



A systematic review of the measurement properties of the European Organisation for Research and Treatment of Cancer In-patient Satisfaction with Care Questionnaire, the EORTC IN-PATSAT32

Koen I. Neijenhuijs^{1,2} · Femke Jansen^{2,3} · Neil K. Aaronson⁴ · Anne Brédart⁵ · Mogens Groenvold^{6,7} · Bernhard Holzner⁸ · Caroline B. Terwee⁹ · Pim Cuijpers¹ · Irma M. Verdonck-de Leeuw^{1,2,3}

Received: 7 December 2017 / Accepted: 30 April 2018 / Published online: 7 May 2018

© The Author(s) 2018

Abstract

Purpose The EORTC IN-PATSAT32 is a patient-reported outcome measure (PROM) to assess cancer patients' satisfaction with in-patient health care. The aim of this study was to investigate whether the initial good measurement properties of the IN-PATSAT32 are confirmed in new studies.

Methods Within the scope of a larger systematic review study (Prospero ID 42017057237), a systematic search was performed of Embase, Medline, PsycINFO, and Web of Science for studies that investigated measurement properties of the IN-PATSAT32 up to July 2017. Study quality was assessed, data were extracted, and synthesized according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology.

Results Nine studies were included in this review. The evidence on reliability and construct validity were rated as sufficient and of the quality of the evidence as moderate. The evidence on structural validity was rated as insufficient and of low quality. The evidence on internal consistency was indeterminate. Measurement error, responsiveness, criterion validity, and cross-cultural validity were not reported in the included studies. Measurement error could be calculated for two studies and was judged indeterminate.

Conclusion In summary, the IN-PATSAT32 performs as expected with respect to reliability and construct validity. No firm conclusions can be made yet whether the IN-PATSAT32 also performs as well with respect to structural validity and internal consistency. Further research on these measurement properties of the PROM is therefore needed as well as on measurement error, responsiveness, criterion validity, and cross-cultural validity. For future studies, it is recommended to take the COSMIN methodology into account.

Keywords Cancer · Patient satisfaction · Systematic review · Psychometry · Validity · Reliability · Patient-reported outcome measure · PROM · EORTC · IN-PATSAT32

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00520-018-4243-9>) contains supplementary material, which is available to authorized users.

✉ Irma M. Verdonck-de Leeuw
IM.Verdonck@vumc.nl

¹ Department of Clinical, Neuro- and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

² Cancer Center Amsterdam, Amsterdam, The Netherlands

³ Department of Otolaryngology-Head and Neck Surgery, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands

⁴ Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

⁵ Psycho-Oncology Unit, Institut Curie, Paris, France

⁶ The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen University Hospital, Copenhagen, NV, Denmark

⁷ Department of Public Health, University of Copenhagen, Copenhagen, Denmark

⁸ Department of Psychiatry, Psychotherapy and Psychosomatics, CL-Service, Medical University of Innsbruck, Innsbruck, Austria

⁹ Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, The Netherlands

Introduction

The evaluation of patient health care experiences is relevant for improving health care [1] and as a patient-reported outcome measure (PROM) in clinical cancer trials [2]. While multiple PROMs are available to measure patient satisfaction with care [3–7], these PROMs lack international validations [8]. To assess patient satisfaction with health care and to enable cross-cultural comparison of patient health care experiences, the Quality of Life Group of the European Organisation for Research and Treatment of Cancer (EORTC) developed the IN-PATSAT32 [8].

The IN-PATSAT32 is a 32-item PROM assessing hospitalized cancer patients' satisfaction with care. It includes 11 multi-item scales designed to assess: doctors' technical skills (three items), nurses' technical skills (three items), doctors' interpersonal skills (three items), nurses' interpersonal skills (three items), doctors' information provision (three items), nurses' information provision (three items), doctors' availability (two items), and nurses' availability (two items), other hospital staff's interpersonal skills and information provision (three items), waiting time (two items), and hospital access (two items). Three single-item scales address the exchange of information, comfort, and general satisfaction.

The initial development and validation study of the IN-PATSAT32 was carried out in 647 patients from eight European countries and Taiwan and yielded good psychometric results [8]. Multitrait item scaling (MIS) indicated that the structure of the IN-PATSAT32 coincided in most part to the hypothesized structure of items and subscales. Internal consistency and test-retest reliability were satisfactory ($\alpha = .80-.96$; ICC = .70–.85, respectively). Multi- and single-item scales showed evidence of convergent validity when compared to the the Oberst Perception of Care Quality and Satisfaction Scale [4] and divergent validity with the EORTC QLQ-C30 [9]. Finally, validity was supported by the ability of the PROM to distinguish between patients with different levels of intention to recommend the hospital to others [8].

