Neill, 2015; Sloan & Knowles, 2017). In Kolovos, Kaitelidou, Lemonidou, Sachlas, and Sourtzi's (2016) study, patients described their participation as communicating their preferences for care with the nursing staff and Sahlsten, Larsson, Sjostrom, and Plos (2008) emphasized the importance of adjusting information and knowledge sharing to patients' needs, highlighting the patient-centeredness of engagement. Consumerism, or involving patients more actively in their care, currently underpins many contemporary perspectives on improving healthcare quality (Carman, Lawrence, & Siegel, 2019).

Background

Patient engagement initiatives have been associated with patient outcomes such as improved mental and physical health, better post-discharge health-related quality of life, clinical safety, participation in self-managing behaviors, decreased healthcare utilization, and subsequent

lower spending (Black, Varaganum, & Hutchings 2014; Doyle, Lennox & Bell, 2013; Duke, Lynch, Smith, & Winstanley, 2015). However, patients often are not engaged in their care at their preferred level (Jerofke-Owen & Dahlman, 2019; Rozenblum et al., 2011). Consequently, when patients feel that their care needs are unmet, due to not being heard by nurses or perceiving that nurses do not care about them, patients can become disengaged in their care (Latimer, Chaboyer, & Gillespie, 2014).

Not all patients will have the assertiveness necessary to share their engagement preferences with their providers without being asked, placing value on the necessity of a valid and reliable measure to assess patient preferences for engagement in care (Etkind, Bone, Lovell, Higginson, & Murtagh, 2018). Nursing staff should not assume they know what patients' needs or values are without asking patients, as very different needs can be identified by patients themselves, when compared to nurses' assumptions (Florin, Ehrenberg & Ehnfors, 2005). Incorporating patients' preferences for engagement provides a way to help prioritize what is important to the patient (Mangin, Stephen, Bismath, & Risdon, 2016) and leads to increased adherence to the careplan (Turner-Stokes, Rose, Ashford, & Singer, 2015) and improved satisfaction levels (Suhonen, Valimaki, & Leino-Kilpi, 2005).

Prior systematic reviews (Mavis et al., 2014; Phillips, Street, & Haesler, 2016) have focused on instruments that measure the patient experience of engagement or the success of the engagement process (Bolvin et al., 2018; Graffigna & Barelo, 2018), rather than on instruments that measured patient preferences for engagement. The identification of a valid and reliable instrument that can be used to assess patient preferences for engagement in healthcare is necessary so that healthcare providers can tailor engagement interventions to patient desire rather than assumptions.

The Review

Aims

The aim of this systematic review is to identify, critically appraise, and summarize instruments that measure patients' preferences for engagement in healthcare.

Design

A psychometric systematic review was conducted using guidelines for measurement property evaluation from the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist. The authors of this systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, & Altman, 2009) for reporting. The authors received no funding for this work.

Search Methods

The following electronic databases were searched from their inception up to March 30, 2019: PubMed, Cumulative Index to Nursing and Allied Health Professions (CINAHL), PsycINFO, and Embase. The search strategies were developed in PubMed by two health librarians, using a combination of Medical Subject Terms (MeSH) and keywords, and then modified to fit the parameters of the other databases. Hand searching was also conducted using the reference lists of studies that were included in the review. Articles were limited to English language. The complete search strategy can be found in Supplementary File 1. The review was registered at PROSPERO (registry number CRD42018109253).

Eligibility criteria.

Inclusion criteria were:

- Type of participants: adults (aged 19 years and over)

- Type of outcomes: self-reported patient preferences for engagement in healthcare
- Type of studies: studies that had described the development of instruments that measure patients' preferences for engagement in healthcare and evaluated at least one psychometric property from the COSMIN checklist
- Type of instrument: self-reported quantitative instrument

Exclusion criteria were:

- Instrument not available in English language (primary language of authors) even though article written in English language
- Instrument disease-specific or setting-specific (i.e. rehabilitation units, mental health units) and not able to edit a word or two in items to make it usable across a larger sample
- Grey literature

Search Outcome

References for 3,478 articles were exported to EndNote (Clarivate Analytics, 2019) and 638 duplicates were removed. The detailed selection process can be found in Figure 1. Abstracts were screened and examined independently by the researchers (██████████). Full-texts of articles were examined if abstracts were deemed suitable by at least one author. Differences in selections were discussed and consensus was reached without the introduction of a third reviewer.

Quality Appraisal

Assessment of the Methodological Quality

COSMIN checklists (Mokkink et al., 2018; Terwee et al., 2017) were used to evaluate the methodological quality of the following measurement properties for instruments in each single study: content validity, structural validity, internal consistency, reliability, hypotheses testing for

construct validity, measurement error, cross-cultural validity/measurement invariance, criterion validity, and responsiveness. Each subscale of a multidimensional instrument was evaluated separately. The COSMIN rating scale of ‘very good’, ‘adequate’, ‘doubtful’ or ‘inadequate’ was used to rate each measurement property. The overall score for each measurement property was obtained by considering the lowest rating given to any item within the given measurement property evaluation box. Three researchers ( ) independently rated the studies, ratings were compared, and differences were resolved by discussion.

Assessment of the Measurement Properties

The extracted data on measurement properties for each study were rated against the ‘updated criteria for good measurement properties’ using Terwee’s quality criteria (Prinsen et al., 2018; Terwee et al., 2007). Extracted measurement property data were evaluated as positive (+), negative (-), or indeterminate (?) based on how they related to the quality criteria for each of the nine measurement properties, with the exception of content validity, for which there is no criteria. Content validity is rated based on relevance of items, comprehensiveness of items, and comprehensibility rather than on statistical analyses (Terwee et al., 2017) and therefore is given a methodology quality rating only.

Data Abstraction

The following data were extracted independently by the researchers ( ) from the included studies: name of instrument, language version, author(s), year, country, study aim(s), dimensions and number of items, response options, delivery method, total sample size, sample characteristics, setting, and measurement properties. Any discrepancy in data was resolved through discussion. For example, on occasion a measurement property was missed by one researcher and had to be confirmed by going back to the full-text articles.

Synthesis

After the methodological quality was determined for each of the single studies and the measurement properties were evaluated, the results were qualitatively summarized to determine a final modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) rating for the quality of evidence for each instrument (Prinsen et al., 2018). In cases where instruments were tested in multiple studies and results were not consistent, the conclusion was based on the majority of consistent results and a downgrade was applied for inconsistency (Mokkink et al., 2018). The modified GRADE approach examines: risk of bias (methodological quality), inconsistency, imprecision (total sample size), and indirectness of instruments across studies to determine quality of evidence. Instruments were scored as high, moderate, low, or very low.

Results

Identified Instruments

Sixteen single studies were identified that tested the following eight instruments: Patient Preferences for Patient Participation Scale (4Ps), Autonomy Preference Index (API), Control Preferences Scale (CPS), 10-item Decisional Engagement Scale (DES-10), Desire to Participate in Medical Decision Making (DPMDM), Healthcare Empowerment Questionnaire (HCEQ), Communication Preferences of Patients with Chronic Illness (KOPRA), and Patient Attitudes & Beliefs Scale (PABS) as listed in Figure 1.

Description of the Instruments and Studies

Table 1 presents the characteristics of the eight instruments for measuring patients' preferences for engagement in healthcare. Sample sizes varied from 65 to 1592. The earliest instrument, the API, originated in 1989, and the most recent, the DES-10, in 2016. Four

instruments were tested in more than one language (De las Cuevas & Penate, 2016; Giordano et al., 2008; Mohebbi et al., 2018; Simon et al., 2010). Eight of the studies were conducted in Europe, seven in the United States, two in Canada, and one in Asia. All studies but one, which was conducted in a medical-surgical inpatient setting (Simon et al., 2010), were conducted in ambulatory, clinic-type settings. Some instruments were solely tested using distinct samples: HCEQ in persons older than 75 years (Gagnon, Hebert, Dube & Dubois, 2006), DPMDM and DES-10 in patients with diabetes (Golin, DiMatteo, Leake, Duan, & Gelberg, 2001; Hoerger, Chapman, Mohile, & Duberstein, 2016), 4Ps in patients with chronic obstructive pulmonary disease (COPD) or congestive heart failure (Luhr, Eldh, Nilsson, & Holmefur, 2018), and the KOPRA in patients with chronic back pain or ischemic heart disease (Farin, Gramm & Kosiol, 2011).

