Quality of primary care from the patient’s point of view. A systematic review. There is currently no acceptable tool for practice or research purposes.

Jeremy Derriennic  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest  
https://orcid.org/0000-0003-1659-5072

Patrice Nabbe  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest

Marie Barais  
Université de Bretagne Occidentale

Sophie Lalande  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest

Delphine Le Goff  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest

Thomas Pourtau  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest

Benjamin Penpennic  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest

Jean-Yves Le Reste  
Université de Bretagne Occidentale Faculté de Médecine et des Sciences de la Santé de Brest

Research article

Keywords: Patient-satisfaction, patient-experience, primary care, multi-disciplinary health-care centres, systematic review

Posted Date: September 9th, 2019

DOI: https://doi.org/10.21203/rs.2.14226/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background To run an inventory of instruments which assess quality of care from the patients' experiences in primary care, in the context of multi-disciplinary health-care centres and which appraise their measurement quality, taking into account the methodological quality of their validation studies.

Method Systematic review using Medline, Pascal, PsyNINFO, Google Scholar, Cochrane, Scopus and CAIRN. For each instrument identified, the level of evidence was assessed using the COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist; the appraisal of the psychometric quality of the measurement property using three possible quality scores and the best-evidence synthesis based on the number of studies, their methodological and psychometric quality, and the direction and consistency of the results. Details of the subscales used to capture patients' experiences of primary care were extracted and synthesized by grouping them into 9 dimensions defined by the Institute Of Medicine (IOM).

Results 29 articles describing 29 instruments were found. Constructs captured by the assessment tools included illustrated a diversity of conceptualizations of patients' experiences of primary care. There was no clear consensus across the assessment tools included about what patient experience in primary care ought to measure. There is an overall lack of evidence of their measurement quality, either because validation is missing or because methods are poor.

Conclusion Due to the lack of evidence, the choice for the most appropriate instrument is difficult. Improvement and validation of existing instruments, and the use of COSMIN-guidelines could help make evaluations more effective.

Background

Primary care is changing from mono practice to multi-professional clinics with the support of public authorities. Multi-professional clinics are the rule rather than the exception. In the context of an aging population, where the prevalence of multi-morbidity is growing (1), disease centered care had to give way to patient-centered care. Multi-professional clinics should help meet this new demand for primary care (2). The application of new primary health care models, such as the Chronic care model (3) or the Patient-centered medical home (4,5), appears to be the favored option.

Today, evaluation of multi-professional structures focuses on their economic impact. Public authorities concluded that, beyond productivity, quality of care and patient satisfaction remain to be assessed (6). Accordingly, quality of care became a priority issue for public health authorities and the World Health Organization (WHO) (7).

The WHO defines quality of care as “the extent to which health care services provided to individuals and patient populations improve desired health outcomes. In order to achieve this, health care must be safe, effective, timely, efficient, equitable and people-centered”. In 1980, Donabedian (1919–2000) was already affirming the need for patient-centered care (8). He distinguished three facets from the patient's perspective needed in order to give the patient a full role in ensuring the quality of health care. First, by defining what quality is, second, by evaluating quality, third, by providing information that permits others to evaluate it (9). More recently, the Institute Of Medicine (IOM) established patient-centered care as a requirement of modern medicine (10).

Information about quality of care could be classified under three categories (8): “structure”, “process”, and “outcome”. Structure denotes the attributes of the settings in which care occurs. This includes the attributes of material resources (such as facilities, equipment, and money), of human resources (such as the number of personnel and their qualifications), and of organizational structure (such as medical staff organization and methods for peer reviewing). Process denotes what is actually done in giving and receiving care. It includes the patient's activities in seeking care and obtaining it, as well as the practitioner's activities in making a diagnosis and recommending or implementing treatment. Outcome denotes the effects of care on the health status of patients. Improvements in the patient's knowledge and salutary changes in the patient's behaviour are included under a broad definition of health status, as is the degree of the patient's satisfaction with the care provided.

The IOM conducted consultations with professionals and patients (10) highlighting nine domains of care as priorities and requiring assessment by patients:

Respect for patients' values, preferences, and expressed needs: This is a major concern of patient-centered care. Some patients wish to avoid risk; others may choose a risky intervention despite a relatively low likelihood of benefit. Patients’ preferences are likely to change over time and to depend on the clinical problems in question. Therefore, the enterprise of shared decision making is a dynamic one, changing as patients and circumstances change.

Coordination and integration of care: Many patients depend on those who provide care to coordinate services to ensure that accurate and timely information reaches those who need it at the appropriate time. Patient-centered care addresses the need to manage smooth transitions from one setting to another or from a health-care to a self-care setting.

Information, communication, and education: With respect to their health, people want to know what is wrong (diagnosis) or how to stay well, what is likely to happen and how it will affect them (prognosis) and how manage this. Patients are diverse in the way they prefer to interact with caregivers: some seek ongoing, personal, face-to-face relationships; others prefer to interact only when unavoidable. Common to all interactions is the desire for trustworthy information that is attentive, responsive, and tailored to an individual's needs.

Physical comfort: Patients experience pain, shortness of breath, or some other discomfort. They need not undergo such suffering, especially at the end of life. Many patients fail to receive state-of-the-art pain relief or respiratory management. Attention to physical comfort implies timely, tailored, and expert management of such symptoms.

Emotional support - relieving fear and anxiety: Suffering is more than just physical pain and other distressing symptoms; it also encompasses significant emotional and spiritual dimensions. Patient-centered care stresses that special attention is paid to anxiety due to uncertainty, fear of pain, disability, loneliness and social impact.
Involvement of family and friends: This dimension of patient-centered care focuses on accommodating family and friends on whom patients may rely, involving them, as appropriate, in decision making, supporting them as caregivers, making them welcome and comfortable in care and recognizing their needs and contributions.

Timeliness: Timeliness is an important characteristic of any service and is a legitimate and valued focus of improvement in health care. However, long waits are the rule rather than the exception. In addition to emotional distress, physical harm may result, for example, from a delay in diagnosis or treatment that results in preventable complications.

Efficiency: In an efficient health care system, resources are used to get the best value for money. It is also true for most improvements in safety, which result in fewer injuries, continually reduce the burden of illness, injury, and disability, and improve the health and functioning of the people.

Equity/accessibility: The aim of equity is to secure the benefits for all patients. Equity in care implies universal access.

In the face of growing interest in patients' perceptions of their care, many researchers are interested in patient satisfaction and experience. Satisfaction is the patient's appreciation of the care received (9). Satisfaction surveys show an optimistic and limited picture of care (11). They do not really assess what has to be improved (12). Faced with this, researchers became more interested in the patient's experience (13).

Patient Reported Outcomes Measures (PROMs) are tools used to measure patient-reported outcomes (14). PROMs are standardized, validated questionnaires that are completed by patients during the perioperative period to ascertain perceptions of their health status, perceived level of impairment, disability, and health-related quality of life. They allow the efficacy of a clinical intervention to be measured from the patients' perspective. Questionnaires are given to patients both pre- and post-operatively to allow comparison of outcomes pre- and post-procedure. In addition to outcomes relating to interventions, PROMs measure patients' perceptions of their general health or their health in relation to a specific disease. PROMs are a means of measuring clinical effectiveness and safety.

In parallel, Patient Reported Experience Measures (PREMs) gather information on patients' views of their experience whilst receiving care (14). They are an indicator of the quality of patient care, although do not measure it directly. PREMs are most commonly in the form of questionnaires. In contrast to PROMs, PREMs do not look at the outcomes of care but the impact of the process of the care on the patient's experience e.g. communication and timeliness of assistance. They differ from satisfaction surveys by reporting objective patient experiences, removing the ability to report subjective views.

In the context of multi-morbidity, the two measures are intertwined, and the distinction has little significance (14).

The USA were the first to conduct a standardized national survey of patient care assessment: The Hospital Consumer Assessment of Health Care Providers and Systems (HCAHPS) (15). In the UK, the UK government launched a survey of primary care patients in 2006 (16). The National Health Service (UK) is now encouraging local surveys in consultation with patients, in addition to national surveys (17). Canada (18), is also experimenting with patient surveys. No data exists in France in the context of primary care.