Over a decade after the initial development of the IN-PATSAT32, it is of interest to investigate whether these initial good results regarding the measurement properties of the IN-PATSAT32 are confirmed in other studies, to ensure that it performs as expected in diverse clinical and cultural settings. The aim of the current study was to perform a systematic review of the measurement properties of the IN-PATSAT32, as tested in individual validation studies. Evaluating measurement properties requires weighing many variables on both the level of the study and on the level of the PROM. Therefore, the current study used the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) criteria for assessing measurement properties of PROMs [10–13].

Methods

Literature search strategy

The literature search was part of a larger systematic review (Prospero ID 42017057237 [14]) investigating the validity of 39 different PROMs measuring quality of life of cancer survivors included in an eHealth application called “Oncokompas” (Amsterdam, the Netherlands) [15–18]. The databases Embase, Medline, PsycINFO, and Web of Science were searched for publications that investigated measurement properties of these 39 PROMs including the EORTC IN-PATSAT32. The search terms were the PROM's name, combined with search terms for cancer, and a precise filter for measurement properties [19]. The first search was performed in July 2016. The full search terms can be found in Appendix A. An additional search (up to July 2017) was performed using the same search terms, and a subsequent manual search in Google Scholar and Pubmed for missing records, to search for recent studies.

Inclusion and exclusion criteria

Studies were included that reported original data on cancer patients, and on at least one of the following measurement properties of the IN-PATSAT32 as defined by the COSMIN taxonomy [10–13]: internal consistency, reliability, measurement error, structural validity, hypothesis testing (for construct validity), criterion validity, cross-cultural validity, and responsiveness. Validation studies on other PROMs, which also reported original data on the IN-PATSAT32, were included. Studies that were only available as abstracts or conference proceedings were excluded, as well as non-English publications. Titles and abstracts, and the selected full-texts were reviewed by two independent raters (KN and FJ). Disagreements were discussed until verbal agreement on consensus.

Data extraction

Two independent researchers (KN and FJ) extracted information from eligible papers on each of the measurement properties defined by the COSMIN taxonomy [10]. Relevant data included the type of measurement property, its outcome, and information on methodology. Disagreements were discussed until verbal agreement on consensus.

Data synthesis

Data synthesis consisted of three steps. First, the quality of the methodology of the included studies was rated using the 4-point scoring system of the COSMIN checklist [10–12]. Methodological aspects regarding design requirements and

preferred statistical methods, specific to each measurement property under consideration, were rated as either “poor”, “fair”, “good”, or “excellent”. The methodological quality was operationalized per measurement property per study as the lowest score they received on any of the methodological aspects. The final ratings can be found in Appendix B.

Second, criteria for good measurement properties were applied to the results of the included studies, following the COSMIN guidelines for systematic reviews of PROMs [13]. Each measurement property in each individual study was rated as sufficient (+), insufficient (–), or indeterminate (?), according to predefined criteria. For indeterminate ratings, the methodological rating was non-applicable. All of these ratings were qualitatively summarized to determine the overall rating of the measurement property. If all studies indicated a sufficient, insufficient, or indeterminate rating for a specific measurement property, the overall rating of this measurement property was accordingly. If there were inconsistencies between studies, explanations were explored (e.g., differences in methodological quality). If explanations were found, they were discussed until consensus was reached and taken into account during interpretation. If no explanations were found, the overall rating would be inconsistent (\pm).

Third, we used the modified GRADE approach [13] to rate the quality of the evidence available for the measurement properties of the IN-PATSAT32. This approach takes into account (i) methodological quality, (ii) directness of evidence, (iii) inconsistency of results, and (iv) precision of evidence. The overall quality of evidence was rated as high, moderate, low, or very low. Measurement properties that were rated as indeterminate in the previous step did not receive a rating as there was no evidence to rate. All ratings (methodological quality, measurement property rating, and GRADE rating) were rated by one researcher (KN), whose ratings were checked by a second independent researcher (AH). Discrepancies in ratings were discussed until verbal agreement on consensus.

Results

Search results

The initial search identified 980 abstracts of which 10 were relevant to the IN-PATSAT32 (Fig. 1). Three abstracts and one full-text were excluded for not providing unique information on a measurement property. One study not captured by the search, but known to the authors, was added before data extraction. The search update up to July 2017 identified three more abstracts of which one was excluded for not providing unique information on a measurement property. No full texts were excluded from this search update. In total, nine studies were included in this review (see Supplementary Table 1).

These nine studies reported on the structural validity (six studies), internal consistency (five studies), reliability (two studies), and hypothesis testing (six studies) of the IN-PATSAT32, but lacked information on measurement error, criterion validity, responsiveness, and cross-cultural validity. We were able to calculate measurement error for two studies.