The DES-10 and the CPS were the only instruments that were not multi-dimensional. All instruments, except the CPS (Degner & Sloan, 1992), utilized a Likert scale for item responses ranging from 4-10 points. The CPS used a card sort technique with five cards that is methodologically based on unfolding theory. The number of items in the instruments ranged from 10 to 32.

Methodological Quality and the Measurement Properties Results

The methodological quality and measurement properties results are presented in Table 2. No instruments included in this review were evaluated for measurement error or responsiveness by their authors, so the authors did not assess those measurement properties. Criterion validity was not evaluated for the instruments, as there is no current gold standard instrument for assessing patient preferences for engagement in care (Mokkink et al., 2018). Cross-cultural

validity was not assessed for the four studies that tested a different language version of an instrument, as those study teams did not conduct multiple group factor analysis.

Content validity. The construct being measured was clearly described for all eight instruments, but only the development of the CPS (Degner & Sloan, 1992), DPMDM (Golin et al., 2001), and the PABS (Arora, Ayanian, & Guadagnoli, 2005) were based on theory. Items for the 4Ps, DPMDM, KOPRA, and PABS were developed with input from the target population (Arora et al., 2005; Farin et al., 2011; Golin et al., 2001; Luhr et al., 2018), strengthening their content validity, whereas the API, CPS, DPMDM, and DES-10 were constructed from input from healthcare professionals only (Degner & Sloan, 1992; Ende, Kazis, Ash, & Moskowitz, 1989; Golin et al., 2001; Hoerger et al., 2016). Cognitive interviews were conducted with patients of the target populations to determine the relevance and comprehensibility of the items for the 4Ps and KOPRA (Farin et al., 2011; Luhr et al., 2018), strengthening the methodological quality for those two instruments. The evaluation of the CPS, DES-10, and the PABS (Arora et al., 2005; Degner & Sloan, 1992; Hoerger et al., 2016) was limited because no content validity testing was reported. Overall, the 4Ps, DPMDM, and KOPRA (Farin et al., 2011; Golin et al., 2001; Luhr et al., 2018) were rated the highest in content validity quality, as their study teams were the most thorough at developing the construct within the target population and pilot testing the instrument before use. None of the studies qualified for a ‘very good’ methodological quality rating because it was not assumable that proper measures to ensure comprehensiveness and comprehensibility of the instruments were undertaken.

Structural validity. The structure of the 4Ps was tested (Luhr et al., 2018) using Rasch modeling. Less than 200 subjects were used for the Rasch model, thus the methodological

quality was downgraded to ‘adequate’. The quality of the measurement property was (-) because one Z-standardized infit value was not > -2 .

Most of the authors used classical test theory (CTT) to test structural validity, however, a factor structure could not be tested for the CPS because it is scored using a card sort method. For the API, Ende et al. (1989) did not provide fit indices, Simon et al. (2010) demonstrated an adequate fit for the German version, and Bonfils, Adams, Mueser, Wright-Berryman and Salyers (2015) and Morandi et al. (2017) demonstrated some questionable fit indices.

Many instruments were scored ‘inadequate’ due to the methods used for testing structural validity. For example, Hoerger et al. (2016) used confirmatory factor analysis (CFA) to test the theoretical factor structure of the DES-10, but only tested for configural invariance between groups and did not complete a multiple group comparison between groups in CFA. Golin et al. (2001) intended to do exploratory factor analysis (EFA), but used Principal Components Analysis (PCA) in a small sample size. For the HCEQ, Gagnon et al. (2006), Mohebbi et al. (2018), and Arora et al. (2005) used PCA but interpreted their analysis as if it resulted from EFA (Park, Dailey, & Lemus, 2002).

The KOPRA (Farin et al., 2011) was scored as ‘inadequate’ due to lack of clarity of EFA or PCA was used, inconsistencies in treating data as both continuous and ordered categorical in analyses, poor fit indices, and an inadequate sample size ($n=333$) given the large number of initial items tested ($n=105$).

Internal consistency. The majority of the scales were tested for internal consistency and scored ‘very good’ for methodological quality. All reported Cronbach alpha estimates were ≥ 0.7 , which is acceptable for new instruments, except for the Persian version of the HCEQ (Mohebbi et al., 2018). Methodological quality was ‘very good’ for the API, except for in one

study (Sung, Raker, Myers, & Clark, 2010), when only one Cronbach alpha score was provided for a multi-dimensional instrument.

Reliability. While reliability measures were provided in eight of the articles, the majority of authors did not provide details regarding the stability of the patients following the initial measure, similarity of test conditions, or justification for the time interval ($n=5$) and were therefore graded as ‘doubtful’. ICC values were given for the 4Ps, API (Morandi et al., 2017), and CPS (Sung et al., 2010), and weighted Kappa for the CPS (Giordano et al. 2008), however they were not of acceptable values. Ende et al. (1989) and Golin et al. (2001) used Pearson r correlation coefficient to measure reliability and did not provide ICC or weighted Kappa measures.

Hypothesis testing. All studies were rated as ‘very good’ for methodological quality because appropriate statistical methods were used and important characteristics of the subgroups were described. Only one study team stated a priori hypotheses (Golin et al., 2001) relating to the relationship of DPMDM scores with other measures of health value and social support, age, attendance of a diabetes education class, and measures of autonomy. Rather than using structural equation modeling where multiple relationships can be examined at once, eleven different analyses were conducted, introducing Type 1 error into the results. Arora et al. (2005) examined if patients’ scores on the PABS varied with stage of readiness within the Transtheoretical Model of Health Behavior Change (theoretical framework of the instrument) and found that they did, adding strength to the instrument. Other study teams performed numerous univariate and regressions analyses to see whether patient demographic variables or illness factors influenced preferences.

Qualitative summary of evidence and overall grade of evidence quality. The final modified GRADE scores can be found in Table 3. The strength of the 4Ps instrument was its development, which was based on extensive patient input and tested in the intended population; however, development of items was not based on theory. Weaknesses included a small sample size and unfavorable fit indices and reliability measures. It was given an overall rating of ‘Moderate’.

Medical professionals developed items on the API and the instrument was tested in five studies. Methodology for hypothesis testing was ‘very good’ for the API; however, findings were inconsistent. There were also inconsistencies in the factor structure of the instrument across the five studies and the use of the clinical vignettes. Sung et al. (2010) provided one Cronbach alpha internal consistency estimate, which was inconsistent with the two-factor structure of the instrument. Lastly, intraclass correlation coefficients did not reach significant values (≥ 0.70). The API was given an overall rating of ‘low’ because of the inconsistencies in the structure, the inconsistencies in the reported Cronbach alpha values, and the inadequate reliability measures.

The scoring procedures used in the five separate studies that tested the Control Preferences Scale were inconsistent. The CPS was intended to be scored using unfolding theory (Degner & Sloan, 1992), in which patients compare two cards together until the entire preference order across the set of five cards is unfolded. Giordano et al. (2008) and De las Cuevas and Penate (2016) both used a different, simultaneous administration method and Sung et al. (2010) had patients select one card when presented with the five choices, essentially using a 5-point Likert scale. The methodological quality of hypothesis testing was ‘very good’ for all five studies; however, findings were inconsistent. The CPS does not have a structure to test nor does

it have individual items that can be used to determine Cronbach's alpha estimates (although one study team did turn the cards into five separate items and calculated internal consistency). Due to the unacceptable measurement properties and the inconsistencies in scoring, the CPS was given an overall rating of 'low'.