Measuring patient perception of their experience of primary care, in its fully multidimensional sense, requires a robust instrument of measurement. Reviews have shown that many instruments have been developed over time (19,20). These existing reviews have not systematically appraised the measurement properties of the instruments found. Therefore, a systematic review was needed to identify the instruments which measured quality of primary care from the patient's perspective and also evaluated the measurement properties of these instruments.

Methods

To minimize potential sources of bias, this systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Search Strategy

We conducted a systematic literature search in Medline, Pascal, Cockrane, Scopus, Cairn, PsyCINFO and Google Scholar between 1990 and November 2018. 1990 was chosen as the start date as multi-professional practice only emerged as the main practice model in that year. It would therefore be very unlikely that relevant instruments developed before 1990 would explore multi-professional domains for quality of primary care. Different Medical Subject Headings (MeSH) and keywords in four domains were used, including: “questionnaire”, “patient satisfaction”, “patient experience”, “primary health care”, with the help of the Health library of Brest University. The search was restricted to English or French Language articles. Reference lists were screened to identify additional relevant studies. The search strategy for Medline can be found in an additional file [see additional file 1].

Selection of eligible articles

The search aimed to include all articles that described the development or evaluation of instruments that measured quality of primary care. Articles that evaluated instruments which measured other constructs (such as quality of life, health status, preferences...) were not included. The inclusion criteria are presented in Table 1.

Titles and abstracts were independently screened by two researchers (JD and TP). If the title and abstract did not clearly indicate whether the inclusion criteria had been met, full-text was obtained and reviewed by the same researchers. Where necessary, a third reviewer (JYLR) was consulted for a final decision. The bibliography of each article included was then checked, while following the same inclusion process, to add their specific bibliography to the pool of articles included.
Data extraction

For each article included we extracted data on the methods and results for measurement properties and interpretability (see Table 2). In the case of an article describing the evaluation of multiple instruments, the data extraction was performed separately for each instrument investigated.

For each instrument identified within the articles included, the data extracted was as follows:

i. The measurement characteristics, i.e. underlying measurement model, number of subscales and items, response scale, and score range,

ii. Dimensions for quality of care: for each questionnaire, we identified which dimensions for quality of care were measured, and

iii. Measurement properties.

The measurement properties within each questionnaire are categorized within three domains, according to the COSMIN taxonomy: (1) reliability (including internal consistency, reliability and measurement error), (2) validity (including content validity, structural validity, criterion validity, cross-cultural validity and hypothesis testing (construct validity)), and (3) responsiveness. For each article included, the data was extracted by one team member (TP) and checked by a second project team member (JD); differences of opinion between these two were discussed until consensus was reached. Where there was any doubt, a third researcher was consulted (JYLR). In addition, interpretability was also described. Interpretability is the degree to which one can assign qualitative meaning to quantitative scores. This means that investigators should provide information about clinically meaningful differences in scores between subgroups, floor and ceiling effects, and the minimally important change.

Quality appraisal

For each instrument, the quality of nine measurement properties and interpretability were appraised (see Table 2) and described in the validation studies in two ways: first, the quality of the methods used to evaluate the measurement properties of the instrument (from here on referred to as the appraisal of methodological quality) was appraised; secondly, the measurement properties, based on the results of the validation studies, were appraised. Data from these two appraisals were combined to provide a best-evidence synthesis of the measurement properties for each instrument included.

Appraisal of methodological quality

The COSMIN criteria (21,23) were used to assess methodological quality. The COSMIN checklist describes how nine different measurement properties should ideally be evaluated and provides scoring criteria for the methodological quality appraisal. The quality of the methods used to evaluate each measurement property is scored on a four-point rating scale: “excellent”, “good”, “fair”, or “poor”. An additional box was used to assess requirements for studies that used Item Response Theory (IRT). For interpretability, two aspects were evaluated, i.e. floor and ceiling effects, minimally important change and minimally important difference values. More information on COSMIN and the checklist items can be found on the website: https://www.cosmin.nl/.

Appraisal of the measurement properties

Criteria developed by Terwee et al (22) and Schellingerhout et al (24,25) (see table 3) were used to rate the measurement property of an instrument within a particular study with three possible quality scores: a positive rating (labeled +), an inconclusive rating (labeled ?), and a negative rating (labeled -).

Best-evidence synthesis

Some studies evaluated the same measurement properties for a specific questionnaire. In that event, the results from the different articles were synthesized, as suggested by Terwee et al (22). The quality of a particular measurement property was determined using the method recommended by Schellingerhout and colleagues (24,25). The appraisal of methodological quality of the studies (see 2.4.1), the appraisal of the measurement property (see 2.4.2), the number of studies assessing the property, and the consistency of the results in the case of multiple validation studies were taken into account. For this overall rating, five levels of evidence were applied (see table 4). Four members of the research team (TP, BP, JD, JYLR) rated the methodological quality and measurement property of each article, with discrepancies discussed until consensus was reached. One team member then performed the best-evidence synthesis (JD) and a second checked it (JYLR).

Results

Studies included

Electronic searches identified 2775 articles. Another 236 articles were identified by the citation check of all articles that were eligible for inclusion in this systematic review. Title and abstract screening resulted in the exclusion of 2797 records. The remaining 214 full-text articles were retrieved and assessed for eligibility. In total, 37 articles met the inclusion criteria, of which 21 were derived from the primary search and 16 from the citation check. After removing duplicates, 29 articles were included in the synthesis. The main reason for exclusion was that articles did not assess the psychometric properties of the related instrument (97 articles). The 29 articles included described the development and/or evaluation of 29 instruments that assessed patients’ experiences in primary care. Fig 1 provides the PRISMA flow chart of the inclusion process with the complete list of reasons for exclusion at each step.

Fig 1. Flow diagram of article selection process.

Overview of studies
Table 5 gives an overview of measurement tools and Table 6 provides an overview of the studies included. In total, 29 studies were included in the review, reporting on 29 different tools. 23 studies reported on the initial development and validation of the tools. The others reported on further development of an existing tool (with a different sample, assessing a different psychometric property) or compared several instruments. Most of the studies were from the UK (N = 12) and USA (N = 7). All studies reported on the validation of tools in English. The number of items in the tools included featured from 6 to 84 items. Of the 29 tools, 19 used a five-point Likert scale for response categories. Samples size of the studies varied from N = 21 to N = 190038 patients.

Table 5 should appear here

Table 6 should appear here

Constructs captured by the measurement tools included

Details of the subscales used to capture patients’ experiences of primary care were extracted from the included tools (see Table 5). Eight studies did not report any subscales. Constructs captured by the tools included illustrated a diversity of conceptualizations of patients’ experiences of primary care, captured by a wide range of different subscales (See Table 7). There was no clear consensus across the tools included about which aspects of the patient experience in primary care should be measured. Each measurement tool included captured a different conceptualization of patients’ experiences of primary care, with approximately 58 constructs. A synthesis of these constructs is shown in Table 7, grouped into the 9 dimensions of patient-centered care as defined by IOM.

Domain 1: Respect for patients’ values, preferences, and expressed needs (15 constructs)
Domain 2: Coordination and integration of care (5 constructs)
Domain 3: Information, communication, and education (6 constructs)
Domain 4: Physical comfort (1 construct)
Domain 5: Emotional support - relieving fear and anxiety (2 constructs)
Domain 6: Involvement of family and friends (1 construct)
Domain 7: timeliness (2 constructs)
Domain 8: efficiency (9 constructs)
Domain 9: equity/accessibility (8 constructs)
Others: (10 constructs)

No tool captured all of the 9 domains. Figure 2 gives an overview of the percentage of the tools included which takes into account the domains covered. The domains most frequently assessed were “respect for patients' values, preferences, and expressed needs” (59%). The least frequently assessed domains were “physical comfort” and “involvement of family and friends” (3.45%).

Fig 2. Percentage of tools which covered the domains defined by IOM.

Table 7 should appear here

Quality of design, methods and reporting

Table 8 provides an overview of the assessment of the methodological quality of the studies included using the COSMIN criteria and checklist with 4-point scale ratings. While most studies used the “classical test theory” (CTT), one study used the “item response theory” (IRT). No single study assessed all the measurement properties. Across the all the measurement tools included, only an average of three psychometric properties, of the nine possible psychometric properties available, had actually been assessed.