Structural validity

Six studies reported on structural validity. Methodological quality of these studies was rated as “good” [20], “fair” [21], or “poor” [22–25] (Table 1). The poor ratings were due to using Multitrait Item Scaling (MIS) instead of confirmatory or exploratory factor analysis (CFA/EFA). The fair score was due to lack of information about the handling of missing values. Results of the MIS analyses were consistent across studies, as well as with the original validation study. However, MIS is an indirect way of testing structural validity. Therefore, no conclusions can be drawn on basis of these studies. Two articles [20, 21] presented results of principal component analyses (PCA). Hjörleifsdóttir [20], of “good” quality, extracted four components with an eigenvalue > 1 , with a balanced distribution of explained variance. Pishkuhi [21], of “fair” quality, extracted five components with an eigenvalue > 1 , and one of those components explained most of the variance. The factor structures found in these two studies were inconsistent with the 11 subscale (and three single-item scales) model as reported in the initial study [8], leading to an insufficient rating.

Internal consistency

Five studies reported on internal consistency of the IN-PATSAT32, and their methodological quality was rated as “good” [20], “fair” [21], or “poor” [22–25]. The main reason for the poor ratings was that the unidimensionality of the scales was not tested appropriately. The values for Cronbach’s alpha of five studies [21–25] are presented in Supplementary Table 2. One other study [20] presented Cronbach’s alpha values for scales they had established: nurse satisfaction ($\alpha = .95$), doctor satisfaction ($\alpha = .93$), information satisfaction ($\alpha = .91$), and service satisfaction ($\alpha = .67$). However, as these scales do not represent the subscales recommended for this questionnaire [8], this study is not included in Supplementary Table 2, nor further taken into account. All but one subscale (hospital access) showed Cronbach’s alpha values that would qualify for a sufficient rating. However, as none of the studies provided any evidence of unidimensionality for the subscales, Cronbach’s alpha cannot be properly interpreted [26]. The inconsistency of Cronbach’s alpha coefficients across studies is noteworthy for the subscale hospital access ($\alpha = .36–.86$).

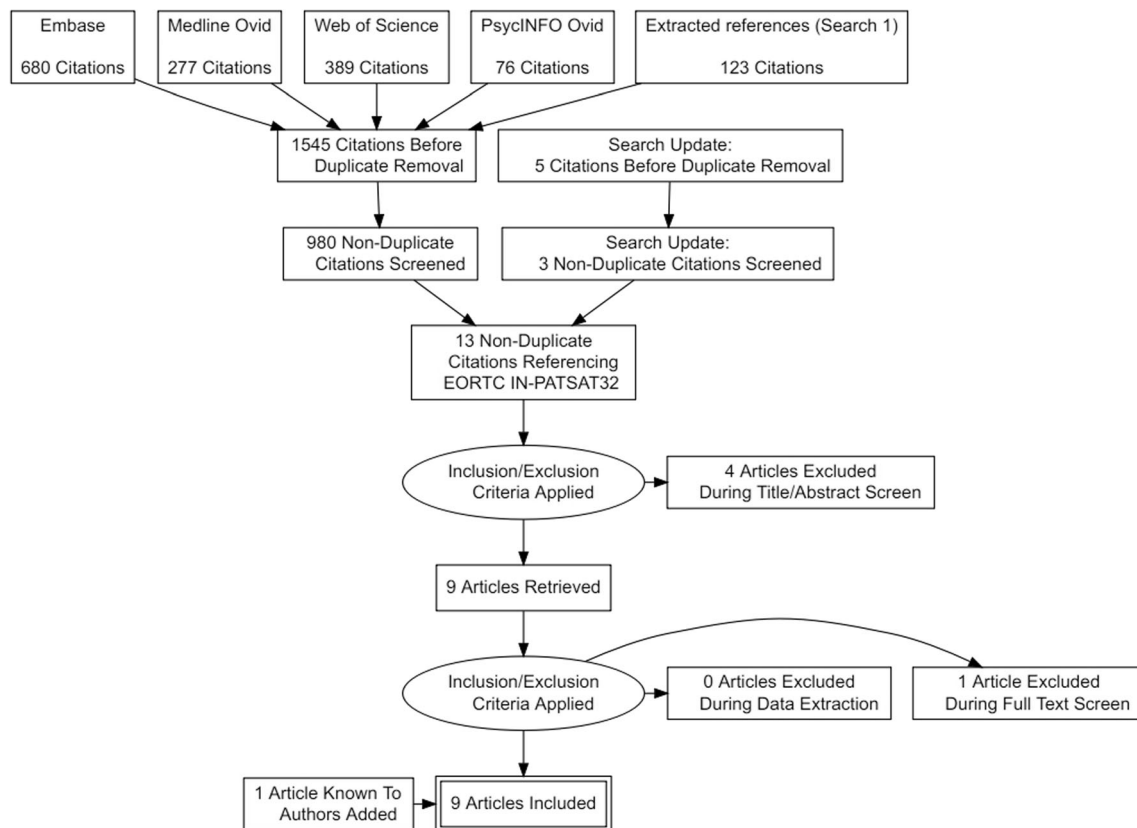


Fig. 1 PRISMA diagram

Reliability

Two studies [21, 25] reported on test-retest reliability (see Supplementary Table 3). Methodological qualities were rated as “fair” due to lack of information about the handling of missing values [21, 25], not reporting the type of correlation coefficient [21], and a short time interval (30 min) [25]. One study [21] showed high test-retest correlations ($r > .85$), leading to a sufficient rating on test-retest reliability. However, as the type of correlation coefficient was not reported, it is unclear whether these values represent appropriate estimates of test-retest reliability [27, 28]. The other study [25] showed acceptable test-retest correlations ($ICC > .70$), except for doctors’ availability ($ICC = .64$) and general comfort ($ICC = .67$), leading to a sufficient rating.