The development of the DES-10 was based on professional input; however, no content validity or pilot testing was conducted with the initial version of the instrument. Hoerger et al (2016) demonstrated acceptable fit indices and internal consistency for the DES-10 and there was strong evidence for hypothesis testing. Therefore, the DES-10 was given an overall rating of 'moderate'. Strengths of the DPMDM included content validity and demonstration of internal consistency. The sample size was small ($n=65$) and hypothesis testing showed inconsistencies and questionable methodology. Therefore, the DPMDM was given an overall rating of 'low'. The Healthcare Empowerment Questionnaire was tested in English (Gagnon et al., 2006) and in Persian (Mohebbi et al., 2018); however, the only measurement property that was reported for both versions was Cronbach's alpha internal consistency estimate. Cronbach's alpha for the "involvement in decisions" dimension in the Persian version was $< .7$, which is not ideal for a new instrument and ICC values for subscales were primarily $< .7$. The structure of the instrument was incorrectly tested in both instances and no hypothesis testing was conducted. The HCEQ was given a quality level rating of 'low'.

Strengths of the KOPRA instrument included integration of patient feedback during the development of the instrument and demonstration of internal consistency. Weaknesses included the fit indices, small sample size, and absence of reliability and hypothesis testing. The KOPRA was given an overall rating of 'low', as little quality evidence supported its use in practice. Strengths of the PABS included item development based on the Transtheoretical Model of

Health Behavior Change, acceptable fit indices following CFA, and evidence of internal consistency. Patients' responses on the PABS did correlate appropriately with their stage of readiness within the Transtheoretical Model of Health Behavior Change (Arora et al., 2005). Weaknesses included absence of thorough content validity testing and treating of PCA as EFA. The PABS was given an overall rating of 'low'.

Discussion

To the best of our knowledge, this is the first systematic review of the psychometric properties of instruments that measure patients' preferences for engagement in healthcare. This systematic review identified, critically appraised, and summarized 16 studies that estimated the psychometric properties of eight instruments. Four of the studies were conducted to validate three of the instruments in different languages and cultures. Our critical appraisal of the sixteen studies revealed that none of the author teams reported all nine measurement properties that are part of the psychometric analysis using COSMIN methodology. The psychometric properties of the 4Ps, DES-10, DPMDM, KOPRA, and PABS have only been assessed in single studies. Additional evaluations of the psychometric properties of those instruments are necessary before recommending their widespread use. While the API, CPS, and HCEQ were tested in multiple studies, there were inconsistencies in their measurement properties, administration methods, and instrument structures.

Each of the eight instruments had downgrading due to methodological issues and/or insufficient measurement properties. Common methodological issues uncovered were that instruments were not theoretically or conceptually derived or developed with input from patients themselves, incorrect usage and reporting of PCA or EFA, use of Pearson correlation instead of intra-class correlation or kappa measures when reporting reliability, failure to provide detail if

testing conditions and patients were stable in retesting situations, and absence of a priori hypotheses to test convergent validity. The 4Ps was the only instrument not downgraded to doubtful or lower for methodological quality of measurement, yet its reported structural validity and reliability measurement properties were insufficient. However, the appropriateness of measuring reliability for patient preference for engagement instruments is questionable, given that patients' preferences may change over time, and thus should be measured each time a patient interacts with the health care system.

It was alarming that the majority of the reviewed instruments were not theoretically driven and end-users were not involved in their development. Without a sound theoretical or conceptual framework, the structural validity of an instrument or cross-study comparisons of results could be biased (Martinez, Lewis & Weiner, 2015). The meaningful involvement of consumers in research is now best practice worldwide (Manaf, Petermann, Mason-Lai, & Vandall-Walker, 2018), as consumers can co-design instruments to ensure their needs are met, and the final instrument is useable and acceptable. COSMIN criteria place the most weight on the measurement property of content validity, as it is crucial that instruments are comprehensible and relevant to the study population they are intended for (Mokkink et al., 2018). Future research must focus on the development and validation of patient preference instruments using theoretical frameworks that include the concept of consumerism, to ensure that the instruments are measuring all dimensions of patient engagement. Instruments must also involve consumers in the development and initial testing of items to meet current demands for patient-centered landscapes (Forsythe et al., 2019).

Applicability and clarity of items is important when evaluating the relevancy of an instrument. Henrikson, Davison and Berry (2011) conducted an interview study with 20 men

being treated for prostate cancer to examine perceptions of the meaning of the five card choices of the CPS and applicability to their clinical situation. They found confusion among the men about what each of the cards meant and difficulty applying the cards to situations in which a multidisciplinary team was treating the men, as the cards specifically used the term “doctor”. The CPS was not the only instrument in our study to use the term “doctor” or “physician” within the item text. The API used the term “doctor” in 5 items; the only time “nurse” was used in the API was in the high blood pressure vignette for the following item: “how often the nurses should wake you up to check your temperature and blood pressure” (Ende et al., 1989). The scope of nursing practice has advanced beyond taking vital signs and this item reflects a narrow and dated view of the role of the nurse from thirty years ago when the instrument was developed. The DPMDM used the term “doctor” in ten of eleven items. The 4Ps tool was the only tool reviewed that used the terminology “healthcare staff” within items. Given the value placed on delivering care within multidisciplinary teams, future studies should give consideration to using instruments with more inclusive language or develop new instruments that would be more cognizant of a team approach to care.

Two of the instruments, the API and the CPS, were created before the 21st century, when the provider-patient relationship was viewed more paternalistically, where patient participation did not extend beyond decision-making. Interestingly, the API and CPS are still being used in current studies to assess patient preferences as demonstrated by the studies included in this review (Bonfils et al., 2015; De las Cuevas & Penate, 2016; Morandi et al., 2017). The delivery of value-based care became more prolific after the publication of the Institute of Medicine’s (2001) *Crossing the Quality Chasm*. Healthcare systems have been urged to integrate patients into care in new innovative ways (Agency for Healthcare Research and Quality, 2018, Patient-

Centered Outcomes Research Institute, 2014). The KOPRA and 4Ps were two instruments in this study that examined patient preferences for engagement in numerous capacities beyond decision-making including having dialogue with health care staff, sharing knowledge, partaking in planning, managing self-care, encouraging question asking, and treating/getting to know oneself as a person; however there was mixed evidence for methodological quality and properties for the two instruments. Instruments need to include items that reflect a full range of patient engagement activities, so that the full continuum of engagement can be assessed by the provider to better inform tailoring of engagement interventions.

Thirteen of the studies used convenience sampling and tested the instrument in specific samples: The 4Ps was tested with patients who had either a diagnosis of COPD or heart failure (Luhr et al. 2018), the API was tested with patients who had pelvic floor disorders and patients who a significant mental health disorder in three of the five studies (Bonfils et al. 2015; Morandi et al. 2017, Sung et al. 2010), the CPS in oncology patients in two studies (Beaver et al., 1996; Degner & Sloan, 1992) and inpatients with mental health disorders, pelvic floor disorders, and multiple sclerosis (De las Cuevas & Penate, 2016; Giordano et al., 2008; Sung et al., 2010), the DES-10 in oncology patients (Hoerger et al., 2016), the DPMDM was tested with patients who had type 2 diabetes (Golin et al., 2001), the HCEQ was tested in patients older than 75 years and Persian women of reproductive age (Gagnon et al., 2006; Mohebbi et al., 2018), and the KOPRA was tested in patients with chronic back pain or chronic ischemic heart disease (Farin et al., 2011). Only three of the studies used a heterogenous sample of medical patients (Arora et al., 2005; Ende et al., 1989; Simon et al., 2010). Ideally, an instrument measuring patient preference for engagement should be broad enough to be used with patients in any health care setting, who may be experiencing varying healthcare needs. Future studies should be conducted to examine if

the validity of patient engagement instruments are consistent in other populations, providing stronger evidence for the generalizability of the initial findings. As a result of this review, the authors developed and are currently testing a new tool with a heterogeneous sample to measure patient preferences pertaining to a full range of patient engagement activities ([REDACTED] [REDACTED]).