Table 8 should appear here

For interpretability, all the studies reported the way in which missing items had been handled. 11 studies reported the percentage of respondents with the highest possible score and the lowest possible score. Neither minimally important change (MIC) nor minimally important difference (MID) were assessed in any study.

For generalizability, most studies reported the sampling method and description. The most common sampling was convenience sampling. Most studies included patients with a wide age range. Gender distribution had been achieved in all the studies. All the studies had been conducted in Western countries.

Overall results on the best evidence synthesis of the instruments included

The best available evidence (Table 9) was unknown for 50% or more of the instruments across all their measurement properties.

Table 9 should appear here.
Discussion

Heterogeneity of the constructs

Many questionnaires have been created to capture patients’ experience in primary care. After examination, there appears to be a huge diversity in the definitions of quality of primary care. This lack of consensus resulted in failure to choose an instrument. Questionnaires identified by Lévêque and al (20) captured constructs which analyzed the following: community orientation (equity, community participation), patient-centered care (global care, family-centered care, cultural sensitivity, patient-doctor relationships, respect, communication), attributes of clinical care (technical quality, accessibility, continuity, care management, comprehensibility) and structural dimensions (information management, multi-disciplinarity). This review has found all the constructs. It is, therefore, able to offer a more comprehensive definition of the quality of primary care and one which is also in line with previous studies (20,55). By classifying the tools according to the domains they capture, designers of a study into the quality of primary care are able to choose the questionnaire best suited to fulfill its research objectives.

Lack of validity

Content validity is a widely explored property (22). No Gold Standard can, by nature, be designed for a tool which assesses quality of care. Some studies use the following additional question in their psychometric assessment: “Are you satisfied with the quality of care you have just received?” as a “gold standard”. It is a coherent approach to statistical analysis since it makes it possible to compare a subjective datum, theoretically containing an infinite number of domains, with a test containing a finite number of domains. This question, however, can only be used in statistical analysis, since it is not sufficiently descriptive to constitute an evaluative tool (56).

Some properties were not applicable, including cross-cultural validity that aims to analyze the validity of the translation of a tool. Only questionnaires translated into French have their transcultural validity evaluated.

Lack of reliability

To evaluate the reliability of an instrument, it is necessary to measure its internal consistency as well as its measurement error (57). In this review, many authors used internal consistency as the unique indicator of reliability. This is an inadequate assessment. However, a validity analysis of a tool can only be undertaken after an analysis of reliability. A tool can only be valid if it is reliable (58). The lack of scientific methodology for assessing the reliability of most of the studies included in this review makes interpretation of their validity questionable.

Lack of insight into the ability to measure change and to interpret change

No study evaluated the responsiveness of its questionnaires. Responsiveness measures the ability of an instrument to detect change in the data measured over time. For an investigator who wants to study the effects of an intervention on quality of care, it is highly important to know the responsiveness of the instrument. Finally, it is relevant to know, for a given tool, the Minimally Important Change (MIC) or the Minimally Important Difference (MID).

Strengths and limitations of the review

This review was based on a published methodology and followed all the standards required for a systematic literature review (59). It appears to be the first to perform an evaluation of psychometric properties analysis, by the COSMIN method, applied to self-report questionnaires, from the perspective of patients, in the field of the quality of primary care.

First, two raters (or four when necessary) evaluated the eligibility of articles, extracted the data, and performed the quality appraisal for each measurement property. Therefore, the results are robust. Second, to provide an unbiased appraisal of the measurement quality of the instruments included, all the results and the methodological quality of all their validation studies were taken into account. In addition, the rating of methodological quality was based on the widely accepted COSMIN standards. Third, due to the high number of instruments included, an insight is provided into overall trends regarding property measurement evaluations, their quality, and the overall quality of instruments. This insight makes it possible to provide general recommendations on how to improve instruments, and their validation studies, when assessing the quality of patients’ experiences in primary care.

The study had some limitations. Given the number of initial search criteria, even where an article had excellent sensitivity, in order to be eligible for inclusion it had to describe a study that aimed to develop, or validate, the patients’ experiences by means of a primary care evaluation instrument. Consequently, it was possible to miss relevant articles if development or validation of an instrument was not explicitly mentioned in either its title or its abstract. In addition, with one of the aims of the study being psychometric analysis, all the questionnaires found which lacked a psychometric analysis were excluded from this review. A selection bias is therefore possible for tools without any psychometric analysis. The methodological analysis was performed using the COSMIN criteria. This methodology has been put in place to minimize information biases. Nevertheless, the possibility of this type of bias remains.

Interest and future implications

This is the first systematic review of the literature on a patient self-assessment tool which is concerned with the quality of primary care and which includes an analysis of its psychometric properties. It enables designers of primary care quality studies to understand the strengths and limitations of existing instruments in terms of captured constructs and psychometric properties. It also reveals, for a given questionnaire, the weaknesses of its psychometric properties and possibly creates the opportunity to design a study to reinforce these properties. Despite a growing interest in evaluating the quality of care and the abundance of instruments validated within the framework of hospital structures, particularly in the context of their accreditation, few psychometric evaluations of instruments have been developed in primary care.
Future research will need to develop or validate a generic tool for assessing quality of care in primary care. This instrument should be reliable, valid, reactive and interpretable so that it can be used within a healthcare system and so that it allows comparability, both in space (comparing the quality of care of two healthcare homes to define the optimal organization) and over time (comparing the results before and after an intervention to measure the benefit). This review could be the starting point for such work and provide a solid foundation. From this point, a researcher will be able to identify the most suitable questionnaire for his/her work and supplement or reinforce the psychometric analysis. Later work could involve translating it, adapting it to different cultures and finally testing it in each environmental location. This can also be the starting point for the creation of a new and better validated health assessment questionnaire on the quality of primary care.

Conclusions

This systematic review shows that many patient self-assessment tools, which are concerned with the quality of primary care and which include an analysis of their psychometric properties, exist. However, comparison across instruments highlighted a wide variability in terms of captured constructs and psychometric properties. It also reveals, for a given questionnaire, the weaknesses of its psychometric properties. High quality studies are needed to reinforce these properties or to develop a generic tool for assessing quality of care in primary care. This instrument should be reliable, valid, reactive and interpretable.

List Of Abbreviations

COSMIN: COnsensus-based Standards for the selection of health status Measurement Instruments.
IOM: Institute Of Medicine.
WHO: World Health Organization.
PROMs: Patient Reported Outcomes Measures.
PREMs: Patient Reported Experience Measures.
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
IRT: Item Response Theory.
CTT: Classical Test Theory.
QOPC: Quality Of Primary Care
MIC: Minimally Important Change.
MID: Minimally Important Difference.

Declarations

Ethics approval and consent to participate: Not applicable.
Consent for publication: Not applicable.
Availability of data and material: All data generated or analysed during this study are included in this published article
Competing interests: The authors declare that they have no competing interests.
Fundings: The authors declare that there was no source of funding for the research.
Authors'contributions: JD and JYLR were responsible for study conception. JD and JYLR developed the systematic review protocol. TP, BP, PN and MB contributed to the development of the protocol. JD conducted the literature search. JD, TP, BP and JYLR scanned selected titles and abstracts, assessed full-text versions independently and performed the narrative synthesis. JD wrote the first draft of the manuscript. PN, MB, SL and JYLR revised the manuscript critically. All authors read and approved the final manuscript.

Acknowledgements: We thank Georges Buzet (Library, University of Western Brittany, Brest, France) for his assistance in performing the search and for his contribution during the selection of eligible studies and Dr. C. B. Terwee (Department of Epidemiology and Biostatistics and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands) for her advice on questions concerning the interpretation and application of the COSMIN guidelines.