Measurement error

While none of the studies presented results regarding measurement error, the standard error of measurement (SEM) and smallest detectable change (SDC) could be calculated for the two studies reporting test-retest reliability [21, 25]. Methodological quality was “good”, due to the need to calculate measurement error indirectly (Table 2). Since no minimal important change (MIC) was reported, a criterion for

good measurement error could not be applied. While there is no evidence for or against good measurement error, the SDC could be compared to the maximum range of the subscales. The SDC represents the minimum change score over time of which we can be certain does not represent measurement error. Most SDC scores were between 20 and 30, representing 20–30% on the 100-point scale. There were a few notable outliers: Doctor availability (29.17–46.40), waiting time (25.05–44.70), and hospital access (29.39–34.48).

Construct validity (hypothesis testing)

Known-group comparison

Three studies performed known-group comparison, a comparison between groups that are known to show differences on the measured construct. Known group differences were investigated with respect to age [23], educational level [23], tumor stage [24], time since diagnosis [24], and satisfaction with care [22]. The methodological quality of these studies was rated as “fair” [22] or “poor” [23, 24]. The poor scores were due to not providing a priori hypotheses, while the fair score was due to lack of information about the handling of missing values (Table 3). The

Table 1 Structural validity of the IN-PATSAT32

Reference	Methodology	Outcome	Rating structural validity	Quality
Arraras et al., 2009 [22]	Multitrait Item Scaling	Most items exceeded correlations of .4 with other items in their own scale, except for items 29 and 30 (Hospital Access). Most items had a higher correlation with other items in their own scale than items in other scales, except for items 14 (Nurse Interpersonal Skills), 21, 22 (Nurse Availability), 24 (Other Staff Interpersonal Skills), and 30 (Hospital Access).	Indeterminate	Poor
Hjörleifsdóttir et al., 2010 [20]	Multitrait Item Scaling	All items exceeded correlations of .4 with other items in their own scale. The weakest scale was ‘satisfaction with service and care organization’, in which 50% of the items correlated higher with other items in their own scale than other items in other scales. The strongest scale was ‘satisfaction with nurses’ conduct’, in which 92% of items correlated higher with other items in their own scale than other items in other scales.	Indeterminate	n/a
Hjörleifsdóttir et al., 2010 [20]	Principal Component Analysis	Four components were extracted with an eigenvalue > 1, explaining 67.4% of variance. The components can be identified as: Satisfaction with nurses (24.7% variance), satisfaction with doctors (21% variance), satisfaction with information (13.6% variance), and satisfaction with service (8% variance).	Insufficient	Good
Obtel et al., 2017 [25]	Multitrait Item Scaling	All items exceeded correlations of .4 with other items in their own scale. All items had higher correlations with other items in their own scale than items in other scales.	Indeterminate	Poor
Pishkuhi et al., 2014 [21]	Multitrait Item Scaling	All items exceeded correlations of .8 with other items in their own scale. All items had higher correlations with other items in their own scale than items in other scales.	Indeterminate	n/a
Pishkuhi et al., 2014 [21]	Principal Component Analysis	Five components were extracted with an eigenvalue > 1, explaining 71.1% of variance. The components can be identified as: Satisfaction with nurses (45.4% variance), satisfaction with services and care organization (9.5% variance), satisfaction with doctors (8.1% variance), satisfaction with doctors’ information provision (4.7% variance), and satisfaction with nurses’ information provision (3.2% variance).	Insufficient	Fair
Zhang et al., 2014 [23]	Multitrait Item Scaling	All items exceeded correlations of .4 with other items in their own scale. Fifty percent of items had a higher correlation with other items in their own scale than items in other scales.	Indeterminate	Poor
Zhang et al., 2015 [24]	Multitrait Item Scaling	All items exceeded correlations of .4 with other items in their own scale. Six out of 29 items had a significantly lower correlation with items in their own scale than items in other scales.	Indeterminate	Poor

Table 2 Measurement error (standard error of measurement and smallest detectable change) of the IN-PATSAT32