Limitations and Strengths

There were limitations and strengths to our review. This review was limited to studies published in English, therefore instruments developed and evaluated in other languages or cultures may have been overlooked. Although a meticulous search of four databases was conducted to identify instruments, there are other databases and the grey literature that were not searched, and some existing instruments may have been missed. It is possible that relevant articles may have been excluded if there was no mention of evaluation of psychometric properties in the abstract. Two health librarians were actively involved in the search methodology and both authors independently screened articles and conducted handsearching, strengthening the search process.

To the best of our knowledge, this is the first systematic review that identifies, critically appraises, and summarizes the psychometric evidence of instruments that measure patient preferences for engagement in healthcare. This review was conducted using COSMIN guidelines and two reviewers independently screened all of the articles and evaluated all of the included instruments. A statistician was also consulted to assist with evaluating the quality of the methodology and measurement properties of the instruments. We were limited to evaluating measurement properties of the instruments that were published.

Conclusion

This systematic review included eight instruments to measure patient preferences for engagement in their healthcare that were tested in sixteen single studies. All measurement properties were not tested or reported for any one instrument, and the methodological quality and assessment of measurement properties were limited or conflicting. Special care should be given to involving consumers in the instrument development process of future instruments, given our culture is now more expecting and accepting of patient engagement in care and research. The 4Ps and DES-10 instruments had the highest overall GRADE scores; however, each still had some underlying methodological issues, insufficient measurement properties, and was tested in homogenous samples. Cautious use of these instruments is recommended until further high quality studies are conducted. The use of COSMIN guidelines when developing future psychometric studies may strengthen the methodology used and improve the confidence that the true measurement properties are close to the estimates provided by the researchers. Future research must also encompass a broader focus for patient engagement beyond communicative processes such as decision-making and focus on theoretical frameworks for consumerism. Patient preference for engagement instruments should be tested in multiple studies using heterogeneous samples, as the underlying construct of engagement should be applicable to all types of patients regardless of diagnosis or setting. Without a robust way to identify patients' preference for engagement, health care professionals may use ineffective strategies to engage patients in their care.

Conflicts of Interest

No conflict of interest has been declared by the authors.

References

- Agency for Healthcare Research and Quality [AHRQ]. (2018). Engaging patients and families in their health care. Retrieved from <https://www.ahrq.gov/professionals/quality-patient-safety/patient-family-engagement/index.html>
- Arora, N. K., Ayanian, J. Z., & Guadagnoli, E. (2005). Examining the relationship of patients' attitudes and beliefs with their self-reported level of participation in medical decision-making. *Medical Care*, 43(9), 865-872.
- Beaver, K., Luker, K. A., Owens, R.G., Leinster, S., Degner, L., & Sloan, J.A. (1996). Treatment decision making in women newly diagnosed with breast cancer. *Cancer Nursing*, 19(1), 8-19.
- Black, N., Varaganum, M., & Hutchings, A. (2014). Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Quality and Safety*, 23, 534-542. doi: 10.1136/bmjqs-2013-002707
- Bolvin, A., L'Esperance, A., Gauvin, F., Dumez, V., Macaulay, A. C., Lehouxx, P., & Abelson, J. (2018). Patient and public engagement in research and health system decision making: A systematic review of evaluation tools. *Health Expectations*, 21(6), 1075-1084. doi: 10.1111/hex.12804
- Bonfils, K. A., Adams, E. L., Mueser, K. T., Wright-Berryman, J. L., & Salyers, M. P. (2015). Factor structure of the Autonomy Preference Index in people with severe mental illness. *Psychiatry Research*, 228(3), 526-530. doi: 10.1016/j.psychres.2015.06.004
- Carman, K. L., Dardess, P., Maurer, M., Sofaer, S., Adams, K., Bechtel, C., & Sweeney, J. (2013). Patient and family engagement: A framework for understanding the elements and developing interventions and policies. *Health Affairs*, 32(2), 223-231. doi:

10.1377/hlthaff.2012.1133.

Carman, K., Lawrence, W., & Siegel, J. (2019). The 'new' health care consumerism. Retrieved from <https://www.healthaffairs.org/doi/10.1377/hblog20190304.69786/full/>

Clarivate Analytics. (2019). EndNote X8. Retrieved from <https://endnote.com/>

De las Cuevas, C., & Penate, W. (2016). Validity of the Control Preferences Scale in patients with emotional disorders. *Patient Preference and Adherence*, 10, 2351-2356.
doi: 10.2147/PPA.S122377

Degner, L. F., & Sloan, J. A. (1992). Decision making during serious illness: What role do patients really want to play? *Journal of Clinical Epidemiology*, 45(9), 941-950.

Doyle, C., Lennox, L., & Bell, D. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*, 3(1).
doi:10.1136/bmjopen-2012-001570

Drenkard, K., Swartwout, E., Deyo, P., & O'Neil, M.B. (2015). Interactive Care Model: A framework for more fully engaging people in their healthcare. *Journal of Nursing Administration*, 45(10), 503-510. doi: 10.1097/NNA.0000000000000242.

Duke, C. C., Lynch, W. D., Smith, B., & Winstanley, J. (2015). Validity of a new patient engagement measure: The Altarum Consumer Engagement (ACE) Measure. *The Patient*, 8(6), 559-568. doi: 10.1007/s40271-015-0131-2

Ende, J., Kazis, L., Ash, A., & Moskowitz, M.A. (1989). Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. *Journal of General Internal Medicine*, 4(1), 23-30.

Etkind, S., Bone, A. E., Lovell, N., Higginson, I. J., & Murtagh, F. E. M. (2018). Influences on

- care preferences of older people with advanced illness: A systematic review and thematic synthesis. *Journal of the American Geriatric Society*, 66, 1031-1039. doi: 10.1111/jgs.15272.
- Farin, E., Gramm, L., & Kosiol, D. (2011). Development of a questionnaire to assess communication preferences of patients with chronic illness. *Patient Education & Counseling*, 82(1), 81-88. doi:<https://dx.doi.org/10.1016/j.pec.2010.02.011>
- Florin, J., Ehrenberg, A., & Ehnfors, M. (2005). Patients' and nurses' perceptions of nursing problems in an acute care setting. *Journal of Advanced Nursing*, 51(2), 140-149. doi: [10.1111/j.1365-2648.2005.03477.x](https://doi.org/10.1111/j.1365-2648.2005.03477.x)
- Forsythe, L. P., Carman, K. L., Szydowski, V., Fayish, L., Davidson, L., Hickam, D. H., . . . Anyanwu, C. U. (2019). Patient engagement in research: Early findings from the Patient-Centered Outcomes Research Institute. *Health Affairs*, 38(3), 359-367.
- Gagnon, M., Hibert, R., Dube, M., & Dubois, M. F. (2006). Development and validation of an instrument measuring individual empowerment in relation to personal health care: the Health Care Empowerment Questionnaire (HCEQ). *American Journal of Health Promotion*, 20(6), 429-435. doi:10.4278/0890-1171-20.6.429
- Giordano, A., Mattarozzi, K., Pucci, E., Leone, M., Casini, F., Collimedaglia, L., & Solari, A. (2008). Participation in medical decision-making: Attitudes of Italians with multiple sclerosis. *Journal of Neurological Sciences*, 275, 86-91.
- Golin, C. E., DiMatteo, M. R., Leake, B., Duan, N., & Gelberg, L. (2001). A diabetes-specific measure of patient desire to participate in medical decision making. *Diabetes Educator*, 27(6), 875-886.
- Graffigna, G., & Barelo, S. (2018). Spotlight on the Patient Health Engagement model (PHE