References

1. Le Reste JY, Nabbe P, Manceau B, Lygidakis C, Doerr C, Lingner H, et al. The European General Practice Research Network presents a comprehensive definition of Multimorbidity in Family Medicine and Long-Term Care, following a systematic review of relevant literature. J Am Med Dir Assoc. 2013;14(5): 319–25.
2. Cutler DM. Where Are the Health Care Entrepreneurs? The Failure of Organizational Innovation in Health Care. Innov Policy Econ. 2011;11(1): 1–28.
3. Wagner EH. Chronic disease Management: what will it Take to improve care for chronic illness? Eff Clin Pract. 1998;1: 2–4.
4. Rittenhouse DR, Shortell SM. The patient-centered medical home: will it stand the test of health reform? JAMA. 2009;301(19): 2038–40.
5. Rittenhouse DR, Shortell SM, Gillies RR, Casalino LP, Robinson JC, McCurdy RK, et al. Improving chronic illness care: findings from a national study of care management processes in large physician practices. Med Care Res Rev. 2010;67(3): 301–20.
6. Kruk ME, Kelley E, Syed SB, Tarp F, Addison T, Akachi Y. Measuring quality of health-care services: what is known and where are the gaps? Bull World Health Organ. 2017;95(6): 389.
7. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. Bull World Health Organ. 2017;95(5): 368–74.
8. Donabedian A. The quality of care: how can it be assessed? JAMA. 1988;260(12): 1743–8.
9. Donabedian A. The lichfield lecture. Quality assurance in health care: consumers’ role. Qual Health Care. 1992;1(4): 247–51.
10. Crossing the quality chasm: A New Health System for the 21st Century. Institute of Medicine (US). Committee on Quality of Health Care in America. National Academies Press (US). 2001.
11. Fan VS, Burman M, McDonell MB, Fihn SD. Continuity of care and other determinants of patient satisfaction with primary care. J Gen Intern Med. 2005;20(3): 226–33.
12. Jenkinson C, Coulter A, Bruster S, Richards N, Chandola T. Patients’ experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care. Qual Saf Health Care. 2002;11(4): 335–9.
13. Salisbury C, Wallace M, Montgomery AA. Patients’ experience and satisfaction in primary care: secondary analysis using multilevel modelling. BMJ. 2010; 341:c5004.
14. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Educ. 2017;17(4): 137–44.
15. Goldstein E, Elliott MN, Lehrman WG, Hambarsoomian K, Giordano LA. Racial/Ethnic differences in patients’ perceptions of inpatient care using the HCAHPS Survey. Med Care Res Rev. 2010;67(1): 74–92.
16. Roland M. Measuring and improving patient experience in primary care. Prim Health Care Res Dev. 2012;13(02): 103–5.
17. Asprey A, Campbell JL, Newbold J, Cohn S, Carter M, Davey A, et al. Challenges to the credibility of patient feedback in primary healthcare settings: a qualitative study. Br J Gen Pract. 2013;63(608): 200–8.
18. Accreditation Canada. Programme d’accréditation Qmentum international. Available from: https://accreditation.ca/intl-fr/accreditation/qmentum/ Cited 23 Jan 2019.
19. Evans RG, Edwards A, Evans S, Elwyn B, Elwyn G. Assessing the practising physician using patient surveys: a systematic review of instruments and feedback methods. Fam Pract. 2007;24(2): 117–27.
20. Lévesque J-F, Haggerty J, Beninguissé G, Burge F, Gass D, Beaulieu M-D, et al. Mapping the coverage of attributes in validated instruments that evaluate primary healthcare from the patient perspective. BMC Fam Pract. 2012;13: 20.
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4): 539–49.
22. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. 2007;60(1): 34–42.
23. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7): 737–45.
24. Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. BMC Med Res Methodol. 2011;11(1): 87.
25. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res. 2012;21(4): 659–70.
26. Safran DG, Karp M, Collin K, Chang H, Li A, Ogren J, et al. Measuring patients’ experiences with individual primary care physicians. J Gen Intern Med. 2006;21(1): 13–21.
27. Flocke SA. Measuring attributes of primary care: development of a new instrument. J Fam Pract. 1997;45(1): 64–74.
28. Haggerty JL, Burge F, Beaulieu M-D, Pineault R, Beaulieu C, Lévesque J-F, et al. Validation of instruments to evaluate primary healthcare from the patient perspective: overview of the method. Health Policy. 2011;7(Spec Issue): 31–46.
29. Baker R. Development of a questionnaire to assess patients’ satisfaction with consultations in general practice. Br J Gen Pract. 1990;40(341): 487–90.
30. Baker R, Whitfield M, Bs B. Measuring patient satisfaction: a test of construct validity. Qual Health Care. 1992;1(2): 104–9.
31. Kinnersley P, Stott N, Peters T, Harvey I, Hackett P. A comparison of methods for measuring patient satisfaction with consultations in primary care. Fam Pract. 1996;13(1): 41–51.
32. Gasquet I, Villeninot S, dos Santos AC, Vallet Q, Verdier A, Kovess-Masfety V, et al. Adaptation culturelle et validation de questionnaires de satisfaction à l’égard du système de santé français. Sante Publique. 2003;15(4): 383.
33. Wensing M, Mainz J, Grob R. A standardised instrument for patient evaluations of general practice care in Europe. Eur J Gen Pract. 2000;6(3): 82–7.
34. Maurice-Szumburski A, Michel P, Loundou A, Auquier P. Validation of the generic medical interview satisfaction scale: the G-MISS questionnaire. Health Qual Life Outcomes. 2017;15(1): 36.
35. Mead N, Bower P, Roland M. The General Practice Assessment Questionnaire (GPAQ)—Development and psychometric characteristics. BMC Fam Pract. 2008;9(1): 13.

36. Roland M, Roberts M, Rhenius V, Campbell J. GPAQ-R: development and psychometric properties of a version of the general practice assessment questionnaire for use for revalidation by general practitioners in the UK. BMC Fam Pract. 2013;14: 160.

37. Ramsay J, Campbell JL, Schroter S, Green J, Roland M. The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties. Fam Pract. 2000;17(5): 372–9.

38. Grogan S, Conner M, Willits D, Norman P. Development of a questionnaire to measure patients' satisfaction with general practitioners' services. Br J Gen Pract. 1995;45(399): 525–9.

39. Haddad S, Potvin L, Roberge D, Pineault R, Remondin M. Patient perception of quality following a visit to a doctor in a primary care unit. Fam Pract. 2000;17(1): 21–9.

40. Greco M, Powell ROY, Sweeney K. The Improving Practice Questionnaire (IPQ): a practical tool for general practices seeking patient views. 2003:440–8.

41. Thayaparan AJ, Mahdi E. The Patient Satisfaction Questionnaire Short Form (PSQ–18) as an adaptable, reliable, and validated tool for use in various settings. Med Educ Online. 2013;18: 21747.

42. Meakin R, Weinman J. The "Medical Interview Satisfaction Scale" (MISS–21) adapted for British general practice. Fam Pract. 2002;19(3): 257–63.

43. Wolf MH, Putnam SM, James SA, Stiles WB. The Medical Interview Satisfaction Scale: development of a scale to measure patient perceptions of physician behavior. J Behav Med. 1978;1(4): 391–401.

44. Bamford C, Jacoby A, Bamford M. Development of patient satisfaction questionnaires: I. Methodological issues. Qual Heal Care. 1992;1: 153–7.

45. Safran DG, Kosinski M, Tarlov AR, Rogers WH, Taira DH, Lieberman N, et al. The Primary Care Assessment Survey: tests of data quality and measurement performance. Med Care. 1998;36(5): 728–39.

46. Shi L, Starfield B, D Jiahong Xu AN, Baltimore M, Carolina S. Validating the Adult Primary Care Assessment Tool. The Journal of Family Practice. 2001;50(2): 161–75.

47. Bowman MA, Herndon A, Sharp PC, Dignan MB. Assessment of the patient-doctor interaction scale for measuring patient satisfaction. Patient Educ Couns. 1992;19(1): 75–80.

48. van der Feltz-Cornelis CM, Van Oppen P, Van Marwijk HW, De Beurs E, Van Dyck R. A patient-doctor relationship questionnaire (PDRQ–9) in primary care: development and psychometric evaluation. Gen Hosp Psychiatry. 2004;26: 115–20.

49. Howie JG, Heaney DJ, Maxwell M, Walker JJ. A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. Fam Pract. 1998;15(2): 165–71.

50. Steine S, Finset A, Laerum E. A new, brief questionnaire (PEQ) developed in primary health care for measuring patients' experience of interaction, emotion and consultation outcome. Fam Pract. 2001;18(4): 410–8.

