



A systematic review of the measurement properties of the Body Image Scale (BIS) in cancer patients

Heleen C. Melissant^{1,2} · Koen I. Neijenhuijs^{1,2} · Femke Jansen^{2,3} · Neil K. Aaronson⁴ · Mogens Groenvold^{5,6} · Bernhard Holzner⁷ · Caroline B. Terwee⁸ · Cornelia F. van Uden-Kraan^{1,2} · Pim Cuijpers¹ · Irma M. Verdonck-de Leeuw^{1,2,3}

Received: 29 November 2017 / Accepted: 2 March 2018 / Published online: 12 March 2018
© The Author(s) 2018

Abstract

Introduction Body image is acknowledged as an important aspect of health-related quality of life in cancer patients. The Body Image Scale (BIS) is a patient-reported outcome measure (PROM) to evaluate body image in cancer patients. The aim of this study was to systematically review measurement properties of the BIS among cancer patients.

Methods A search in Embase, MEDLINE, PsycINFO, and Web of Science was performed to identify studies that investigated measurement properties of the BIS (Prospero ID 42017057237). Study quality was assessed (excellent, good, fair, poor), and data were extracted and analyzed according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology on structural validity, internal consistency, reliability, measurement error, hypothesis testing for construct validity, and responsiveness. Evidence was categorized into sufficient, insufficient, inconsistent, or indeterminate.

Results Nine studies were included. Evidence was sufficient for structural validity (one factor solution), internal consistency ($\alpha = 0.86–0.96$), and reliability ($r > 0.70$); indeterminate for measurement error (information on minimal important change lacked) and responsiveness (increasing body image disturbance in only one study); and inconsistent for hypothesis testing (conflicting results). Quality of the evidence was moderate to low. No studies reported on cross-cultural validity.

Conclusion The BIS is a PROM with good structural validity, internal consistency, and test-retest reliability, but good quality studies on the other measurement properties are needed to optimize evidence. It is recommended to include a wider variety of cancer diagnoses and treatment modalities in these future studies.

Keywords Cancer · Body image · PROM · Systematic review · Measurement properties · Body Image Scale

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00520-018-4145-x>) 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, Faculty of Behavioural and Movement Sciences, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, van der Boerhorststraat 1, 1081BT Amsterdam, The Netherlands

² Cancer Center Amsterdam (CCA), VU University Medical Center, De Boelelaan 1118, 1081HV Amsterdam, The Netherlands

³ Department of Otolaryngology-Head and Neck Surgery, Amsterdam Public Health Research Institute, VU University Medical Center, De Boelelaan 1117, 1081HV Amsterdam, The Netherlands

⁴ Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands

⁵ The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen University Hospital, 20D, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark

⁶ Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen, Denmark

⁷ Department of Psychiatry, Psychotherapy and Psychosomatics, Medical University of Innsbruck, Anichstraße 35, 6020 Innsbruck, Austria

⁸ Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, VU University Medical Center, De Boelelaan 1089a, 1081HV Amsterdam, The Netherlands

Introduction

Patients with cancer are often faced with invasive treatments, with a temporal or permanent impact on appearance. Cancer patients may have to deal for example with scars or amputated body parts following surgery, skin burns due to radiation therapy, or hair loss due to chemotherapy. These appearance changes can negatively affect body image. Body image is a multi-dimensional construct and comprises cognitive, behavioral, and affective aspects of appearance [1]. For instance, altered body appearance after cancer treatment can be accompanied with feelings of shame, negative self-esteem, or social avoidance [2, 3]. For some patients, negative aspects of body image are persistent and remain prevalent years after treatment [4, 5] and can negatively impact quality of life. Therefore, body image is considered to be an essential factor of health-related quality of life (HRQOL) in cancer patients [6, 7]. Monitoring HRQOL (including body image) in clinical practice is important to identify patients who may benefit from supportive care, and patient-reported outcome measures (PROMs) are often used for that purpose [8, 9].

The Body Image Scale (BIS) is a PROM developed to measure body image in all types of cancer patients. This is in contrast to other PROMs that aim to measure body image in non-cancer populations (e.g., Appearance Schemas Inventory-Revised (ASI-R)) [10] or in cancer patients with specific types of cancer or treatment (e.g., Breast Impact of Treatment Scale (BITS) in breast cancer patients, Sexual Adjustment and Body Image Scale (SABIS-g) in gynecologic cancer patients, and Body Image Screener for Cancer Reconstruction (BICR) for patients after breast reconstruction) [11–13]. The initial development and validation study of the BIS showed good measurement properties concerning internal consistency, known-group comparison and responsiveness among English-speaking breast cancer patients [14]. Since then, the BIS was validated in several other languages such as Dutch, Greek, and Portuguese [15–17] and across diverse cancer populations, e.g., in advanced cancer patients and colorectal cancer patients [18, 19]. Recently, Muzzatti et al. (2017) presented a review of PROMs measuring body image in cancer patients, including the BIS, and concluded that the measurement properties of these PROMs require more thorough investigation [20]. With respect to the BIS specifically, they concluded that the measurement properties were adequate, except for inconsistent results regarding structural validity and lacking evidence for criterion validity. However, not all measurement properties were taken into account (i.e., measurement error and responsiveness). Moreover, no guideline was used to interpret results, and the methodological quality of the extracted studies was not assessed. Therefore, the aim of this current study was to conduct a systematic review specifically focusing on the measurement properties of the BIS in cancer patients, following the

COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology.

The COSMIN methodology is based on taxonomy and definitions of measurement properties for PROMs [21] including content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. The current study will add important information to the previous review [20], which is of high importance when considering the use of the BIS in clinical trials and practice as well as for interpretation of BIS outcomes.