- model): a psychosocial theory to understand people's meaningful engagement in their own health care. *Patient Preferences and Adherence*, 12, 1261-1271. doi: 10.2147/PPA.S145646.
- Gruman, J., Rovner, M. H., French, M. E., Jeffress, D., Sofaer, S., Shaller, D., & Prager, D. J. (2010). *Patient Education and Counseling*, 78(3), 350-356. doi: 10.1016/j.pec.2010.02.002
- Henrikson, N. B., Davison, B. J., & Berry, D. L. (2011). Measuring decisional control preferences in men newly diagnosed with prostate cancer. *Journal of Psychosocial Oncology*, 29, 606-618. doi: 10.1080/07347332.2011.615383.
- Higgins, T., Larson, E., & Schnall, R. (2017). Unraveling the meaning of patient engagement: A concept analysis. *Patient Education and Counseling*, 100(1), 30-36. doi: 10.1016/j.pec.2016.09.002.
- Hoerger, M., Chapman, B. P., Mohile, S. G., & Duberstein, P. R. (2016). Development and psychometric evaluation of the Decisional Engagement Scale (DES-10): A patient-reported psychosocial survey for quality cancer care. *Psychological Assessment*, 28(9), 1087-1100. doi:10.1037/pas0000294
- Institute of Medicine. (2001). Crossing the quality chasm. Access from <http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2001/Crossing-the-Quality-Chasm/Quality%20Chasm%202001%20%20report%20brief.pdf>
- Jerofke-Owen, T. A., & Garnier-Villarreal, M. (2020). Development and psychometric analysis of the Patient Preferences for Engagement Tool. *Nursing Research* [publish ahead of print] DOI: 10.1097/NNR.0000000000000423
- Jerofke-Owen, T., & Dahlman, J. (2019). Patients' perspectives on engaging in their

- healthcare while hospitalized. *Journal of Clinical Nursing*, 28(1-2), 340-350.
doi.org/10.1111/jocn.14639
- Kolovos, P., Kaitelidou, D., Lemonidou, C., Sachlas, A., & Sourtzi, P. (2016). Patients' perceptions and preferences of participation in nursing care. *Journal of Research in Nursing*, 21(4), 290-303. doi.org/10.1177/1744987116633498
- Latimer, S., Chaboyer, W., & Gillespie, B. (2014). Patient participation in pressure injury prevention: Giving patient's a voice. *Scandinavian Journal of Caring Sciences*, 28(4), 648-656. doi: 10.1111/scs.12088.
- Luhr, K., Eldh, A. C., Nilsson, U., & Holmefur, M. (2018). Patient preferences for patient participation: Psychometric evaluation of The 4Ps tool in patients with chronic heart or lung disorders. *Nordic Journal of Nursing Research*, 38(2), 68-76. doi:10.1177/2057158517713156
- Manafo, E., Petermann, L., Mason-Lai, P., & Vandall-Walker, V. (2018). Patient engagement in Canada: a scoping review of the 'how' and 'what' of patient engagement in health research. *Health Research Policy and Systems*, 16, 5. doi:10.1186/s12961-018-0282-4
- Mangin, D., Stephen, G., Bismah, V., & Risdon, C. (2016). Making patient values visible in healthcare: A systematic review of tools to assess patient treatment priorities and preferences in the context of multimorbidity. *BMJ Open*, 6(6), doi: 10.1136/bmjopen-2015-010903.
- Martinez, R. G., Lewis, C. C., & Weiner, B. J. (2014). Instrumentation issues in implementation science. *Implementation Science*, 9, 118. doi: 10.1186/s13012-014-0118-8
- Mavis, B., Rovner, M.H., Jorgenson, S., Coffey, J., Anand, N., Bulica, E., . . . Ernst, A. (2015).

- Patient participation in clinical encounters: a systematic review to identify self-report measures. *Health Expectations*, 18(6), 1827-1843. doi: 10.1111/hex.12186
- Mohebbi, B., Tol, A., Shakibazadeh, E., Yaseri, M., Sabouri, M., & Agide, F.D. (2018). Testing psychometric of Health Empowerment Questionnaire (HCEQ) among Iranian reproductive age women: Persian version. *Ethiopian Journal of Health Sciences*, 28(3), 341-346. doi: 10.4314/ejhs.v28i3.12.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement*. *PLoS Med* 6(7). doi:10.1371/journal.pmed1000097
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, 27(5), 1171-1179. doi: 10.1007/s11136-017-1765-4
- Morandi, S., Golay, P., Vazquez-Montes, M., Rugkasa, J., Molodynski, A., Yeeles, K., & Burns, T. (2017). Factorial structure and long-term stability of the Autonomy Preference Index. *Psychological Assessment*, 29(1), 110-115. doi:https://dx.doi.org/10.1037/pas0000327
- Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4), 562–577. doi.org/10.1111/j.1468-2958.2002.tb00824.x
- Patient-Centered Outcomes Research Institute [PCORI] (2014). Advisory panel on patient engagement. Retrieved from <http://www.pcori.org/content/advisory-panel-patient-engagement>
- Phillips, N. M., Street, M., & Haesler, E. (2016). A systematic review of reliable and valid tools

- for the measurement of patient participation in healthcare. *BMJ Quality & Safety*, 25(2), 110-117. doi:10.1136/bmjgs-2015-004357
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27, 1147-1157. doi: 10.1007/s11136-018-1798-3.
- Rozenblum, R., Lisby, M., Hockey, P.M., Levtizion-Korach, O., Salzberg, C.A., Lipsitz, S., & Bates, D.W. (2011). Uncovering the blind sport of patient satisfaction: An international survey. *BMJ Quality and Safety*, 20(11), 959-965. doi: 10.1136/bmjqs-2011-000306.
- Sahlsten, M. J., Larsson, I. E., Sjostrom, B., & Plos, K. A. (2008). An analysis of the concept of patient participation. *Nursing Forum*, 43(1), 2-11. doi: 10.1111/j.1744-6198.2008.00090.x.
- Simon, D., Kriston, L., Loh, A., Spies, C., Scheibler, F., Wills, C., & Härter, M. (2010). Confirmatory factor analysis and recommendations for improvement of the Autonomy-Preference-Index (API). *Health Expectations*, 13(3), 234-243. doi:10.1111/j.1369-7625.2009.00584.x
- Sloan, A. G., & Knowles, A. (2017). Patient engagement: qualitative narratives illustrate patient engagement behaviors. *Journal of Communication in Healthcare*, 10(2), 78-87. <https://doi.org/10.1080/17538068.2017.1340541>
- Suhonen, R., Valimaki, M., & Leino-Kilpi, H. (2005). Individualize care, quality of life and satisfaction with nursing care. *Journal of Advanced Nursing*, 50(3), 283-292.
- Sung, V.W., Raker, C.A., Myers, D.L., & Clark, M.A. (2010). Treatment decision-making and information-seeking preferences in women with pelvis floor disorders. *International*

Urogynecology Journal, 21, 1071-1078. doi: 10.1007/s00192-010-1155-8

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., de Vel, H. C. W., Bouter, L. M., Alonso, J. . . .

Mokkink, L. B. (2018). COSMIN methodology for assessing the content validity of PROMs. User manual version 1.0. Retrieved from <https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf>

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J. . . de

Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42.

Turner-Stokes, L., Rose, H., Ashford, S., & Singer, B. (2015). Patient engagement and

satisfaction with goal planning: Impact on outcome from rehabilitation. *International Journal of Therapy and Rehabilitation*, 22(5), 210-216.