51. Marcinowicz L, Rybczuk M, Grebowski R, Chlabicz S. A short questionnaire for measuring the quality of patient visits to family practices. Int J Qual Heal Care. 2010;22(4): 294–301.

52. Baker R. The Reliability and Criterion Validity of a Measure of Patients’ Satisfaction with their General Practice. Fam Pract. 1991;8(2): 171–7.

53. Vuković M, Gvozdenović BS, Gajić T, Stamatović Gajić B, Jakovljević M, McCormick BP. Validation of a patient satisfaction questionnaire in primary health care. Public Health. 2012;126(8): 710–8.

54. Ware JE, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. Eval Program Plann. 1983;6(3): 247–63.

55. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. Soc Sci Med. 1998;47(10): 1573–88.

56. Starfield B. New paradigms for quality in primary care. Br J Gen Pract. 2001;51(465): 303–9.

57. McCrae RR, Kurtz JE, Yamagata S, Terracciano A. Internal consistency, retest reliability, and their implications for personality scale validity. Pers Soc Psychol Rev. 2011;15(1): 28–50.

58. John OP, Soto CJ. The importance of being valid: reliability and the process of construct validation. Cambridge: 2007. pp. 461–94.

59. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009;6(7): e1000097.

Tables
Table 1. Eligibility criteria.

**Inclusion criteria**

1- The article had to investigate a self-report questionnaire,

2- The article had to describe a primary study in which the development or evaluation of one or more instruments occurred,

3- Instruments under investigation:
   
   a. Were developed with the aim of measuring the process of health care delivered to a patient (with or without caregiver) by at least two primary health care providers,
   
   b. Were developed or evaluated in terms of their ability to measure patients’ experience of primary care. To guarantee a focus on quality of care, these instruments should assess at least three dimensions of quality of care.

4- The article had been peer-reviewed,

5- The article was written in English or French.

**Exclusion criteria**

To guarantee that the instrument under investigation measured in-patient experience of primary care, the following three exclusion criteria were applied:

1- Articles investigating instruments that measure a health outcome such as quality of life, health status, burden of diseases, handicap ... any that did not include the process,

2- Articles about instruments were evaluated in healthcare establishments and not in general practitioner centered settings, such as emergency, medical home care,

3- Articles about instruments evaluated in a restricted sample (aging, specific condition, specific gender...).
| Measurement property | Definition |
|----------------------|------------|
| **I. Reliability**   |            |
| **Internal consistency** | The degree to which items in a (sub)scale are inter-correlated, thus measuring the same construct. |
| **Reliability**      | The extent to which subjects can be distinguished from each other, despite measurement errors (relative measurement error). |
| **Measurement error/Agreement** | The degree to which the scores on repeated measurements are close to each other (absolute measurement error). |
| **II. Validity**     |            |
| **Content validity** | The degree to which the instrument is an adequate reflection of the construct to be measured. |
| **Construct validity** | The degree to which the scores of the instrument are an adequate reflection of the dimensionality of the construct to be measured. |
| **Structural validity** | The degree to which the scores of the instrument are consistent with hypotheses, based on the assumption that the instrument validly measures the construct to be measured. |
| **Hypotheses testing** | The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument. |
| **Cross-cultural validity** | The degree to which the scores of the instrument are an adequate reflection of a ‘gold standard’. |
| **Criterion validity** | The degree to which the instrument is an adequate reflection of the same construct. |
| **III. Responsiveness** | The ability of the instrument to detect changes over time in the construct measured. |
| **Responsiveness**   | Interpretability is the degree to which one can assign qualitative meanings- that is, clinical or commonly understood connotations- to an instrument’s quantitative scores or change in scores. |
| **Interpretability** |            |
Table 3. Quality criteria for results on measurement properties based on Terwee et al. (22).

| Measurement property | Criteria for appraisal of the results on measurement properties evaluation |
|----------------------|--------------------------------------------------------------------------------|
| Internal consistency | + Cronbach’s alpha(s) are ≥ 0.70.                                               |
|                      | ? Not able to score because of unclear or missing information, e.g. the dimensionality is not known or Cronbach’s alpha(s) are not presented. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Reliability          | + ICC agreement/weighted Kappa ≥ 0.70 OR IC Consistency/ICC without approach stated/Pearson’s r ≥ 0.80 OR un-weighted Kappa/or Kappa without approach stated ≥ 0.8. |
|                      | ? Not able to score because of unclear or missing information, e.g. neither ICC, Kappa, nor Pearson’s r is determined. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Measurement error/   | + MIC ≥ SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable. |
| Agreement            | ? Not able to score because of unclear or missing information, e.g. SEM, SDC not calculated, or MIC not defined. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Content validity     | + Target group and/or experts considered all items to be relevant AND considered the items to be complete. |
|                      | ? Not able to score because of unclear or missing information, e.g. no results on item relevance according to experts’ reports. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Construct validity   |                                                                                   |
| Structural validity  | + For exploratory factor analyses: Factors chosen explain at least 50% of variance OR factors chosen explain less than 50% of variance but the choice is justified by the authors. For confirmatory factor analyses: (The goodness-of-fit indicators fulfil the following requirements: (CFI or TLI or GFI or comparable measure >0.90) AND (RMSEA or SRMR < 0.08)) AND (results confirm models with the original factor structure OR results confirm a model with slight changes if these changes are justified by the authors. |
|                      | ? For exploratory factor analyses: Not able to score because of unclear or missing information, e.g. explained variance not mentioned. For confirmatory factor analyses: Not able to score because of unclear or missing information, e.g. no fit indices are presented. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Hypothesis testing   | + (At least 75% of the results are in accordance with the hypotheses AND, if calculated, the correlation with an instrument measuring the same construct is ≥ 0.50) AND correlations with related constructs are higher than with unrelated constructs if calculated. |
|                      | ? Not able to score because of unclear or missing information, e.g. no correlations with related construct are calculated. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Cross-cultural validity | + The original factor structure is confirmed AND no important DIF found. If only one of these properties is investigated: either the factor structure is confirmed OR no important DIF found. |
|                      | ? Not able to score because of unclear or missing information, e.g. no confirmative factor analyses is performed nor is the DIF investigated. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Criterion validity   | + Correlations with chosen gold standard ≥ 0.70; OR AUC ≥ 0.80, OR (specificity AND sensitivity ≥ 80). |
|                      | ? Not able to score because of unclear or missing information. |
|                      | - Criteria for ‘+’ not met.                                                      |
| Responsiveness       | + Correlations of change scores of the target instrument with an instrument measuring the same construct are ≥ 0.40 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70 AND Correlations of change scores of the target instrument with an instrument... |
measuring a related construct are higher than with an unrelated construct if calculated.

? Not able to score because of unclear or missing information, e.g. no correlations of change score with related constructs are calculated or no AUC investigated.

- Change score correlation with an instrument measuring the same construct < 0.40 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR change score correlations with related constructs are lower than with unrelated constructs.

**Interpretability**

| Rating | Criteria |
|--------|----------|
| No quality scoring performed |

**Item response theory (IRT)**

+ At least limited evidence for uni-dimensionality or positive structural validity AND no evidence for violation of local independence: Rasch: standardized item-person fit residuals between -2.5 and 2.5; OR IRT: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3s < 0.37 AND no evidence for violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND adequate model fit: Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > -2 and <2 OR IRT: G2 >0.01. Optional additional evidence: Adequate targeting; Rasch: adequate person-item threshold distribution; IRT: adequate threshold range. No important DIF for relevant subject characteristics (such as age, gender, education), McFadden's R2 < 0.02.

? Model fit not reported.

- Criteria for ‘+’ not met.