Methods

The Body Image Scale

The 10-item Body Image Scale was developed by Hopwood et al. in 2001 to measure affective, behavioral, and cognitive body image symptoms. Patients can indicate body image symptoms on a 4-point scale (0 “not at all” to 3 “very much”). The total score ranges from 0 to 30 and can be calculated by summing up the 10 items. A higher score means a higher level of body image disturbance [14].

Literature search strategy

This study was part of a larger systematic review (Prospero ID 42017057237) [22], investigating the validity of 39 PROMs measuring quality of life of cancer survivors included in an eHealth application called “Oncokompas” [23–25]. Before the actual search, a search for reviews and meta-analyses of the measurement properties of each of the 39 PROMs was performed. This search did not yield any relevant results for the BIS.

The databases Embase, MEDLINE, PsycINFO, and Web of Science were systematically searched for publications directly investigating aspects of measurement properties of the BIS. Search terms were the measurement instrument’s name and its acronym, combined with search terms (text words and key words) for cancer, and a precise filter for measurement properties (Appendix A) [26]. The search was performed in July 2016 and updated in July 2017 to verify new publications. Search results were checked for duplications.

Inclusion and exclusion criteria

Studies were included that reported on original data about at least one measurement property as defined in the COSMIN taxonomy [21] related to the BIS. Validation studies of other PROMs that reported original data on the BIS (as comparison instrument) were also included. The COSMIN taxonomy [21] distinguishes nine measurement properties for PROMs: (1)

structural validity (degree to which scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured), (2) internal consistency (degree of interrelatedness among items), (3) reliability (the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions), (4) measurement error (systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured), (5) hypothesis testing for construct validity (degree to which the scores are consistent with hypotheses on known-groups comparison, and on relations to scores of other PROMs (convergent and divergent validity)), (6) criterion validity (degree to which the scores are an adequate reflection of a gold standard), (7) responsiveness (the ability of a PROM to detect change over time in the construct to be measured), (8) cross-cultural validity (degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version), (9) content validity (degree to which the content of a PROM is an adequate reflection of the construct to be measured). In the present review study, we did not evaluate content validity because no protocol existed to evaluate this measurement property.

We excluded studies that were conference proceedings, studies without full-text available, publications in other languages than English, and studies that investigated populations without cancer. Full-text publications were reviewed by two independent raters (KN and FJ). Disagreements regarding inclusion and exclusion were discussed until consensus was reached.

Data extraction

Two independent extractors (KN and FJ) who identified eligible studies extracted information on each of the measurement properties defined in the COSMIN taxonomy [21]. Relevant data included the study population, sample size, the method, information on missing values, type of measurement property, and its outcome. Disagreements were discussed until consensus was reached.

Data analyses

Data analyses were performed in three steps to accomplish adequate interpretation of the results, following the COSMIN methodology [27].

First, we rated the methodological quality of the included studies, based on the COSMIN checklist for assessing the methodological quality of studies on measurement properties [28]. Methodological aspects regarding design requirements and preferred statistical methods, specific to the measurement properties under consideration were rated on a 4-point scale: “excellent,” “good,” “fair,” or “poor.” In accordance with COSMIN recommendations, overall methodological quality

per measurement property of the BIS was obtained by taking the lowest rating of any of the methodological aspects assessed [29].

Second, criteria for good measurement properties were applied to the results of the included studies, following the COSMIN guidelines for systematic reviews of PROMs [27, 30]. Each measurement property in each individual study was rated as “sufficient” (+), “insufficient” (−), or “indeterminate” (?). For example, hypothesis testing for construct validity is rated as “sufficient” if at least 75% of the results are in accordance with the hypotheses. These results were qualitatively summarized to obtain an overall rating of the measurement property across all included studies: sufficient (+), insufficient (−), “inconsistent” (±) or indeterminate (?). If all studies indicated sufficient or insufficient results, the overall rating was accordingly. If there were inconsistencies between studies, explanations were explored. If no explanations were found, the overall rating would be inconsistent. The overall rating would be indeterminate if not enough information was available [27].

In the third step, this overall rating of evidence was supplemented by a level of quality of the evidence, using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach from the COSMIN methodology to grade the confidence in the total body of evidence available for the measurement properties [27]. Quality of the evidence was graded as high, moderate, low, or very low. This grade was based on (i) risk of bias, (ii) indirectness, (iii) inconsistency of results, and (iv) imprecision of studies. Each study was rated by a single rater (HM), whose ratings were checked by a second independent rater (KN). Discrepancies in ratings were discussed until consensus was reached.