Reference	DrInt		NInt		Nava	SInt	WT	HA	IE	HC	OA	Rating	Quality		
	DrTech	DrInfo	DrAva	NTech										NInfo	
Pishkuhi et al., 2014 [21]												?	Fair		
	SEM	9.53	10.82	7.99	10.53	7.88	6.47	6.05	8.67	6.80	9.03	10.60			
	SDC	26.42	29.98	22.13	29.17	21.85	17.93	16.77	24.58	18.85	25.05	29.39			
Obtel et al., 2017 [25]													?	Fair	
	SEM	7.83	8.09	8.38	16.81	7.84	12.08	8.80	9.10	10.92	16.13	12.44	10.70	14.40	14.88
	SDC	21.69	22.41	23.23	46.60	21.74	33.49	24.38	25.22	30.26	44.70	34.48	29.66	39.93	41.24

DrTech doctor technical skills, *DrInt* doctor interpersonal skills, *DrInfo* doctor information provision, *DrAva* doctor availability, *NTech* nurse technical skills, *NInt* nurse interpersonal skills, *NInfo* nurse information provision, *Nava* nurse availability, *SInt* other staff interpersonal skills, *WT* wait times, *HA* hospital access, *IE* information exchange, *HC* hospital comfort, *OA* overall satisfaction, ? = Indeterminate

known-group comparisons investigated by Arraras [22] were based on a priori hypotheses and provide sufficient evidence of construct validity. Due to not providing a priori hypotheses, the results of Zhang [23, 24] were rated as indeterminate.

Convergent validity

Four studies reported on convergent validity and compared the IN-PATSAT32 to the EORTC QLQ-INFO25 (measuring patient perceptions of information received and their

Table 3 Known-group validity of the IN-PATSAT32

Reference	Comparison groups	Outcome	Rating	Quality
Arraras et al., 2009 [22]	Low vs. high score on the Oberst perception of care quality and satisfaction scale	Significant differences in all IN-PATSAT32 areas except nurse availability. Patients with higher Oberst scores had greater care satisfaction.	Sufficient	Fair
Arraras et al., 2009 [22]	Low vs. high score on item investigating intention to recommend the hospital or ward to others	Significant differences in all IN-PATSAT32 areas except nurse availability. Patients with higher intention to recommend the hospital or ward had greater care satisfaction.	Sufficient	Fair
Zhang et al., 2014 [23]	Patients < 58 years vs. patients ≥ 58 years	Patients < 58 years scored significantly higher than patients ≥ 58 years, except on nurse availability and hospital comfort	Indeterminate	Poor
Zhang et al., 2014 [23]	Patients who finished lower than compulsory education vs. patients who finished compulsory or higher education	Patients who had finished compulsory education scored significantly higher than patients who had not finished compulsory education.	Indeterminate	Poor
Zhang et al., 2015 [24]	Patients who finished lower than compulsory education vs. patients who finished compulsory or higher education	Patients who had finished compulsory education scored significantly higher on technical skills, interpersonal skills, information provision, and availability of both doctors and nurses. Effect sizes were small (< .50) in for all scales.	Indeterminate	Poor
Zhang et al., 2015 [24]	Patients with metastatic vs. non-metastatic tumors	Patients with metastatic tumors scored significantly higher on nurses' conduct, other hospital staffs' interpersonal skills information provision scales. Effect sizes were small (< .50) except for nurses' interpersonal skills (- .55), nurses' information provision (- .57), and nurses' availability (- .51).	Indeterminate	Poor
Zhang et al., 2015 [24]	Patients with > 2 months diagnostic time vs. patients with < 2 months diagnostic time	Patients with > 2 months diagnostic time scored significantly higher on nurses' conduct, other hospital staffs' interpersonal skills information provision scales. Effect sizes were small (< .5) except for nurses' technical skills (- .55), and nurses' interpersonal skills (- .50).	Indeterminate	Poor

Table 4 Convergent validity of the IN-PATSAT32

Reference	Comparison instrument	Correlations	Rating	Quality
Aboshaiqah et al., 2016 [31]	EORTC QLQ-C15-PAL	IN-PATSAT32 general satisfaction correlated with physical function ($r = .21$), emotional function ($r = .32$), and global health status ($r = .26$).	Insufficient	Poor
Arraras et al., 2009 [22]	Oberst patients' perception of care quality and satisfaction scale	Oberst medical care scale correlated with the IN-PATSAT32 doctor scales (.62–.71). The Oberst information adequacy scale correlated with the IN-PATSAT32 doctor information provision (.70) and nurses' information provision (.62) scales. The Oberst quality of nursing scale correlated with the IN-PATSAT32 nurse scales (.60–.69). The Oberst self-care information scale correlated with doctors' (.60) and nurses' (.61) information provision.	Sufficient	Fair
Arraras et al., 2010 [32]	EORTC QLQ-INFO25	Doctors' information provision (.61), nurses' information provision (.46), other staff interpersonal skills (.42) correlated with the QLQ-INFO25 item regarding information satisfaction. Single items regarding information provision of the IN-PATSAT32 correlated with QLQ-INFO25 items measuring similar constructs (.30–.61), with more similar theoretical items correlating higher ($> .40$).	Sufficient	Fair
Asadi-lari et al., 2015 [29]	EORTC QLQ-INFO25	Doctors' information provision (.23), nurses' information provision (.39), and other staff interpersonal skills (.20) correlated with the QLQ-INFO25 item regarding information satisfaction. Single items regarding information provision of the IN-PATSAT32 correlated with the QLQ-INFO25 items measuring similar constructs (.15–.41).	Insufficient	Good