+ = positive result for a measurement property

? = result of measurement property is unknown

- = negative result for a measurement property

---

**Table 4. Levels of evidence for the best-evidence synthesis.**

| Level of evidence | Rating | Criteria |
|-------------------|--------|----------|
| Strong            | +++ or -- | Consistent findings in multiple studies of good methodological quality OR one study of excellent methodological quality |
| Moderate          | ++ or -- | Consistent findings in multiple studies of fair methodological quality OR one study of good methodological quality |
| Limited           | + or - | One study of fair methodological quality |
| Conflicting       | +/-    | Conflicting findings |
| Unknown           | ?      | Only studies of poor methodological quality |

A plus sign (+) indicates positive results for a measurement property evaluation and a minus sign (-) indicates negative results for a measurement property evaluation, e.g. + stands for limited evidence of positive results and − stands for strong evidence of negative results for a measurement property.
| Tool                                      | Authors                      | Items | Sub-scales                                                                                                                                                                                                 | Response scale | Language |
|------------------------------------------|------------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|----------|
| Ambulatory Care Experience survey (ACES) | Safran et al (2006)          | 39    | Eleven subscales: organizational access, visit-based continuity, integration, clinical team, office staff, communication, whole-person orientation, health promotion, interpersonal treatment, patient trust, relationship and duration. | 6-Point likert Scale | English   |
| Components of Primary Care Index (CPCI)  | Flocke et al (1997)          | 19    | Seven subscales: comprehensiveness of care, accumulated knowledge, interpersonal communication, coordination of care, first contact, continuity of care and longitudinally.                                               | 5-Point likert Scale | English   |
| Consultation Satisfaction Questionnaire  | Baker et al (1990)           | 18    | Four subscales: general satisfaction, professional care, depth of relationship and perceived time.                                                                                                           | 5-Point likert Scale | English   |
| Consumer Satisfaction Survey VF (CSS-VF) | Gasquet et al (2003)         | 39    | Nine subscales: access to primary care, access to secondary care, communication and competence of general practitioner, communication of specialist, competence of specialist, choice and continuity, interpersonal care, general satisfaction and finances. | 5-Point likert Scale | French    |
| EUROPEP (28,33)                          | Comité EUROPEP (1998)        | 23    | Five subscales: relationship, technical aspects of care/competence, information and support, organization of care and access.                                                                               | Gradual scale from 1 to 5 points | French    |
| Generic Medical Interview Satisfaction Scale 16 items VF (G-MISS-16-VF) | Maurice-Szamburski et al (2017) | 16    | Three subscales: pain, communication and compliance.                                                                                                                                                        | 5-Point likert Scale | French    |
| General Practice Assessment Questionnaire (GPAQ) | Mead et al (2008)            | 46    | Five subscales: access, office staff, continuity of care, communication and medical care.                                                                                                                      | Gradual scale varying from 2 to 6 points | English   |
| General Practice Assessment Questionnaire for Revalidation (GPAQ-R) | Roland et al (2013)          | 46    | Five subscales: access, office staff, continuity of care, communication and medical care.                                                                                                                      | Gradual scale varying from 2 to 6 points | English   |
| General Practice Assessment Survey (GPAS) | Ramsay et al (2000)          | 53    | Nine subscales: access, technical aspects of care, communication, humanity, trust, accumulated knowledge, medical care, appointments and premises.                                                            | 5-Point likert Scale | English   |
| Grogan Patient Satisfaction questionnaire 40 Items (Grogan-PSQ-40) | Grogan et al (1995)          | 40    | Five subscales: general practitioner, access, nurses, appointment and facilities.                                                                                                                          | 5-Point likert Scale | English   |
| Haddad Patient Satisfaction questionnaire 22 Items (Haddad-PSQ-22) | Haddad et al (2000)          | 22    | Three subscales: relationship, technical aspects of care and outcomes.                                                                                                                                  | 5-Point likert Scale | English   |
| Improving Practice Questionnaire (IPQ)   | Greco et al (2003)           | 27    | No subscales reported (expected subscales were: facilities, office staff and general practitioner).                                                                                                         | 5-Point likert Scale | English   |
| Marshall Patient Satisfaction questionnaire 18 Items (Marshall-PSQ-18) | Marshall et al (1994)        | 18    | No subscales reported (expected subscales were: general satisfaction, technical aspect of care, communication, relation, finances, time and access.                                                           | 5-Point likert Scale | English   |
| Medical Interview Satisfaction Scale 21 items (MISS-21) | Meakin et al (2002)          | 21    | Four subscales: communication comfort, distress relief, compliance intent and rapport.                                                                                                                     | 5-Point likert Scale | English   |
| Medical Interview Satisfaction Scale 26 items (MISS-26) | Wolf et al (1978)            | 26    | Three subscales: cognitive satisfaction, affective satisfaction, and behavioral satisfaction.                                                                                                               | 5-Point likert Scale | English   |
| Scale Name | Author(s) | Item Count | Description | Scale Range | Language |
|------------|-----------|------------|-------------|-------------|----------|
| Medical Interview Satisfaction Scale 29 items (MISS-29) | Wolf et al (1981) | 29 | Four subscales: communication comfort, distress relief, rapport and compliance intent. | Scale 7 points | English |
| Newcastle MAAG Patient Satisfaction Survey Accessibility (NMPSSA) | Eccles et al (1992) | 12 | No subscales reported (expected subscales were: access and patient reception. | 5-Point likert Scale | English |
| Newcastle MAAG Patient Satisfaction Survey Interpersonal aspects of care (NMPSSIAC) | Eccles et al (1992) | 11 | Three subscales: Listening, information and global satisfaction. | 5-Point likert Scale | English |
| Primary Care Assessment Survey (PCAS) | Safran et al (1998) | 51 | Eleven subscales: Finances, access, longitudinal continuity, visit-based continuity, knowledge of the patient, preventive counseling, integration, communication, physical examination, interpersonal treatment and trust. | Scale and range scale from 1 to 100. | English |
| Primary Care Assessment Tool (PCAT) | Shi et al (2001) | 84 | Nine subscales: first contact/accessibility, first contact/using, care in progress, coordination of services, services available, services received, family centered care, community orientation and cultural competencies. | 4-Point likert Scale | English |
| Patient-Doctor Interaction scale (PDIS) | Bowman et al (1992) | 19 | No subscales reported (aim was to explore patient-doctor interactions). | 5-Point likert Scale | English |
| Patient Doctor relationship Questionnaire (PDRQ-9) | Van der Feltz-Cornelis et al (2004) | 9 | No subscales reported (aim was to assess patient-doctor relationship). | Scale 5 points | English |
| Patient Enablement Instrument (PEI) | Howie et al (1998) | 6 | No subscales reported (aim was to understand the feelings of patients after a consultation). | Scale 3 points | English |
| Patient Experience Questionnaire (PEQ) | Steine et al (2001) | 18 | Five subscales: communication, emotions, outcomes, barriers and auxiliary staff. | 5-Point likert Scale | English |
| Quality of Visit to Family Practician (QVFP) | Marcinowicz et al (2010) | 30 | Three subscales: doctor-patient relationship and consultation outcome, barriers and difficulties, accessibility to care. | 5-Point likert Scale | English |
| Surgery Satisfaction Questionnaire (SSQ) | Baker et al (1991) | 17 | Six subscales: general satisfaction, continuity, access, medical care, premises and availability. | 5-Point likert Scale | English |
| Visit-specific Satisfaction Questionnaire VF (VSQ-VF) | Gasquet et al (2003) | 9 | No subscales reported (aim was to assess the global satisfaction after a visit). | 5-Point likert Scale | English |
| Vukovic Patient Satisfaction questionnaire 20 Items (Vukovic-PSQ-20) | Vukovic et al (2012) | 20 | No subscales reported (aim was to assess the global satisfaction). | 5-Point likert Scale and dichotomous score | English |
| Ware Patient Satisfaction questionnaire 55 items (Ware-PSQ-55) | Ware et al (1983) | 55 | Seven subscales: Access to care, financial aspects, availability of resources, continuity of care, technical quality, interpersonal manner and overall satisfaction. | 5-Point likert Scale | English |
Table 6. Characteristics of studies included