Results

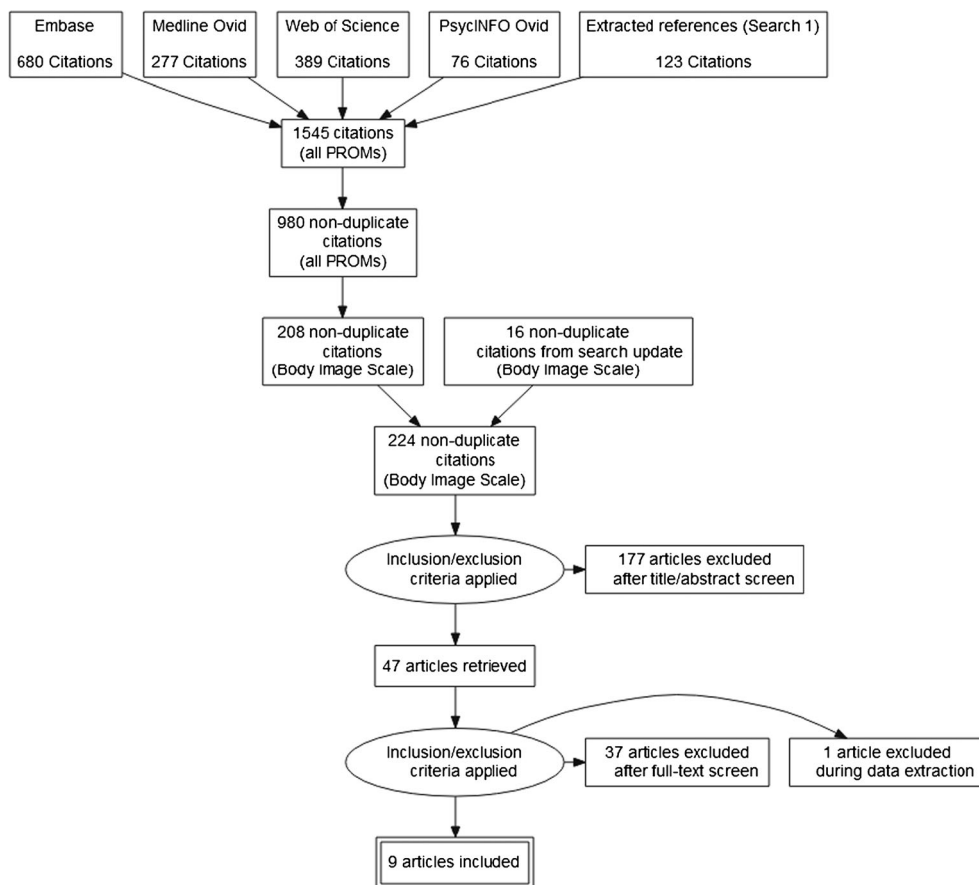
Search results

In total, 980 non-duplicate abstracts were screened, of which 208 abstracts concerned the BIS. The 2017 search update resulted in 16 extra abstracts on the BIS. Having applied inclusion and exclusion criteria, 177 studies were excluded after title/abstract screening. Of the remaining 47 studies, 37 were excluded after full-text screening and one was excluded during data extraction. In total, we included nine studies that investigated measurement properties of the BIS in cancer patients (see Fig. 1).

Study characteristics

Table 1 summarizes the characteristics of the included studies. One study described the development and validation of the BIS in English [14]. Six studies examined validity of the translated BIS in other languages (Greek, Spanish, Korean,

Fig. 1. Flow diagram of the systematic search



Portuguese, Dutch, and Turkish) [15–17, 31–33]. In one study, screening of body image in patients with advanced

cancer (locally advanced, recurrent, or metastatic) was specifically the focus [18]. One study validated the BIS in colorectal

Table 1 Characteristics of included studies

| Reference | Main aim of study | Population | Sample size |
|-----------------------------|---|---|-------------|
| Anagnostopoulos et al. [16] | Examining reliability and validity of Body Image Scale in Greek | Breast cancer patients treated with mastectomy or breast-conserving surgery; Greece | 70 |
| Gómez-Campelo et al. [31] | Validation of Body Image Scale in Spanish | Breast and gynecological cancer patients; Spain | 100 |
| Hopwood et al. [14] | Development and validation of Body Image Scale in English | Breast cancer patients; UK | 682 |
| Karayurt et al. [32] | Validation of Body Image Scale in Turkish | Ostomy patients; Turkey | 100 |
| Khang et al. [33] | Validation of Body Image Scale in Korean | Breast cancer patients treated with mastectomy, breast-conserving surgery or oncoplastic surgery; South Korea | 155 |
| Moreira et al. [17] | Validation of Body Image Scale in Portuguese | Postoperative breast cancer patients; Portugal | 173 |
| Rhondali et al. [18] | To examine the construct of body image dissatisfaction and its measurement using a single question in patients with advanced cancer | Advanced cancer; USA | 81 |
| Van Verschuer et al. [15] | Validation of Body Image Scale in Dutch | Breast cancer patients who have received breast conserving treatment or mastectomy; The Netherlands | 209 |
| Whistance et al. [19] | Validation of Body Image Scale for colorectal patients undergoing surgery | Colorectal cancer patients undergoing surgery; UK | 82 |

cancer patients undergoing surgery [19]. The study populations were breast cancer patients [14, 17, 33], colorectal cancer patients [19], patients with an ostomy (included because 82% of the population were cancer patients) [32], or a mixed cancer population (including breast, gynecological, gastro-intestinal, genitourinary, head and neck, hematologic, and respiratory cancer) [18, 31]. We report on the results based on data extracted from nine studies addressing structural validity, internal consistency, reliability, hypothesis testing for construct validity, and responsiveness. Although none of the studies reported on measurement error, this could be calculated for three studies. None of the studies presented results on cross-cultural validity or criterion validity.

Measurement properties

Structural validity

In total, seven studies examined structural validity using exploratory factor analyses (EFA) [14, 16, 17, 19, 31–33] and

three studies performed an additional confirmatory factor analysis (CFA) [16, 31, 32] (Table 2).

Two studies of excellent [14] and good [33] quality concluded that, over the total study sample, the BIS has a one-factor solution. In subgroup analyses, a two-factor structure was found among breast cancer patients after mastectomy [14] and breast cancer patients after surgery with immediate breast reconstruction [33]. Three fair quality studies also reported a one-factor solution [17, 31, 32] and one fair quality study reported a two-factor solution [16] among breast cancer patients after breast-conserving surgery (BCS) or mastectomy. In the poor quality study [19], a multi-trait item analysis was performed.

Based on these findings, structural validity of the BIS overall was rated sufficient (+) because two studies of at least good quality and three studies of fair quality support unidimensionality of the scale. It should be noted that in some studies, a two-factor solution was also found. The quality of evidence of structural validity was graded as moderate due to inconsistent findings.