information needs) [29], the Oberst patients' perception (measuring the quality of care received and how well the care meets patients' expectations [4]) [22], and the EORTC QLQ-C15-PAL (measuring quality of life of patients with incurable cancer [30]) [31]. The methodological quality of these studies was rated as either "good" [29], "fair" [22, 32], or "poor" [31]. The poor score was due to not providing a priori hypotheses [31]. The fair scores were due to lack of information about the handling of missing values [32], or due to lack of information about a priori hypotheses [22] (Table 4). Two studies [22, 32], of "fair" quality, demonstrated moderate correlations ($r > .40$) with related constructs, indicative of sufficient convergent validity. Asadi-lari [29], of "good" quality, and Aboshaiqah [31], of "poor" quality, found low correlations ($r < .40$) for most of the constructs that were hypothesized to be related to the IN-PATSAT32, indicating insufficient convergent validity.

Divergent validity

Four studies reported on divergent validity and compared the IN-PATSAT32 scales to scales of the EORTC QLQ-C30 (measuring health-related quality of life in cancer patients [9]). Their methodological quality was rated as "fair" [21, 22] or "poor" [23, 24]. The poor scores were due to not providing a priori hypotheses. The fair score of Arraras [22] was due to the lack of detail in formulated a priori hypotheses, while the fair score of Pishkuhi [21] was due to lack of information about the handling of missing values. One study of "fair" quality found no significant correlations [21], and one study of "fair" quality [22] and two studies of "poor" quality [23, 24]

found correlations smaller than .40, indicative of sufficient divergent validity.

Data synthesis

The synthesized ratings of the measurement properties can be found in Table 5. Internal consistency was rated indeterminate as no tests of unidimensionality were reported. Measurement error was rated indeterminate as no MIC was reported and could not be calculated with the available data. Structural validity was rated insufficient with evidence of low quality. Test-retest reliability and construct validity (hypothesis testing) were judged to be sufficient, both with evidence of moderate quality. The indeterminate findings [23, 24] for construct validity were not taken into account in this synthesis, as they did not provide evidence for or against construct validity. Studies of "poor" quality were outweighed by studies with better quality. One study of "good" quality provided insufficient evidence on convergent validity for construct validity [29], while three studies of "fair" quality provided sufficient evidence on known-group comparison and convergent validity for construct validity [21, 22, 32].

Discussion

This systematic review investigated the current evidence up to July 2017 regarding the measurement properties of the EORTC IN-PATSAT32 [8]. Nine studies were included in this review. The evidence on reliability and construct validity were

Table 5 Ratings of measurement properties

Measurement property	Rating of measurement property	Quality of evidence
Structural Validity	Insufficient	Low
Internal Consistency	Indeterminate	
Reliability	Sufficient	Moderate
Measurement Error	Indeterminate	
Construct Validity	Sufficient	Moderate

rated as sufficient and of moderate quality evidence. The evidence on structural validity was rated as insufficient and of low quality. The evidence on internal consistency was indeterminate, as the assumption of unidimensionality was not investigated. Measurement error, responsiveness, criterion validity, and cross-cultural validity were not reported in the studies reviewed.

With respect to structural validity, the developers of the IN-PATSAT32 postulated an a priori scale structure and provided support for that structure in their original validation study [8]. In the studies that reported on structural validity [20–25], MIS or PCA was applied instead of CFA. The findings of the PCA analyses [20, 21] are of particular interest as they revealed fewer scales compared to the original 11-scale (and three separate single-item) factor structure [8].

Future studies investigating structural validity may inform their theorized factor structures based on these results. They may consider performing CFAs to test the posited 11-scale structure, but also two factor structures which seem plausible, given the results of the reported PCAs [20, 21]:

1. A first-order factor structure where the relevant items load on one of four factors: (i) satisfaction with nurses; (ii) satisfaction with doctors; (iii) satisfaction with services and care; and (iv) information provision;
2. A second-order factor structure where all items load on the originally developed scales. The originally developed scales will then load on the relevant second-order factors: (i) satisfaction with nurses; (ii) satisfaction with doctors; (iii) satisfaction with services and care; and (iv) information provision.