Table 1

Characteristics of study instruments and samples

Instrument (Language)	Author(s) (year) (Country)	Study Aim(s)	Dimensions (number of items)	Response options	Delivery method	Total N	Sample characteristics	Setting
4Ps (Swedish)	Luhr, Eldh, Nilsson, & Holmefur (2018) (Sweden)	To perform a psychometric evaluation of the tool	4 (12) Having dialogue with health care staff (3 items), sharing knowledge (3 items), partaking in planning (3 items), and managing self-care (3 items)	4-point Likert scale	Written and mailed surveys	108	50 patients with COPD and 60 patients with CHF. 55.6% male, mean age 69. Diagnosed with chronic illness for a mean duration of 4 years	Ambulatory care clinics for patients with chronic heart or lung disease

Autonomy Preference Index (English)	Ende, Kazis, Ash, & Moskowitz (1989) (United States)	To develop and test an instrument for measuring patients' preferences for autonomy (desire to make medical decisions and desire to be informed)	2 (23) Information seeking (8 items) and decision making (6 general items and 9 items related to 1 of 3 clinical vignettes representing different levels of illness severity: URI, HTN, and MI)	5-point Likert scale for 14 items 10-point Likert scale for 9 vignette items	Written survey	312	31% were 26-50, 32% were 51-65 and 33% were >65; 62% female, 55% Caucasian, 42% married, 32% had more than high school education, 85% had an income less than \$30,000	Primary care clinic
Autonomy Preference Index (German)	Simon, Kriston, Loh, Spies, Scheibler, Wills, & Harter (2010) (Germany)	Validation of the German version of the API	2 (11) Information seeking (7 items) and decision-making (4 items). Did not test vignettes	5-point Likert scale	Written survey	1,592	Mean age 47.7 (SD=18.4), 51.9% male 11.7% primary care outpatients, 37.4% inpatients and 50.9% emergency department patients	Patients treated for depression in primary care, med/surg inpatient units, and the emergency department

The pelvic floor disorders Autonomy Preference Index (English)	Sung, Raker, Myers, & Clark, (2010) United States	Validate API and CPS in sample of women with pelvic floor disorders. Test-retest 2 weeks later	2(23) Information seeking (8 items) and decision making (6 general items and 9 items related to clinical vignettes adapted to pelvic floor disorders)	5-point Likert scale for 14 items 10-point Likert scale for 9 vignette items	Written survey	110	100% women with pelvic floor disorder. Mean age 62, 97% white, 12.8% did not complete high school	Outpatient clinic

Autonomy Preference Index (English)	Bonfils, Adams, Mueser, Wright-Berryman, Salyers (2015) (United States)	Assess the factor structure of the instrument in a sample of individuals with severe mental illness	2 (11) Information seeking (7 items) and decision making (4 items). Did not test vignettes	5-point Likert scale	Written survey	293	Patients diagnosed with mental health disorder. 54.9% male, 51.9% white, 56.5% living independently. 72.3% had not completed any college and 86.0% were unemployed. Average age 50.2 years. 51.2% had diagnosis of schizophrenia.	Interviewed – outpatient services from mental health or primary care/integrated care services

Autonomy Preference Index (English)	Morandi et al. (2017) (United Kingdom)	Determine best factorial structure of API and examine long-term stability (reassessed at 6 and 12 months)	3 (23) Information seeking (8 items), decision making (6 general items), and decision making (9 items related to 1 of 3 clinical vignettes representing different levels of depression)	5-point Likert scale for 14 items 10-point Likert scale for 9 vignette items	Interview	285	85.3% with a diagnosis of schizophrenia. 68.1% male. Mean age 39.0 (SD=11.4)	Psychiatric hospitalization
Control Preferences Scale (English)	Degner & Sloan (1992) (Canada)	Develop and test a measure to elicit patient preferences regarding participation in health-care decisions	5 cards portraying different roles patients can assume in treatment decision-making.	Uses unfolding theory to select preference order of cards. 11 possible transitive preference orders – translated to ordinal score	Interview	436 oncology patients and 482 members of general public	Oncology patients: Mean age 59, 52% male, 24% greater than high school education) General public: mean age 42, 45% male, 54% education	Ambulatory oncology clinics (newly diagnosed patients)

				ranging from 1 to 11			greater than high school)	
Control Preferences Scale (English)	Beaver et al. (1996) (UK)	Explore the hypothesis that women with breast cancer had specific preferences about the degree of control they wanted over treatment decision making	5 cards portraying different roles patients can assume in treatment decision-making.	Uses unfolding theory. Make 11 preference orders	Interview	150	150 women newly diagnosed with breast cancer and 200 women with benign breast disease; Women with cancer were an average 2.5 weeks from diagnosis, mean age of 54.8 years. Majority were white British, married. 66% planned to have a lumpectomy. Benign group had a mean age of 39/2 years	Breast cancer patients' preferences measured in hospital ward after admission but before surgery and benign breast disease patients had preferences collected in breast clinic at same hospital.
Control Preferences Scale (Italian)	Giordano et al. 2008 (Italy)	To cross-culturally adapt and validate the	5 cards portraying different roles patients can	Simultaneous administration method: choose	Interview	129	Patients with MS. 71% female, mean age 42.	Outpatient clinics

		Italian version of the CPS	assume in treatment decision-making.	preferred card..turn it over, then the next preferred card out of 4, etc. Procedure continues until 1 card is left.			Disease duration ranged 6 months to 39 years. Only scores for 129 were used	
Control Preferences Scale (English)	Sung, Raker, Myers, & Clark, (2010) United States	Validate API and CPS in sample of women with pelvic floor disorders. Test-retest 2 weeks later	1 item - asked patients for 1 answer when presented with the 5 card choices written out	5-point Likert scale	Interview	110	100% women with pelvic floor disorder. Mean age 62, 97% white, 12.8% did not complete high school	Outpatient clinic
Control Preferences Scale (Spanish)	De las Cuevas & Penate, 2016) (Spain)	Test Spanish translation in mental health population.	5 cards portraying different roles patients can assume in treatment decision-making.	Simultaneous administration method: choose preferred card..turn it over, then the next preferred card out of 4, etc. Procedure continues until 1 card is left.	Interview	621	Patients with depressive disorder and anxiety disorder Mean age 52.4 +/- 13.7. 75% female. 34% had completed secondary studies and 21.7% had a university degree	Community mental health center

Decisional Engagement Scale (DES-10) (English)	Hoerger, Chapman, Mohile, & Duberstein (2016) (United States)	To develop and test a tool to assess patients' engagement in cancer care	1 (10)	10-point Likert scale (disagree to agree)	Electronic survey	376	Oncology diagnosis, Mean age 58.26 (SD = 10.44), across 44 U.S. states, 67.3% male, 78.2% married, 94.1% Caucasian, 66.9% at least some college education	Community (recruited from an online database of 85,000 volunteers)
Desire to Participate in Medical Decision Making Scale (English)	Golin, DiMatteo, Leake, Duan, & Gelberg (2001) (United States)	Develop and test a diabetes-specific scale to test patient desire to participate in medical decision making. * the term 'diabetes' can be substituted with 'health' to make it a general measure	2 (11) Desire for discussion (5 items) and Desire for information (6 items).	4-point Likert scale	Phone interview	65	Type 2 diabetic patients. 67% female, mean age 54, 42% African American, 24% Latino, 31% Caucasian, 3% Asian	General medical clinic

Health Care Empowerment Questionnaire (HCEQ) (English)	Gagnon, Hebert, Dube, & Dubois (2006) (Canada)	Develop and test an instrument to measure individual empowerment in relation to personal health care and services	3 (10) Involvement in decisions (3 items); Involvement in interactions (4 items); Degree of control (3 items)	4-point Likert scale	Verbally	873	Older than 75 years.	In personal homes
Health Care Empowerment Questionnaire (HCEQ) (Persian)	Mohebbi, Tol, Shakibazadeh, Yaseri, Sabouri, & Agide (2018) (Iran)	Test psychometric properties of Persian version of HCEQ	3 (10) Involvement in decisions (3 items); Involvement in interactions (4 items); Degree of control (3 items)	4-point Likert scale	Verbally	549	Reproductive aged women, Mean age 31 (SD 5.2), 82.2% had diploma and higher education, 95.8% married	Ambulatory care
Communication preferences of patients with chronic illness: KOPRA questionnaire (German)	Farin, Gramm & Kosiol (2011) (Germany)	Develop and test a patient-oriented questionnaire to assess communication preferences of chronically ill patients	4 (32) Patient participation and patient orientation (12 items), effective and open communication (10 items), emotionally	5-point Likert scale	Written survey	333 for physician form; 89 for nurse form; 50 for therapist form	Patients with chronic back pain or chronic ischemic heart disease	Inpatient rehab