Table 2 Structural validity of the BIS

| Reference | Methodology | Results | Methodological quality | |
|-----------------------------|-------------------------------------|--|------------------------|--------|
| | | | | Rating |
| Anagnostopoulos et al. [16] | EFA ^a , CFA ^b | Two factor solution: perceived attractiveness accounting for 52.7% of the variance, and body appearance satisfaction accounting for 8.4% of the variance. The two factors were positively intercorrelated ($r = 0.81$). Fit statistics were adequate. RMSEA: 0.058; SRMR: 0.069; CFI: 0.95. | Fair | – |
| Gómez-Campelo et al. [31] | EFA, CFA | One factor solution accounting for 81.03% of the variance with acceptable fit statistics. SRMR: 0.059. | Fair | + |
| Hopwood et al. [14] | EFA | One factor solution in three analyses accounting for 50.1–57.6% of variance. Two-factor solution for mastectomy subgroup: appearance/attractiveness (26.9% of variance) and body satisfaction (18.8% of variance) but results were not reproducible. | Excellent | + |
| Karayurt et al. [32] | EFA, CFA | One-factor solution, fit statistics were acceptable. SRMR: 0.05; CFI: 0.96. | Fair | + |
| Khang et al. [33] | EFA | One-factor solution for global (66.6% of variance), BCS (59.9% of variance), and mastectomy (74.4% of variance) subgroups. Two-factor solution for oncoplastic subgroup (40.2 and 28.6% of variance). | Good | + |
| Moreira et al. [17] | PCA ^c | One-factor solution with eigenvalue of 6.12, explaining 61.2% of variance. | Fair | + |
| Whistance et al. [19] | Multi-trait item scaling | One-factor solution single items each correlated well with the overall ten-item BIS scale with the exception of item 10 ($r = 0.39$). Removal of this item improved the scaling. Factor analysis suggested a one-factor solution, but item 10 had the lowest factor loading (0.41). This analysis was also repeated with item 10 excluded, and the factor loadings of the remaining nine items improved. | Poor | ? |

+ sufficient, ? Indeterminate, – insufficient, NA not applicable, RMSEA root mean square error of approximation, SRMR standardized root-mean-square residual, CFI comparative fit index BCS breast-conserving surgery

^a Exploratory factor analysis

^b Confirmatory Factor Analysis

^c Principal Component Analysis

Internal consistency

All nine included studies reported on internal consistency using Cronbach's alpha (α) (Table 3). In the excellent and good quality studies, values ranged between $\alpha = 0.86$ – 0.96 [14, 15, 19, 33]. These results are sufficient for internal consistency ($\alpha \geq 0.70$ and ≤ 0.95) [27], although in one mastectomy subgroup, a value of $\alpha = 0.96$ was presented, which might reflect overlap of items within the scale. Five studies had fair methodological quality since missing items were not described. Of these studies, four showed sufficient internal consistency [16–18, 32] and one [31] showed insufficient results because of values of $\alpha = 0.97$ in all subgroups.

Based on these findings, internal consistency of the BIS overall was rated as sufficient (+) and the quality of evidence

of internal consistency was graded as moderate because there is moderate evidence for the unidimensionality of the scale.

Reliability

Four studies examined test-retest reliability. The good and fair quality studies reported values of $r = 0.92$ [15] and $r = 0.85$ [32], indicating sufficient results. Two studies had poor quality and therefore indeterminate results because the time interval was considered too long (6 months compared to 2 weeks in the other studies) [33] and because of a small sample size ($n = 19$) [19], reporting values of ICC = 0.67 and $r = 0.89$, respectively. The low value of 0.67 may be an underestimation of the true reliability because of the long time interval. Hence, reliability of the BIS overall was rated as sufficient (+).

Table 3 Internal consistency (Cronbach's α) of the BIS

| Reference | (Sub)groups | Value (α) | Methodological quality | Rating |
|-----------------------------|---------------------------------------|--------------------|------------------------|--------|
| Anagnostopoulos et al. [16] | Satisfaction subscale (7 items) | 0.87 | Fair | + |
| | Attractiveness subscale (3 items) | 0.92 | | |
| | General body image concerns (5 items) | 0.81 | | |
| Gómez-Campelo et al. [31] | Total sample | 0.97 | Fair | – |
| | Breast cancer subgroup | 0.97 | | |
| | Gynecological cancer subgroup | 0.97 | | |
| Hopwood et al. [14] | Total sample | 0.93 | Excellent | + |
| | BCS subgroup | 0.91 | | |
| | Mastectomy subgroup | 0.91 | | |
| | Remaining subgroups ^a | 0.86 | | |
| Karayurt et al. [32] | Total sample | 0.94 | Fair | + |
| Khang et al. [33] | Total sample | 0.94 | Good | + |
| | BCS subgroup | 0.92 | | |
| | Mastectomy subgroup | 0.96 | | |
| | Oncoplastic surgery subgroup | 0.92 | | |
| Moreira et al. [17] | Total sample | 0.93 | Fair | + |
| | BCS subgroup | 0.93 | | |
| | Mastectomy subgroup | 0.92 | | |
| Rhondali et al. [18] | Total sample | 0.88 | Fair | + |
| Van Verschuer et al. [15] | Total sample (time 1) | 0.91 | Good | + |
| | Total sample (time 2) | 0.92 | | |
| Whistance et al. [19] | Total sample (9-item scale) | 0.90 | Good | + |

BCS breast-conserving surgery

^a Breast cancer patients, advanced breast cancer patients, breast cancer patients with oncoplastic surgery, genetic high-risk women following bilateral prophylactic mastectomy

The quality of evidence of reliability was graded as moderate because three out of four studies reported Pearson/Spearman's correlation coefficients [15, 32, 33], while an intraclass correlation coefficient (ICC) would have been more appropriate.