Test-retest reliability was rated as sufficient in the present review although of moderate quality evidence. When this property is examined in future studies, it is important that the intraclass correlation coefficient is used to control for systematic error variance. Without controlling for systematic error variance, test-retest reliability may be overestimated [27, 28].

In the present review, none of the studies reported on *measurement error*. We calculated the standard error of measurement (SEM) and smallest detectable change (SDC) based on the data of two studies. Relating the SDC to the maximum

range of the scale showed that most values were around 20–30% of the scales, although a number of outliers were observed. To interpret these data, information on the minimal important change (MIC) is needed. This should preferably be derived from anchor-based methods. Subsequently, the MIC should be compared to the *measurement error* to determine if the scales can detect small but important changes that are not an artifact of *measurement error*.

Cross-cultural validity was explored in the original validation process [8]. In future studies, this can be investigated further by performing measurement invariance tests for subsamples in CFAs, or by pooling data of multiple international studies to perform measurement invariance tests for language. Unfortunately, it is not possible to assess *criterion validity*, as there is no “gold standard” for assessing patient satisfaction. *Responsiveness* could be investigated through longitudinal studies of changes in patient satisfaction with care.

A limitation of this review is the use of a precise rather than a sensitive search filter regarding measurement properties. The sensitivity of the precise filter was 93% in a random set of PubMed records, while the sensitivity of the sensitive search filter was 97% [19]. The use of the precise filter was a pragmatic choice over the available sensitive filter as the initial search encompassed 39 PROMs (including the IN-PATSAT32), and the sensitive filter would provide too many hits for feasible screening. Although we also performed a manual search and found no missing records, the possibility remains that the precise filter missed validation studies of the IN-PATSAT32. Furthermore, because we included only papers published in English, we may have missed information from studies published in other languages.