			supportive communication(6 items), and communication about personal circumstances (5 items)					
Patient Attitudes and Beliefs Scale (PABS) (English)	Arora, Ayanian, & Guadagnoli (2005) (USA)	Develop and test an instrument to assess patients' attitudes and beliefs about participating in medical decision-making	2 (12) Patients' positive (7 items) and negative (5 items) perceptions about participating in decision-making	5-point Likert scale	Written survey	621	Age (mean) = 45.3 (SD = 17.2); 70.4% women, 43.5% Caucasian, 27.9% married, 30.3% had college degrees	Primary care clinic

			care provider modified model: RMSEA = 0.058 (+); CFI = 0.97 (+); SRMR = 0.056 (+)									
API (Morandi et al. (2017) United Kingdom (English)	285	Very Good	RMSEA .074 [.067, .081] (-), CFI .915 (-) , TLI=.905 (-)	285	Very good	Cronbach's alpha .77 (+) for decision- making general; Cronbach's alpha .86 (+) for decision- making scenarios; Cronbach's alpha .80 (+)for information- seeking	124	Doubtful	ICC .48 for decision- making general at 6 months (-) and .43 at 12 months (-); ICC .44 for decision- making scenarios at 6 months (-) and .31 at 12 months (-); ICC .14 for information- seeking at 6 months (-) and .19 at 12 months (-)			
Pooled or summary result for API (overall rating)	2,482		Different factor structures supported 2 (+) 1(-)	2,298		.74-.83 (+)	267		mixed results (?) (+ (-)	421		Results in line with 6 hypo's (6+) and not in line with 5 hypo's (5-)
CPS (Degner & Sloan, 1992) Canada (English)		Inadequate								918	very good	Results in line with 3 hypo's (3+) Results not in line with 2 hypo's (2-)

CPS (Sung, Raker, Myers, & Clark, 2010) United States (English)								93	Doubtful	ICC .5 (-)	104	very good	Results not in line with 1 hypo (1-)
CPS (De las Cuevas & Penate, 2016) Spain (Spanish)					621	Inadequate	Cronbach's alpha .72 (+)				621	very good	Results in line with 2 hypo's (2+)
CPS (Beaver et al. 1996) United Kingdom (English)											350	very good	Results in line with 3 hypo's (3+) Results not in line with 1 hypo (1-)
CPS (Giordano et al. 2008) Italy (Italian)								35	Adequate	Weighted kappa .65 (-)	140	very good	Results in line with 1 hypo (1+) Results not in line with 1 hypo (1-)
Pooled or summary result for CPS (overall rating)				No factor structure	621		(+)			(-)	2,133		Results in line with 9 hypo's (9+) and not in line with 6 hypo's (5-)
DES-10 (Hoerger, Chapman, Mohile, & Duberstein, 2016) United States (English)		Inadequate	376	Adequate	CFA: CFI = .92, NNFI = .89, IFI = .92, RMSEA = .04 (+) Item-total correlations ranged .30-.61	376	very good	Cronbach's alpha 0.80 (+)			376	Very good	Results in line with 5 hypo's (5+)
DPMDM (Golin, DiMatteo, Leake, Duan, & Gelberg, 2001) United States (English)	22	Very good		Inadequate	?	65	very good	Cronbach's alpha .81 for desire for discussion and .85 for desire for information (+)	42	Doubtful	$r=.71$ Pearson's correlation coefficient (?)	Doubtful	Results in line with 4 hypo's (4+). Results not in line with 7 hypo's (7-)

HCEQ (Gagnon, Hebert, Dube, & Dubois, 2006) Canada (English)	8	Doubtful	873	Inadequate	CFI = .979; HFI = .964; AGFI = .947; RMSEA = .052 (+)	873	very good	Cronbach's alpha .79 for Involvement in decisions (+); .79 for Involvement in interactions (+); and .89 for Degree of control (+); .83 for total scale (+)	38	Doubtful	ICC .62 for Involvement in decisions (-); .70 for Involvement in interactions (+); and .60 for Degree of control (+); .70 for total scale (+)
HCEQ (Mohebbi, Tol, Shakibazadeh, Yaseri, Sabouri, & Agide, 2018 Iran (Persian)				Inadequate	No properties given (?)	549	very good	Cronbach's alpha .62 for Involvement in decisions (-); .71 for Involvement in interactions (+); and .76 for Degree of control (+); .70 for total scale (+)			
Pooled or summary result for HCEQ (overall rating)			873		(+)(?)	1,422		(+) (-)	38		(-)
KOPRA (Farin, Gramm, & Kosiol (2011) Germany (German)	10	Adequate	472	Inadequate	RMSEA PPPO .08, EOC .09, ESC .07, and CPC .12. (-)	472	very good	Cronbach's alpha .92 for Patient participation and orientation (+); .89 for Effective and open communication (+); .84 for Emotionally supportive communication			

							(+); and .80 for Communication about personal circumstances (+)			
PABS (Arora, Ayanian & Guadagnoli, 2005) United States (English)	Inadequate	621	Doubtful	GFI=.941, CFI = .919, RMSEA = .056, all factor loadings significant $p < .05$ (+)	621	very good	Cronbach's alpha for pros was .71 and cons was .72 (+)	621	Very good	Results in line with 1 hypo (1+)

MQ = methodological quality; QM = quality of measurement property; NA = not applicable; + = sufficient; - = insufficient; ? = indeterminate

API= Autonomy Preference Index; CPS = Control Preferences Scale; DPMDM = Desire to Participate in Medical Decision-Making; HCEQ = Healthcare Empowerment

Questionnaire; KOPRA = Communication Preferences of Patients with Chronic Illness; PABS = Patient Attitudes & Beliefs Scale

Table 3

Summary of evidence and overall grade of evidence quality

Instrument	Structural validity	Internal consistency	Reliability	Hypothesis testing	Modified GRADE rating ^a
4Ps	-	+	-	NA	Moderate
API	±	±	±	±	Low
CPS	NA	+	-	±	Low
DES-10	+	+	NA	+	Moderate
DPMDM	?	+	?	±	Low
HCEQ	+	±	-	NA	Low
KOPRA	-	+	NA	NA	Low
PABS	+	+	NA	+	Low

API= Autonomy Preference Index; CPS = Control Preferences Scale; DPMDM = Desire to Participate in Medical Decision-Making; HCEQ = Healthcare Empowerment Questionnaire; KOPRA = Communication Preferences of Patients with Chronic Illness; PABS = Patient Attitudes & Beliefs Scale

^a Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, 27(5), 1171-1179.

High: Confident that the true measurement property lies close to that of the estimate of the measurement property.

Moderate: Moderately confident that the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different.

Low: Confidence is limited. The true measurement property may be substantially different from the estimate of the measurement property.

Very low: Very little confidence in the measurement property estimate: the true measurement property is likely to be substantially different from the estimate of the measurement property.

+ Sufficient

- Insufficient

± Inconsistent (conflicting findings between studies)

? Indeterminate (incomplete data to make a decision)