Measurement error

Although measurement error was not reported in the included studies, we were able to calculate the standard error of measurement (SEM) and the smallest detectable change (SDC) in three studies reporting reliability data and standard deviations. Two studies of good [15] and fair quality ($n = 40$) [32] had an SDC of 4.7 (SEM = 1.7) and 9.1 (SEM = 3.3), respectively. The poor quality study because of the large time interval between the measurements had an SDC of 11.1 (SEM = 4.0) [33]. Interpretation of measurement error is only possible if a SDC score is compared with data on minimal important change (MIC), but this was not reported. Based on these findings, measurement error of the BIS overall was graded as indeterminate (?).

Hypothesis testing for construct validity

Known-groups comparison Eight studies performed known-group comparisons (Table 4). No a priori hypotheses were formulated in four studies [15, 17, 18, 33], and in those cases, we assumed the hypothesis would be that BIS scores are higher (worse) (1) in patients who were treated with a mastectomy compared to patients treated with BCS [34] or breast reconstruction [35], (2) in younger patients compared to older patients [36], (3) in patients with a longer time since treatment [37], and (4) in patients with a stoma vs. without a stoma [38]. Two studies with good quality confirmed their hypotheses [14, 19]. Out of five studies with fair quality [15–17, 31, 33], two studies confirmed the hypotheses [15, 16]. One study had a poor quality [18] because no a priori hypotheses were formulated.

Convergent and divergent validity Six studies reported on convergent validity with other body image-related instruments, psychological function, or HRQOL scales (Table 4). One good quality study [19] showed moderate correlation ($r = 0.40$ to 0.60) with a related construct but failed to confirm their hypotheses on three other constructs, indicating insufficient convergent validity. One study of fair quality [31] found moderate and high correlations ($r > 0.60$) with related constructs, indicating sufficient convergent validity. However, three other fair quality studies [16, 17, 33] presented low correlations ($r < 0.40$) with most of the related constructs, indicating insufficient convergent validity. The poor quality study did not formulate a hypothesis a priori [18]. None of the studies in this review examined divergent validity.

Based on these findings, hypothesis testing for construct validity was rated as inconsistent (\pm) because although three studies showed sufficient evidence ($> 75\%$ of the hypotheses on known-groups and/or convergent validity confirmed) [14, 15, 31], this was contradicted by four studies showing insufficient evidence [16, 17, 19, 33]. Moreover, studies reported inconsistent results in comparison with the same instrument (ASI-R and RSES) [17, 18, 33]. For this reason, and due to the lack of clearly stated a priori hypotheses, quality of evidence of construct validity was graded as low.

Responsiveness

Two studies reported on responsiveness. One study of good quality [14] found a significant increase in body image disturbance for the overall sample ($n = 55$) and for the BCS and mastectomy subgroups 2 weeks to 4 months postoperatively, indicating sufficient responsiveness. The other study had poor quality [19] because of a small sample size ($n = 17$) and found no change in BIS scores from before to after surgical treatment. Based on these findings, responsiveness of the BIS was rated as indeterminate (?). An overall summary of the results for every measurement property of the BIS is shown in Table 5.

Discussion

This systematic review evaluated the measurement properties of the BIS among nine studies identified in a literature search up to July 2017. In summary, evidence on structural validity, internal consistency, and reliability of the BIS was rated as sufficient, and the quality of evidence was moderate. Measurement error and responsiveness were rated as indeterminate, and hypothesis testing for construct validity was rated as inconsistent with a low quality of evidence. None of the studies reported on criterion validity and cross-cultural validity.

For *structural validity*, a one-factor solution was found and evidence was rated as sufficient. However, one fair quality study and subgroup analyses in two good quality studies showed a two-factor structure [14, 16, 33]. Hopwood et al. [14] found a two-factor structure among breast cancer patients after mastectomy, and Khang et al. [33] after surgery with immediate breast reconstruction. These two factors were labeled as “attractiveness” and “satisfaction with body” [14, 16]. However, there was no agreement on which items belonged to which factors precisely. Also, the findings were inconsistent and in the study of Khang et al. [33] based on a relatively small study sample (subgroup $n < 50$). Further research is therefore needed to investigate whether the BIS is a unidimensional construct in all breast cancer patients, regardless of treatment modality.

Table 4 Known-group comparison and convergent validity of the BIS

| Reference | Known-group comparison | | | | Convergent validity | | | |
|-----------------------------|---|---|------------------------|--------|---|---|------------------------|--------|
| | Comparison groups | Results | Methodological quality | Rating | Comparison instrument | Correlations | Methodological quality | Rating |
| Anagnostopoulos et al. [16] | Patients who underwent mastectomy vs. BCS vs. cancer-free women | Compared to women receiving breast-conserving surgery, women receiving mastectomy reported significantly more reduced perceived attractiveness, and greater dissatisfaction with body and appearance. For low-social dysfunction scores, there were no significant differences in general body image concerns among the three groups of women. However, for the high-social dysfunction scores, women who had undergone mastectomy exhibited significantly higher scores on general body image concerns, compared to cancer-free and BCS women's scores. | Fair | + | GHQ-28 - Social dysfunction - Anxiety/insomnia | BIS appearance and attractiveness scale 0.60; 0.38 0.40; 0.26 | Fair | - |
| Gómez-Campelo et al. [31] | Age and time since diagnosis | Significantly higher BIS scores in younger patients. No significant relation between BIS and time since diagnosis. BIS scores were significantly higher in patients who were treated with mastectomy than those treated with BCS. | Fair | - | RSES BDI BAI EORTC QLQ-C30 | -0.73 0.83 0.56 -0.63 | Fair | + |
| Hopwood et al. [14] | Patients who underwent mastectomy vs. BCS | BIS scores were significantly higher in patients who were treated with mastectomy than those treated with BCS. | Good | + | | | | |
| Khang et al. [33] | Patients who underwent mastectomy vs. BCS vs. oncoplastic surgery | BIS scores were significantly higher in patients who were treated with mastectomy than those treated by BCS or oncoplastic surgery. However, the statistical | Fair | - | BESAA RSES HADS total HADS-A HADS-D WHOQOL-BREF - Overall QOL | -0.30 -0.12 0.52 0.50 0.46 -0.22 | Fair | - |