Based on this systematic review, we conclude that with respect to test-retest reliability and construct validity, the IN-PATSAT32 performs as expected in diverse clinical and cultural settings. However, no firm conclusions can be made as to whether the IN-PATSAT32 performs as well with respect to structural validity and internal consistency. Further research on these measurement properties of the EORTC IN-PATSAT32 is therefore needed as well as on measurement error, responsiveness, criterion validity, and cross-cultural validity. For future studies, it is recommended to take the COSMIN methodology into account.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Browne K, Roseman D, Shaller D, Edgman-Levitan S (2010) Analysis & commentary: measuring patient experience as a strategy for improving primary care. *Health Aff* 29(5):921–925. <https://doi.org/10.1377/hlthaff.2010.0238>
- Brédart A, Bottomley A (2002) Treatment satisfaction as an outcome measure in cancer clinical treatment trials. *Expert Rev Pharmacoecon Outcomes Res* 2:597–606. <https://doi.org/10.1586/14737167.2.6.597>
- Ware E, Snyder M, Wright R, Davies A (1983) Defining and measuring patient satisfaction with medical care. *Eval Program Plann* 6: 247–263. [https://doi.org/10.1016/0149-7189\(83\)90005-8](https://doi.org/10.1016/0149-7189(83)90005-8)
- Oberst MT (1984) Patients' Perceptions of care. *Cancer* 53:2366–2375
- Baker R (1990) Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *Br J Gen Pract* 40:487–490
- Rubin H, Ware H, Nelson E, Meterko M (1990) The patient judgments of hospital quality (PJHQ) questionnaire. *Med Care* 28:S17–S18
- Hargraves JL, Hays RD, Cleary PD (2003) Psychometric properties of the consumer assessment of health plans study (CAHPS) 2.0 adult core survey. *Health Serv Res* 38:1509–1528. <https://doi.org/10.1111/j.1475-6773.2003.00190.x>
- Brédart A, Bottomley A, Blazeby JM, Conroy T, Coens C, D'Haese S et al (2005) An international prospective study of the EORTC cancer in-patient satisfaction with care measure (EORTC IN-PATSAT32). *Eur J Cancer* 41:2120–2131. <https://doi.org/10.1016/j.ejca.2005.04.041>
- Bjoridal K, De Graeff A, Fayers PM, Hammerlid E, Van Pottelsberghe C, Curran D et al (2000) A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. *Eur J Cancer* 36:1796–1807. [https://doi.org/10.1016/S0959-8049\(00\)00186-6](https://doi.org/10.1016/S0959-8049(00)00186-6)
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW (2010) 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* 19:539–549. <https://doi.org/10.1007/s11136-010-9606-8>
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al (2012) COSMIN checklist manual. Manual. VU University Medical Center, Amsterdam
- Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 21:651–657. <https://doi.org/10.1007/s11136-011-9960-1>
- Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW et al (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27:1–11. <https://doi.org/10.1007/s11136-018-1798-3>
- Neijenhuijs KI, Verdonck-de Leeuw IM, Cuijpers P, van der Hout A, Melissant HC, de Wit M, Jansen F, Veeger M (2017) Validity and reliability of patient reported outcomes measuring quality of life in cancer patients. PROSPERO: CRD42017057237. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42017057237. Accessed 27 Feb 2018
- van der Hout A, van Uden-Kraan CF, Witte BI, Veerle Coupé VMH, Jansen F, Leemans CR et al (2017) Efficacy, cost-utility, and reach of an eHealth self-management application 'Oncokompas' that facilitates cancer survivors to obtain optimal supportive care: study protocol for a randomized controlled trial. *Trials* 18:228. <https://doi.org/10.1186/s13063-017-1952-1>
- Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM (2015) Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. *J Clin Nurs* 24: 1367–1379. <https://doi.org/10.1111/jocn.12753>
- Jansen F, van Uden-Kraan CF, Van Zwieten V, Witte BI, Leemans CR, Verdonck-de Leeuw IM (2015) Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. *Support Care Cancer* 23:1679–1688. <https://doi.org/10.1007/s00520-014-2514-7>
- Duman-Lubberding S, van Uden-Kraan CF, Jansen F, Witte BI, van der Velden LA, Lacko M et al (2016) Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. *Support Care Cancer* 24:2163–2171. <https://doi.org/10.1007/s00520-015-3004-2>
- Terwee CB, Jansma EP, Riphagen II, De Vet HCW (2009) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 18:1115–1123. <https://doi.org/10.1007/s11136-009-9528-5>
- Hjörleifsdóttir E, Hallberg IR, Gunnarsdóttir ED (2010) Satisfaction with care in oncology outpatient clinics: psychometric characteristics of the Icelandic EORTC IN-PATSAT32 version. *J Clin Nurs* 19:1784–1794. <https://doi.org/10.1111/j.1365-2702.2009.03095.x>
- Pishkuhi MA, Salmaniyan S, Nedjat S, Zendedel K, Lari MA (2014) Psychometric properties of the Persian version of satisfaction with care EORTC-in-patsat32 questionnaire among Iranian cancer patients. *Asian Pac J Cancer Prev* 15:10121–10128. <https://doi.org/10.7314/APJCP.2014.15.23.10121>
- Arraras JI, Vera R, Martínez M, Hernández B, Láinez N, Rico M, Vila M, Chicata V, Asín G (2009) The EORTC cancer in-patient satisfaction with care questionnaire: EORTC IN-PATSAT32. *Clin Transl Oncol* 11:237–242. <https://doi.org/10.1007/s12094-009-0346-6>
- Zhang J, Xie S, Liu J, Sun W, Guo H, Hu Y et al (2014) Validation of EORTC IN-PATSAT32 for Chinese patients with gastrointestinal cancer. *Patient Prefer Adherence* 8:1285–1292. <https://doi.org/10.2147/ppa.s67111>
- Zhang L, Dai Z, Cheng S, Xie S, Woo SML, Luo Z, Wu J, Gao T, Liu J, Zhang K, Zhang J, Jia X, Miller AR, Wang C (2015) Validation of EORTC IN-PATSAT32 for Chinese cancer patients. *Support Care Cancer* 23:2721–2730. <https://doi.org/10.1007/s00520-015-2636-6>
- Obtel M, Serhier Z, Bendahhou K (2017) Validation of EORTC IN-PATSAT 32 in Morocco: Methods and Processes. *Asian-Pac J* 18: 1403–1409. <https://doi.org/10.22034/APJCP.2017.18.5.1403>
- Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 78:98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Bland M, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327: 307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)

28. Bland JM, Altman DG (2007) Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 17:571–582. <https://doi.org/10.1080/10543400701329422>
29. Asadi-lari M, Ahmadi Pishkuhi M, Almasi-Hashiani A, Safiri S, Sepidarkish M (2015) Validation study of the EORTC information questionnaire (EORTC QLQ-INFO25) in Iranian cancer patients. *Support Care Cancer* 23:1875–1882. <https://doi.org/10.1007/s00520-014-2510-y>
30. Groenvold M, Petersen MA, Aaronson NK, Arraras JI, Blazeby JM, Bottomley A, Fayers PM, de Graeff A, Hammerlid E, Kaasa S, Sprangers MA, Bjorner JB, EORTC Quality of Life Group (2006) The development of the EORTC QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *Eur J Cancer* 42:55–64. <https://doi.org/10.1016/j.ejca.2005.06.022>
31. Aboshaiqah A, Al-Saedi TSB, Abu-Al-Ruyhaylah MMM, Aloufi AA, Alharbi MO, Alharbi SSR et al (2016) Quality of life and satisfaction with care among palliative cancer patients in Saudi Arabia. *Palliat Support Care* 14:621–627. <https://doi.org/10.1017/S1478951516000432>
32. Arraras JI, Greimel E, Sezer O, Chie WC, Bergenmar M, Costantini A, Young T, Vlastic KK, Velikova G (2010) An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. *Eur J Cancer* 46:2726–2738. <https://doi.org/10.1016/j.ejca.2010.06.118>