NA Not reported/applicable

Supplementary Material 1

Complete search strategy

Date	Database	Search Strategy	Results
PubMed	3/1/2019	((("Patient Participation"[Mesh] OR "Patient Participation" OR "Patient Involvement" OR "Patient Empowerment" OR "Patient Participation Rates" OR "Patient Participation Rate" OR "Patient Engagement" OR "Patient Preference"[Mesh] OR "Patient Preference" OR "Patient Preferences" OR "Patient Participation/psychology"[Mesh] OR "Patient Preference/psychology"[Mesh])) AND ("Surveys and Questionnaires"[Mesh] OR survey* OR questionnaire* OR "Survey Methods" OR "Survey Method" OR "Survey Methodology" OR "Repeated Rounds of Survey" OR "Questionnaire design" OR "Questionnaire Designs" OR "Baseline Survey" OR "Baseline Surveys" OR Respondent* OR "Randomized Response Technique" OR "Randomized Response Techniques" OR Nonrespondent* OR "screening tool" OR "screening tools" OR scale OR scales OR Instrument*[tiab] OR "Self Report"[Mesh] OR "self report" OR "self-report" OR "Self-Assessment"[Mesh] OR "self-assessment" OR "self-assessments" OR "self assessment" OR "self assessments" OR "Self-Assessments")) NOT (("Observation"[Mesh] OR observ*)) AND ("Evaluation Studies as Topic"[Majr] OR "Evaluation Studies" OR "evaluation study" OR "Evaluation Indexes" OR "Evaluation Report" OR "Evaluation Reports" OR "Evaluation Research" OR "Use-Effectiveness" OR "Use Effectiveness" OR "Pre-Post Tests" OR "Pre Post Tests" OR "Pre-Post Test" OR "Quantitative Evaluation" OR "Quantitative Evaluations" OR "Theoretical Effectiveness" OR "Critique" OR "Critiques" OR "Evaluation Methodology" OR "Evaluation Methodologies" OR "Psychometrics"[Mesh] OR Psychometric* OR "test validity" OR "test sensitivity" OR "test specificity" OR "internal consistency" OR "Validation Studies as Topic"[Mesh] OR Validat* OR reliabil* OR Reliabilit* OR "scale development" OR "scale developed" OR "Reproducibility of Results"[Mesh] OR "reproducibility of results" OR "Reproducibility of Findings" OR "Reliability of Results" OR "Validity of Results" OR "Face Validity" OR "Reliability and Validity" OR "Validity and Reliability" OR "Test-Retest Reliability" OR "Test Retest Reliability" OR "internal consistency" OR "content validity" OR "construct validity") Filters: English; Adult: 19+ years	1,181
CINAHL	3/1/2019	((MH "Consumer Participation") OR "Consumer Participat*" OR "consumer involve*" OR "consumer empower*" OR "consumer participat* rate*" OR "consumer engag*" OR	859

		"Patient Participat*" OR "Patient Involv*" OR "Patient Empower*" OR "Patient Participat* Rate*" OR "Patient Engag*" OR (MH "Patient Preference") OR "Patient Preference*" OR "Consumer preference*" OR (MH "Patient Attitudes") OR "Patient Attitude" OR "Patients: Attitudes and Perceptions") AND ((MH "Surveys") OR survey* OR "Descriptive Survey Method*" OR "Health Survey*" OR "Baseline Survey*" OR "Survey Method*" OR "Repeated Rounds of Survey" OR (MH "Questionnaires") OR "Questionnaire design" OR "Questionnaire Designs" OR (MH "Research Instruments") OR "research instrument*" OR TI instrument* OR AB instrument* OR questionnaire* OR Respondent* OR "Randomized Response Technique*" OR Nonrespondent* OR "screening tool*" OR (MH "Scales") OR scale OR scales OR "rating scale*" OR MH "Self Report" OR "Self Report*" OR "self-report*") NOT ((MH "Observational Methods") OR "observ* method*" OR Observ*) AND ((MH "Evaluation") OR Eval* OR "eval* stud*" OR "eval* index*" OR "eval* report*" OR "eval* research*" OR (MH "Evaluation Research") OR "Use-Effective*" OR "Use Effective*" OR "Pre-Post Test*" OR "Pre Post Test*" OR "Qualit* Eval*" OR "Quant* Eval*" OR "Theoretical Effectiveness" OR Critique* OR "Eval* Method*" OR (MH "Psychometrics") OR Psychometric* OR (MH "Reliability and Validity") OR "Reliab* and Valid*" OR "Validit*and Reliab*" OR "test* valid*" OR "test sensitiv*" OR "test specific*" OR (MH "Internal Consistency") OR "Internal Consistency" OR (MH "Validation Studies") OR "Valid* Research" OR "Valid* Stud*" OR Validat* OR reliab* OR "scale* develop*" OR (MH "Reproducibility of Results") OR Reproduc* OR "reproduce* of result*" OR "Reproduce* of Find*" OR "Reliab* of Result*" OR "Valid* of Result*" OR (MH "Face Validity") OR "face valid*" OR (MH "Test-Retest Reliability") OR "Test-Retest Reliability" OR "Test Retest Reliability" OR (MH "Internal Consistency") OR "Internal Consistency" OR (MH "Content Validity") OR "Content Valid*" OR (MH "Construct Validity") OR "Construct Valid*" OR (MH "Instrument Validation") OR "Instruments Validation" OR "Validation of Instrument" OR "Validation of Instruments" OR (MH "Instrument Construction") OR "Construction of Research Instrument*" OR "Instrument Development" OR "Instrument* Construction") AND Adult Limiters - English	
PscyINFO	3/4/2019	(DE "Client Participation" OR "Client Participat*" OR "Patient Participat*" OR "Patient Involve*" OR "Patient Empower*" OR "Patient Participation Rate*" OR "Patient Engag*" OR "Client Involve*" OR "Client Empower*" OR "Client Participation Rate*" OR "Client Engag*" OR "Consumer Participat*" OR "Consumer Involve*" OR "Consumer Empower*" OR "Consumer	748

		Participation Rate*" OR "Consumer Engag*") AND (DE "Surveys" OR Survey* OR (DE "Likert Scales") OR "Likert Scale*" OR DE "Questionnaires" OR Questionnaire* OR "Descriptive Survey Method*" OR "Health Survey*" OR "Baseline Survey*" OR "Survey Method*" OR "Repeated Rounds of Survey" OR "Questionnaire design*" OR "research instrument*" OR TI instrument* OR AB instrument* OR Respondent* OR "Randomized Response Technique*" OR Nonrespondent* OR (DE "Screening Tests") OR "Screening Test*" OR "screening tool*" OR (DE "Rating Scales") OR "Rating Scale*" OR scale OR scales OR (DE "Self-Report") OR "Self-Report*" OR "self report*" OR (DE "Self-Evaluation") OR "Self-Eval*" OR "self eval*" OR (DE "Self-Monitoring") OR "Self-Monitor*" OR "self monitor*" OR DE "Self-Perception" OR "self-perce*" OR "self perce*" OR "Self-Observ*" OR "Self observ*") NOT (DE "Observation Methods" OR "observ* method*" OR Observ*) AND ((MM "Evaluation") OR Eval* OR "eval* stud*" OR "eval* index*" OR "eval* report*" OR "eval* research*" OR "Use-Effective*" OR "Use Effective*" OR "Pre-Post Test*" OR "Pre Post Test*" OR "Qualit* Eval*" OR "Quant* Eval*" OR "Theoretical Effectiveness" OR Critique* OR "Eval* Method*" OR (DE "Psychometrics") OR Psychometric* OR DE "Test Reliability" OR "Test* Reliab*" OR DE "Test Validity" OR "test valid*" OR "Reliab* and Valid*" OR "Validit* and Reliab*" OR "test sensitiv*" OR "test specific*" OR "Valid* Research" OR "Valid* Stud*" OR Validat* OR reliabil* OR "scale* develop*" OR Reproduc* OR "reproduce* of result*" OR "Reproduce* of Find*" OR "Reliab* of Result*" OR "Valid* of Result*" OR "face valid*" OR "Test-Retest Reliability" OR "Test Retest Reliability" OR "Internal Consistency" OR "Content Valid*" OR "concept valid*" OR "concurrent valid*" OR "construct valid*" OR "convergent valid*" OR "criterion related valid*" OR "discriminant valid*" OR "instrument* valid*" OR "validation of instrument*" OR "instrument construct*" OR "construction of research instrument*" OR "instrument develop*") AND adult Limiters – English	
Embase	3/1/2019	'patient participation'/exp OR 'patient preference'/exp OR 'patient engagement'/exp OR 'patient engagement':ab,ti OR 'patient participation':ab,ti OR 'patient involvement':ab,ti AND 'validation study'/exp OR 'reliability'/exp OR 'psychometry'/exp OR 'validity'/exp AND Restrict to adults ([adult]/lim OR [aged]/lim OR [middle aged]/lim OR [very elderly]/lim OR [young adult]/lim)	690

Figure 1

Flowchart of study selection. API= Autonomy Preference Index; CPS = Control Preferences Scale; DPMDM = Desire to Participate in Medical Decision-Making; HCEQ = Healthcare Empowerment Questionnaire; KOPRA = Communication Preferences of Patients with Chronic Illness; PABS = Patient Attitudes & Beliefs Scale