Table 4 (continued)

| Known-group comparison | | Convergent validity | |
|----------------------------|--|--|--|
| Moreira et al. [17] | <p>Patients who underwent mastectomy vs. BCS: age and time since diagnosis</p> <p>BIS scores were significantly higher in patients who were treated with mastectomy than those treated with BCS. The effect size ($r^2 = .13$) was considered medium. No association with age and time since diagnosis.</p> <p>Significantly higher BIS scores in younger patients.</p> | <p>significance was found only between the mastectomy and oncoplastic surgery subgroups.</p> <p>Fair</p> | <ul style="list-style-type: none"> - General health -0.38 - Physical health domain -0.36 - Psychological domain -0.32 - Bodily image and appearance facet -0.31 - Social relationships domain -0.25 - Environmental domain -0.30 ESS 0.68 DAS24 0.75 ASI-R self-evaluative salience 0.40 ASI-R motivational salience -0.12 WHOQOL-BREF - General health -0.52 - Physical health domain -0.42 - Psychological domain -0.49 - Body image and appearance -0.66 <p>Fair</p> |
| Rhondali et al. [18] | <p>Age</p> <p>Significantly higher BIS scores in younger patients.</p> | <p>Poor</p> | <ul style="list-style-type: none"> ASI-R 0.24 HADS-A 0.52 HADS-D 0.42 ESAS total symptom distress score 0.41 ESAS physical distress subscore 0.35 ESAS psychological distress subscore 0.37 MBSRQ Overall appearance satisfaction item -0.44 <p>Poor</p> |
| Van Verschuier et al. [15] | <p>Patients who underwent mastectomy vs. BCS</p> <p>BIS scores were significantly higher in patients treated with mastectomy than those treated with BCS at both assessment times. The effect size ($d = .47$) was considered moderate.</p> | <p>Fair</p> | <p>+</p> |
| Whistance et al. [19] | <p>Patients with a stoma vs. patients without a stoma</p> <p>BIS scores were significantly higher in patients with a stoma than patients without a stoma.</p> | <p>Good</p> | <ul style="list-style-type: none"> EORTC QLQ-C30 emotion function 0.45 EORTC QLQ-C30 role function <0.40 (exact data not shown) EORTC QLQ-C30 social function <0.40 (exact data not shown) EORTC QLQ-C30 global quality of life <0.40 (exact data not shown) <p>Good</p> |

BCS breast-conserving surgery, *GHQ-28* General Health Questionnaire-28, *RSES* Rosenberg Self-Esteem Scale, *BDI* Beck Depression Inventory, *BAI* Beck Anxiety Inventory, *EORTC QLQ-C30* European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30, *BESAA* Body-Esteem Scale for Adolescents and Adults, *HADS* Hospital Anxiety and Depression Scale, *WHOQOL-BREF* World Health Organization Quality of Life scale-abbreviated version, *ESS* Experience of Shame Scale, *DAS24* Derriford Appearance Scale 24, *ASI-R* Appearance Schemas Inventory—revised, *ESAS* Edmonton Symptom Assessment System

Table 5 Overall rating of the results and levels of evidence of the BIS

| Measurement property | Rating of measurement property | Quality of evidence |
|-------------------------|--------------------------------|---------------------|
| Structural validity | + | Moderate |
| Internal consistency | + | Moderate |
| Reliability | + | Moderate |
| Measurement error | ? | |
| Hypothesis testing | ± | Low |
| Cross-cultural validity | NA | NA |
| Criterion validity | NA | NA |
| Responsiveness | ? | |

Evidence on *reliability* was sufficient because it met the criterion of 0.70 in three out of four studies (range 0.67–0.92). The one study that found a correlation <0.70 had a large time interval (6 months) between the two measurements and was therefore judged as having a poor methodological quality. It is known that body image symptoms can change in the first few months after cancer treatment [14], with patients reporting high deterioration and recovery trajectories [39]. Moreover, body changes (e.g., weight fluctuations or healing of wounds) can occur within half a year. A 7–14-day interval for test-retest reliability is in general considered most appropriate [30].

Measurement error was not reported in any of the included studies, but the SDC could be calculated in three studies. When only taking into account good and fair quality studies, the smallest change in score that can be detected, that is not due to measurement error, ranges between 4.7–9.1 [15, 32], on a total range of 0–30 of the BIS. However, these data are difficult to interpret since no information is available on the anchor points minimal important change (MIC) or minimal important difference (MID). Therefore, further research is needed to establish these anchor points on changes that are important.

Evidence on *hypothesis testing* for construct validity was inconsistent since findings for known-group comparisons and convergent validity were inconsistent. Known-group comparisons in most studies focused on body image issues related to surgical treatment (comparing breast cancer patients treated with mastectomy versus BCS). It is known that other types of treatment may also impact body appearance. For example, cancer survivors who received chemotherapy reported that hair loss and weight gain disrupted their body image [40, 41]. In addition to recommendations to include other cancer populations than breast cancer patients [20], we also recommend to study construct validity of the BIS taking into account the impact of various cancer treatments on body image.

With respect to convergent validity, correlations with other body image scales were inconsistent. There were indications that consciousness of appearance (DAS24) and shame (ESS) are related with body image, with moderate to high correlations [17]. However, correlation with investment in

appearance (ASI-R) was low [17, 18]. Moreover, the relation with self-esteem (RSES) was inconsistent, with only one of two studies finding a high correlation [31, 33]. Given these contradictory findings and the fair quality of these studies, no firm conclusions can be drawn about convergent validity of the BIS. This contradicts the conclusion of Muzzatti et al. presenting adequate convergent validity [20].