| Author (year), scale | Country | Study population | Sample characteristics |
|---------------------|---------|------------------|------------------------|
| Baker et al (1992)  | UK      | Patient base of two medical groups | Sample 1: patients who have changed doctor without moving, n=272 (mean age 40.5 years, 62.4% female); sample 2: patients who have not changed doctor, n=711 (mean age 51 years, 61.4% female); test-retest sample: n=131 |
| (30) SSQ/CSQ        |         |                  |                        |
| Baker et al (1990)  | UK      | Forty consecutive patients visiting eight general practitioners | n=239 |
| (29) CSQ            |         |                  |                        |
| Baker et al (1991)  | UK      | One hundred consecutive patients visiting eight medical offices | n=691 |
| (52) SSQ            |         |                  |                        |
| Bamford et al (1992)| UK      | Three hundred and eighty randomized patients visiting five medical offices | Sample 1: study NMPSSA, n=1463 (58.5% female, 77% under 65 years); sample 2: study NMPSSIAC, n=1558 |
| (44) NMPSSA/NMPSSIAC|        |                  |                        |
| Bowman et al (1992) | USA     | Ninety-one patients who have visited the Family Practice Center de la Wake Forest University on two consecutive Mondays. Patients included were re-contacted one month later | Sample 1: patient submitting the questionnaire to their exit, n=26 (M+1, n=25); sample 2: patients submitting the questionnaire by mail, n=21 (M+1, n=11); sample 3: patients submitting the questionnaire by phone, n=31 (M+1, n=28) |
| (47) PDIS           |         |                  |                        |
| Flocke et al (1997) | USA     | Four thousand four hundred and fifty four patients visited one hundred and thirty eight medical offices in Ohio during a two-day period. | n=2899 (mean age 42 years, 62% female) |
| (27) CPCI           |         |                  |                        |
| Gasquet et al (2003)| France  | Patients randomized from the phone book | n=706 (mean age 44.8 years, 66.1% female) |
| (32) CSS-VF/VSQ-VF  |         |                  |                        |
| Grogan et al (1995) | UK      | Patients randomized from the patient base of a medical office in Norfolk | n=1193 (mean age 52.1 years, 61.1% female) |
| (38) Grogan PSQ     |         |                  |                        |
| Haddad et al (2000) | Canada  | Patients randomized from eleven medical offices in Montreal | n=473 (mean age 46 years, 76% female) |
| (39) Haddad PSQ     |         |                  |                        |
| Haggerty et al (2011)| Canada | Patients randomized from a medical office in Quebec | n=645 (mean age 48 years, SD 14.9, 64.6% female) |
| (28) PCAS/CPCI/EUROPEP|        |                  |                        |
| Kimmersley et al (1996)| UK | Patients visited eight medical offices in south Glamorgan | Sample 1: patients submitted the questionnaire as they left, n=122 (mean age 48.1 years, 66% female); Sample 2: patients submitted the questionnaire by mail, n=74 (mean age 50.2 years, 70% female) |
| (31)                 |         |                  |                        |
| Study | Country | Sample Description | Findings |
|-------|---------|--------------------|----------|
| Shi et al (2001) | USA | Patients visited two medical offices in South Carolina | n=823 |
| Marcinowicz et al (2010) | Poland | Patients visiting eight medical offices in Poland | n=1330 (22.3% aged between 45-55 years, 61.9% female) |
| Maurice-Szamburski et al (2017) | France | Patients visited doctors (all specialties) in Marseille (France) | n=1822 (mean age 37.1 years, 49% female, 28% general practice) |
| Mead et al (2008) | UK | Patients visited one thousand and thirty one medical offices in the UK | Sample 1: patients submitted the questionnaire as they left, n=190038 (mean age 50.3 years, 64.7% female); sample 2: patients submitted their evaluation by email, n=20309 (mean age 54.2 years, 61.4% female) |
| Meakin et al (2002) | UK | Patients visiting eighteen general practitioners | n=151 (mean age 43.3 years, 61.3% female) |
| Ramsay et al (2000) | UK | Consecutive patients visiting fifty-five general practitioners in London | Main sample: n=7765 (mean age 37.7 years, 62.8% female); test-retest sample: n=55 (mean age 46.7 years, 72.7% female) |
| Roland et al (2013) | UK | Patients visiting one hundred and sixty four general practitioners in twenty nine medical offices in the UK | n=7258 (70.8% under 65 years, 64% female) |
| Safran et al (2006) | UK | Patients contacted by phone and email randomized from health insurance database in the UK | n=9625 (mean age 47.2 years, 67% de female) |
| Safran et al (1998) | UK | Patients randomized from employees database in Massachusetts | n=6094 (mean age 48.8 years, 55.8% female) |
| Steine et al (2001) | Norway | Consecutive patients from thirty four general practitioners in Norway | n=1092 (mean age 47 years, SD 20, 67% female) |
| Vukovic et al (2012) | Serbia | Patients visiting a primary health care center in Serbia in 2007 and 2008 | Sample 1: patients included in 2007, n=594; sample 2: patients included in 2008, n=720 (global mean age 55 years, SD 16.56, 65.5%)}
| Reference | Country | Sample Description | Sample Size and Characteristics |
|-----------|---------|--------------------|---------------------------------|
| Vukovic PSQ-20 | female) | Ware et al (1983) USA | Patients from Illinois Sample not described |
| Ware PSQ | | Wensing et al (1998) Europe | Patients visited eight hundred and forty general practitioners in sixteen European countries n=23892 (mean age 50.6 years, 64.9% female) |
| EUROPEP | | Howie et al (1998) Scotland | Patients visiting a general practitioner in Edinburgh n=613 |
| PEI | | Greco et al (2003) UK | Patients from UK n=55687 (mean age 50.56 years, 64.8% female) |
| Marshall et al (1994) US | | Marshall PSQ | Patients visited three medical centers in three American cities (Boston, Chicago and Los Angeles) n=2197 (mean age 55.83 years, SD 16.21, 60% female) |
| Van der Feltz-Cornelis et al (2004) | Netherlands | PDRQ-9 | Consecutive patients visiting five general practitioners in three medical practices in Amsterdam Main sample: n=110; sample test-retest: n=35 |
| Wolf et al (1978) | USA | MISS-26 | Patients from USA n=50 (mean age 36.6, 62% female) |
| Measures       | Respect for patients’ values, preferences, and expressed needs | Coordination and integration of care | Information, communication, and education | Physical comfort | Emotional support - relieving fear and anxiety | Involvement of family and friends | timeliness | efficiency | equity/accessibility | others     |
|---------------|---------------------------------------------------------------|-----------------------------------|----------------------------------------|-----------------|---------------------------------------------|---------------------------------|------------|------------|----------------------------|-----------|
| ACES          | Clinical team, interpersonal treatment, patient trust        | Visit-based coordination of care  | Communication, health promotion        |                 |                                             |                                 |            |            | Organizational access, office staff | Relation |
| CPCI          | Accumulated knowledge                                        | Coordination of care, continuity of care | Comprehensiveness of care, interpersonal communication |                 |                                             |                                 | First contact |            |                             | Longitud |
| CSQ           | Depth of relationship                                        | Depth of relationship             | Perceived time                         | Professional care |                                             |                                 |            |            | Genera satisfac        |          |
| CSS-VF        | Choice, interpersonal care                                   | Continuity of care                | Communication of general practitioner and specialist | Competence of general practitioner and specialist | Access to primary care, access to secondary care, finances | Genera satisfac    |            |            |                             |          |
| EUROPEP       | Organization of care                                         | Information and support           | Technical aspects of care/competence   | Accessibility   |                                             | Relatio          |            |            |                             |          |
| G-MISS-16-VF  |                                                              | Communication, compliance intent  | Pain                                   |                 |                                             |                  |            |            |                             |          |
| GPAQ          | Continuity of care                                           | Communication                     | Medical care                           | Accessibility, office staff | Nurse |            |          |            |                             |          |
| GPAQ-R        | Continuity of care                                           | Communication                     | Medical care                           | Accessibility, office staff | Nurse |            |          |            |                             |          |
| GPAS          | Humanity, trust, accumulated knowledge                       | Continuity of care                | Communication                          | Technical aspects of care | Accessibility, appointments | Premis |            |            |                             |          |
| Grogan-PSQ-40 | Relationship                                                  |                                   |                                        |                 |                                             |                  |            |            | Accessibility, appointments | General    |
| Haddad-PSQ-22 | Relationship                                                  |                                   |                                        |                 |                                             |                  |            |            | Technical aspects of care, |          |
| Questionnaire | Domain | Outcomes |
|---------------|--------|----------|
| IPQ           | Genera satisfaction | Office staff |
| Marshall-PSQ-18 | Relationship | Technical aspects of care |
| MISS-21       | Rapports | Finances, accessibility |
| MISS-26       | Affective satisfaction | General satisfaction |
| MISS-29       | Rapports | General satisfaction |
| NMPSSA        | Rapports | Accessibility, patient reception |
| NMPSSIAC      | Listening | General satisfaction |
| PCAS          | Interpersonal treatment, trust, knowledge of the patient | Physical examination |
| PCAT          | First contact/using | First contact/accessibility, services available |
| PDIS          | Patient-doctor interaction | General feeling |
| PDRQ-9        | Patient-doctor relationship | General satisfaction |
| PEI           | Communication | Emotions |
| PEQ           | Communication | Outcomes |
| QVFP          | Doctor-patient relationship | Consultation outcomes |
| SSQ           | Continuity of care | Medical care |
|                | VSQ-VF                        | Vukovic-PSQ-20                       | Ware-PSQ-55         |
|----------------|------------------------------|-------------------------------------|-------------------|
| Premis         | Global satisfac              | Global satisfac                     |                   |
| Technical quality | Accessibility to care, availability of resources | Overall satisfac |
| Measure      | Authors (Year) | Articles(s) | IRT or CTT | Score A B C D E F G H I |  |
|--------------|----------------|-------------|------------|-------------------------|---|
| ACES         | Safran et al (2005) | (26) | CTT | 0 |  |
| CPCl         | Flocke et al (1997) | (27,28) | CTT | ++ (27) | + (27) |  |
| CSQ          | Baker et al (1990) | (29-31) | CTT | ++ (30) | ++ (29) | 0 (29) |  |
| CSS-VF       | Gasquet et al (2003) | (32) | CTT | + | 0 | + | 0 |  |
| EUROPEP      | Comité EUROPEP (1998) | (33) | CTT | 0 |  |
| G-MISS-16-VF | Maurice-Szamburski et al (2017) | (34) | IRT | ++ | +++ | +++ | 0 |  |
| GPAQ         | Mead et al (2008) | (35) | CTT | +++ |  |
| GPAQ-R       | Roland et al (2013) | (36) | CTT | +++ | +++ |  |
| GPAS         | Ramsay et al (2000) | (37) | CTT | +++ | ++ | 0 |  |
| Grogan-PSQ-40| Grogan et al (1995) | (38) | CTT | 0 | 0 | 0 |  |
| Haddin-PSQ-22| Haddad et al (2000) | (39) | CTT | +++ | ++ | 0 |  |
| IPO          | Greco et al (2003) | (40) | CTT | 0 | 0 |  |
| Marshall-PSQ-18| Marshall et al (1994) | (41) | CTT | +++ |  |
| MISS-21      | Meakin et al (2002) | (42) | CTT | 0 | + |  |
| MISS-26      | Wolf et al (1981) | (31) | CTT | 0 |  |
| MISS-29      | Wolf et al (1978) | (43) | CTT | 0 |  |
| NMPSSA       | Eccles et al (1992) | (44) | CTT | 0 | + |  |
| NMPSSIAC     | Eccles et al (1992) | (44) | CTT | 0 | + |  |
| PCAS         | Safran et al (1998) | (28,45) | CTT | ++ (45) | ++ (45) | +++ (45) |  |
| PCAT         | Shi et al (2001) | (46) | CTT | +++ |  |
| PDIS         | Bowman et al (1992) | (47) | CTT | 0 | + | +++ | 0 |  |
| PDRQ-9       | Van der Feltz-Cornelis et al (2004) | (48) | CTT | 0 | ++ | 0 |  |
| PEI          | Howie et al (1998) | (49) | CTT | 0 |  |
| PEQ          | Steine et al (2001) | (50) | CTT | 0 | +++ |  |
| QVFP         | Marcinowicz et al (2010) | (51) | CTT | + | 0 | +++ |  |
| SSQ          | Baker et al (1991) | (30,52) | CTT | 0 (52) | ++ (30) | 0 (30) | 0 (52) |  |
| VSQ-VF       | Gasquet et al (2003) | (32) | CTT | + | 0 | + | 0 |  |
| Vukovic-PSQ-20| Vukovic et al (2012) | (53) | CTT | 0 | 0 | +++ |  |
| Ware-PSQ-55  | Ware et al (1983) | (54) | CTT | +++ | 0 | 0 | +++ |  |