Evidence for *responsiveness* was indeterminate. Only one study of good methodological quality reported a change in BIS scores postoperatively [14], but no hypotheses were formulated on the expected magnitude of change and no comparison with another instrument was made. More research is needed about the ability of the BIS to detect change in body image symptoms over time.

A limitation of this review is that content validity was not investigated because at the time we conducted our data extraction, no protocol existed to investigate content validity through a systematic review. Recently, this protocol has become available [42]. Another limitation is that a precise filter instead of a sensitive filter was used. The precise filter was a pragmatic choice because a sensitive filter would provide too many hits to feasibly screen since the overall search encompassed 39 PROMs (Prospero ID 42017057237) [22]. There is a small possibility that validation studies of the BIS may have been missed. Lastly, the assessment of quality ratings was performed by one rater. This rating was then checked by a second independent rater, and discussed until consensus was reached. The gold standard practice is to have the assessment done by two raters independently because raters initially may have different opinions and consensus is needed.

This systematic review provides in-depth insight of the current evidence of the BIS as an instrument to measure body image in cancer patients and complements a recent review [20]. For researchers who want to further study the psychometric properties of the BIS, this paper points out future directions. With respect to reliability, this includes examining measurement error and research on minimal important change. Regarding validity, existing evidence on content validity should be summarized and new evidence is needed for cross-cultural validity. Criterion validity is impossible to assess, since a “gold standard” for assessing body image is not

available. Efforts are therefore needed to reach consensus on a measure that could serve as second best. This may comprise body image scores by proxies such as health care providers with vast experience in the targeted study population. Furthermore, it would be valuable to examine structural validity on a possible two-factor structure among cancer subgroups (patients who had reconstructive surgery or amputation of a body part) more thoroughly. High-quality studies exploring convergent validity with investment in appearance (ASI-R) and self-esteem (RSES) are recommended. Finally, responsiveness should be more thoroughly investigated by formulating hypotheses for change scores in the BIS compared to change scores in other instruments. The BIS is mainly tested in a population of patients who are surgically treated for breast cancer. Further research including a wider variety of cancer patients and treatment modalities is recommended. New validation studies with a good methodological quality can further optimize evidence regarding the measurement properties of the BIS.

Funding This work was supported by the Dutch Cancer Society, grant number VUP 2014-7202.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

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

- Cash TF, Smolak L (2011) *Body image: a handbook of science, practice, and prevention*, 2nd edn. Guilford Press, New York
- Lehmann V, Hagedoorn M, Tuinman MA (2014) Body image in cancer survivors: a systematic review of case-control studies. *J Cancer Surviv* 9:339–348. <https://doi.org/10.1007/s11764-014-0414-y>
- Rumsey N, Harcourt D (2004) Body image and disfigurement: issues and interventions. *Body Image* 1:83–97. [https://doi.org/10.1016/S1740-1445\(03\)00005-6](https://doi.org/10.1016/S1740-1445(03)00005-6)
- Falk Dahl CA, Reinertsen KV, Nesvold IL, Nesvold IL, Fosså SD, Dahl AA (2010) A study of body image in long-term breast cancer survivors. *Cancer* 116:3549–3557. <https://doi.org/10.1002/cncr.25251>
- Rhoten BA, Murphy B, Ridner SH (2013) Body image in patients with head and neck cancer: a review of the literature. *Oral Oncol* 49:753–760. <https://doi.org/10.1016/j.oraloncology.2013.04.005>
- Defrank JT, Mehta CC, Stein KD, Baker F (2007) Body image dissatisfaction in cancer survivors. *Oncol Nurs Forum* 34:E36–E41. <https://doi.org/10.1188/07.ONF.E36-E41>
- Taylor-Ford M, Meyerowitz BE, D’Orazio LM, Christie KM, Gross ME, Agus DB (2013) Body image predicts quality of life in men with prostate cancer. *Psychooncology* 22:756–761. <https://doi.org/10.1002/pon.3063>
- Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, Liberty J (2015) Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Ann Oncol* 26:1846–1858. <https://doi.org/10.1093/annonc/mdv181>
- Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, MacGillivray S (2014) What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol* 32:1480–1501. <https://doi.org/10.1200/JCO.2013.53.5948>
- Cash TF, Melnyk SE, Hrabosky JI (2004) The assessment of body image investment: an extensive revision of the appearance schemas inventory. *Int J Eat Disord* 35:305–316. <https://doi.org/10.1002/eat.10264>
- Frierson GM, Thiel DL, Andersen BL (2006) Body change stress for women with breast cancer: the Breast-Impact of Treatment Scale. *Ann Behav Med* 32:77–81. https://doi.org/10.1207/s15324796abm3201_9
- Ferguson SE, Urowitz S, Massey C, Wegener M, Quartey N, Wiljer D, Classen CC (2012) Confirmatory factor analysis of the Sexual Adjustment and Body Image Scale in women with gynecologic cancer. *Cancer* 118:3095–3104. <https://doi.org/10.1002/cncr.26632>
- Fingeret MC, Nipomnick S, Guindani M, Baumann D, Hanasono M, Crosby M (2014) Body image screening for cancer patients undergoing reconstructive surgery. *Psychooncology* 23:898–905. <https://doi.org/10.1002/pon.3491>
- Hopwood P, Fletcher I, Lee A, Al Ghazal S (2001) A body image scale for use with cancer patients. *Eur J Cancer* 37:189–197. [https://doi.org/10.1016/S0959-8049\(00\)00353-1](https://doi.org/10.1016/S0959-8049(00)00353-1)
- van Verschuer VM, Vrijland WW, Mares-Engelberts I, Klem TM (2015) Reliability and validity of the Dutch-translated Body Image Scale. *Qual Life Res* 24:1629–1633. <https://doi.org/10.1007/s11136-014-0907-1>
- Anagnostopoulos F, Myrghianni S (2009) Body image of Greek breast cancer patients treated with mastectomy or breast conserving surgery. *J Clin Psychol Med Settings* 16:311–321. <https://doi.org/10.1007/s10880-009-9176-5>
- Moreira H, Silva S, Marques A, Canavarro MC (2010) The Portuguese version of the Body Image Scale (BIS)—psychometric properties in a sample of breast cancer patients. *Eur J Oncol Nurs* 14:111–118. <https://doi.org/10.1016/j.ejon.2009.09.007>
- Rhondali W, Chisholm GB, Filbet M, Kang DH, Hui D, Cororve Fingeret M, Bruera E (2015) Screening for body image dissatisfaction in patients with advanced cancer: a pilot study. *J Palliat Med* 18:151–156. <https://doi.org/10.1089/jpm.2013.0588>
- Whistance RN, Gilbert R, Fayers P, Longman RJ, Pullyblank A, Thomas M, Blazeby JM (2010) Assessment of body image in patients undergoing surgery for colorectal cancer. *Int J Color Dis* 25:369–374. <https://doi.org/10.1007/s00384-009-0851-7>
- Muzzatti B, Annunziata MA (2017) Body image assessment in oncology: an update review. *Support Care Cancer* 25:1019–1029. <https://doi.org/10.1007/s00520-016-3538-y>
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 63:737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Neijenhuis 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. http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42017057237. Accessed 25 Sep 2017
23. van der Hout A, van Uden-Kraan CF, Witte BI, Coupé VMH, Leemans CR, Cuijpers P, van de Poll-Franse LV, Verdonck-de Leeuw IM (2017) Efficacy, cost-utility and reach of an eHealth self-management application “Oncokompas” that helps cancer survivors to obtain optimal supportive care: study protocol for a randomised controlled trial. *Trials* 18:228. <https://doi.org/10.1186/s13063-017-1952-1>
 24. 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>
 25. Duman-Lubberding S, van Uden-Kraan CF, Jansen F, Witte BI, van der Velden LA, Lacko M, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM (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>
 26. Terwee CB, Jansma EP, Riphagen II, De Vet HC (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>
 27. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. <https://doi.org/10.1007/s11136-018-1798-3>
 28. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (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>
 29. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC (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>
 30. Terwee CB, Bot SD, De Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60:34–42. <https://doi.org/10.1016/J.JCLINEPI.2006.03.012>
 31. Gómez-Campelo P, Bragado-Álvarez C, Hernández-Lloreda MJ, Sánchez-Bernardos ML (2015) The Spanish version of the Body Image Scale (S-BIS): psychometric properties in a sample of breast and gynaecological cancer patients. *Support Care Cancer* 23:473–481. <https://doi.org/10.1007/s00520-014-2383-0>
 32. Karayurt Ö, Edeer AD, Süler G, Dorum H, Harputlu D, Vural F, Ü er C (2015) Psychometric properties of the Body Image Scale in Turkish ostomy patients. *Int J Nurs Knowl* 26:127–134. <https://doi.org/10.1111/2047-3095.12049>
 33. Khang D, Rim HD, Woo J (2013) The Korean version of the body image scale-reliability and validity in a sample of breast cancer patients. *Psychiatry Investig* 10:26–33. <https://doi.org/10.4306/pi.2013.10.1.26>
 34. Al-Ghazal SK, Fallowfield L, Blamey RW (2000) Comparison of psychological aspects and patient satisfaction following breast conserving surgery, simple mastectomy and breast reconstruction. *Eur J Cancer* 36:1938–1943
 35. Fang SY, Shu BC, Chang YJ (2013) The effect of breast reconstruction surgery on body image among women after mastectomy: a meta-analysis. *Breast Cancer Res Treat* 137:13–21. <https://doi.org/10.1007/s10549-012-2349-1>
 36. Metzger LF, Rogers TF, Bauman LJ (1983) Effects of age and marital status on emotional distress after a mastectomy. *J Psychosoc Oncol* 1:17–33. https://doi.org/10.1300/J077v01n03_02
 37. Sneeuw KC, Aaronson NK, Yarnold JR, Broderick M, Regan J, Ross G, Goddard A (1992) Cosmetic and functional outcomes of breast conserving treatment for early stage breast cancer. 1. Comparison of patients’ ratings, observers’ ratings and objective assessments. *Radiother Oncol* 25:153–159
 38. Sharpe L, Patel D, Clarke S (2011) The relationship between body image disturbance and distress in colorectal cancer patients with and without stomas. *J Psychosom Res* 70:395–402. <https://doi.org/10.1016/j.jpsychores.2010.11.003>
 39. Lam WW, Li WW, Bonanno GA, Mancini AD, Chan M, Or A, Fielding R (2012) Trajectories of body image and sexuality during the first year following diagnosis of breast cancer and their relationship to 6 years psychosocial outcomes. *Breast Cancer Res Treat* 131:957–967. <https://doi.org/10.1007/s10549-011-1798-2>
 40. Dua P, Heiland MF, Kracen AC, Deshields TL (2017) Cancer-related hair loss: a selective review of the alopecia research literature. *Psychooncology* 26:438–443. <https://doi.org/10.1002/pon.4039>
 41. Heideman WH, Russell NS, Gundy C, Rookus MA, Voskuil DW (2009) The frequency, magnitude and timing of post-diagnosis body weight gain in Dutch breast cancer survivors. *Eur J Cancer* 45:119–126. <https://doi.org/10.1016/j.ejca.2008.09.003>
 42. Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, de Vet HC, Patrick D, Alonso J, Bouter LM, Mokkink LB (2016) Consensus-based standards and criteria for evaluating the content validity of patient-reported outcome measures: a COSMIN Delphi study. *Qual Life Res* 25:1–1