COSMIN psychometric property boxes: A = internal consistency, B = reliability, C = measurement error, D = content validity, E = structural validity, F = hypothesis testing, G = cross-cultural validity, H = criterion validity, I = responsiveness. 4-point scale rating: +++ = excellent, ++ = good, + = fair, 0 = poor, empty space = COSMIN rating not applicable. IRT = Item response theory; CTT = Classical test theory.
| Instrument      | Articles(s) | Internal consistency | Reliability | Measurement error/Agreement | Content validity | Structural validity/Item response theory (IRT) | Hypothesis testing | Cross-cultural validity | Criterion validity | Responsiveness |
|-----------------|-------------|----------------------|-------------|----------------------------|------------------|-----------------------------------------------|------------------|-------------------------|-------------------|-----------------|
| ACES            | (26)        |                      |             |                            |                  |                                               |                  |                         |                   |                 |
| CPCI            | (27,28)     | -                    |             |                            |                  |                                               |                  |                         |                   |                 |
| CSQ             | (29-31)     | +/-                  |             |                            |                  |                                               |                  |                         |                   |                 |
| CSS-VF          | (32)        | -                    |             |                            |                  |                                               |                  |                         |                   |                 |
| EUROPEP         | (33)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| G-MISS-16-VF    | (34)        | ++                   |             |                            |                  | IRT: ++                                        |                  |                         |                   |                 |
| GPAQ            | (35)        | -                    |             |                            |                  |                                               |                  |                         |                   |                 |
| GPAQ-R          | (36)        | -                    |             |                            |                  |                                               |                  |                         |                   |                 |
| GPAS            | (37)        | -                    |             |                            |                  |                                               |                  |                         |                   |                 |
| Grogan-PSQ-40   | (38)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| Haddad-PSQ-22   | (39)        | ++                   | +           |                            |                  |                                               |                  |                         |                   |                 |
| IPQ             | (40)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| Marshall-PSQ-18 | (41)        | ++                   |             |                            |                  |                                               |                  |                         |                   |                 |
| MISS-21         | (42)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| MISS-26         | (31)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| MISS-29         | (43)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| NMPSSA          | (44)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| NMPSSIAC        | (44)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| PCAS            | (28,45)     | +/-                  |             |                            | +/-             |                                               | +/-             |                         |                   |                 |
| PCAT            | (46)        | -                    |             |                            |                  |                                               |                  |                         |                   |                 |
| PDIS            | (47)        | ?                    | -           |                            |                  |                                               |                  |                         |                   |                 |
| PDRQ-9          | (48)        | ?                    | -           |                            |                  |                                               |                  |                         |                   |                 |
| PEI             | (49)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| PEQ             | (50)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| QVFP            | (51)        | +                    |             |                            |                  |                                               |                  |                         |                   |                 |
| SSQ             | (30,52)     | ?                    | ++          |                            |                  |                                               |                  |                         |                   |                 |
| VSQ-VF          | (32)        | +                    |             |                            |                  |                                               |                  |                         |                   |                 |
| Vukovic-PSQ-20  | (53)        | ?                    |             |                            |                  |                                               |                  |                         |                   |                 |
| Ware-PSQ-55     | (54)        | +                    |             |                            |                  |                                               |                  |                         |                   |                 |

*S = result of best-evidence synthesis. Empty cell = no information available. Rating: +++/-- Strong level of evidence for positive/negative results, ++/- Moderate level of evidence for positive/negative results, +/- Limited evidence for positive/negative results, +/- Conflicting evidence, ? = Unknown, due to poor methodological quality, empty cell = No synthesis possible due to a lack of validation studies for this measurement property.

# = Number of studies on which the best-evidence synthesis regarding the measurement property was based.
Figure 1

Flow diagram of article selection process.
Percentage of instruments discovered domains defined by IOM

Figure 2
Percentage of tools which covered the domains defined by IOM.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- PRISMAchecklist.doc