PROMs in Oncokompas 2.0 - A systematic review of measurement properties

Koen Neijenhuijs

2017-11-08

Contents

Chapter 1

Abstract

The current report presents the results of a systematic review of 29 patient-reported outcome measures (PROMs) used in Oncokompas 2.0.

A five-step cascading search strategy was used. First, we searched for systematic reviews of the 29 PROMs used in cancer populations. Second, for the PROMs that did not turn up (enough) useable data, we searched for individual validation studies in cancer populations. Third, for the PROMs that did not turn up (enough) useable data, we searched for systematic reviews in any population. Fourth, for the PROMs that did not turn up (enough) useable data, we searched for individual validation studies in any population. Fifth, for PROMs that had zero hits on the systematic searches, manual searches of the "PROMs in care" database, Google, and Google Scholar were performed.

Data was extracted following the COSMIN criteria. Data was extracted of 274 studies found in the main systematic searches. For seven PROMs, zero search hits were found, and the manual search resulted in data extraction from six articles, one manual, and one thesis. Two PROMs had zero usable data sources.

Five PROMs were judged to be of high priority recommendation of exploring alternative PROMs. Six PROMs were indicated as medium priority recommendation of additional research into either missing information of important measurement properties, or into replication of the validity in a cancer population. Eleven PROMs were judged to be of low priority recommendation of research into missing information of measurement properties. Seven PROMs were judged to of no priority, and were without doubt suitable for use in Oncokompas 2.0.

Chapter 2

Introduction

Measuring quality of life (QoL) can be an arduous task, as many different aspects of life may have an impact on its' quality. Due to the wide range of of aspects, many different QoL measurement tools have been developed in the past two decades. In oncology alone, more than a decade ago, 439 records of QoL measurement tools could be identified (Garratt et al., 2002). One of the most-used QoL measurement tool is the European Organization for Research and Treatment of Cancer (EORTC)'s QLQ-C30 (Bjordal et al., 2000) and its' various cancer site-specific, symptom-specific, and QoL domain-specific modules (Velikova et al., 2012), of which a total of 21 have been validated in large international samples (EORTC Quality of Life Department, b) and another 22 are ready for use but not yet validated in a large international sample (EORTC Quality of Life Department, a).

This provides researchers and clinicians with a large range of measurement tools to assess QoL in cancer patients. The previously mentioned QoL measurement tools usually cover multiple aspects of QoL to reach either a composite score or a score per aspect. However, at times one may want to delve into aspects of QoL not covered by these measurement tools. For example, one may want to measure fear of cancer recurrence which can have a significant impact on QoL (Custers, 2014).

For the development of the eHealth application Oncokompas 2.0 - an application aimed at facilitating access to supportive cancer care - a test battery to measure relevant aspects of QoL was assembled (Lubberding et al., 2015; Jansen et al., 2015; Duman-Lubberding et al., 2016). For each individual QoL aspect, Patient Reported Outcome Measures (PROMs) were picked based on expert opinion, relevant Dutch guidelines, literature searches, and dissertations. As the length of the test battery was considerable due to the total of 67 determined relevant aspects of QoL, one last criterion for choosing PROMs was their length.

The current report addresses a systematic review of the mesurement properties for the non-in-house developed PROMs used in the eHealth application "Oncokompas" (Table 2.1). First we will cover the methodology used in this systematic review, and second we will cover the search results. The rest of the report describes the findings per QoL area:

1. Psychological Wellbeing
2. Physical Generic Wellbeing
3. Social Life
4. Lifestyle
5. Breast Cancer
6. Intestinal Cancer
7. Head and Neck Cancer

Table 2.1: Quality of Life constructs and PROMs used in Oncokompas 2.0

Measurement Construct	Questionnaire
Psychological wellbeing	
Depression	Hospital Anxiety and Depression Scale
Anxiety	Hospital Anxiety and Depression Scale
Fear of recurrence	Cancer Worry Scale
Subjective cognitive functioning	SF-36: Cognitive function scale
Physical generic wellbeing	
General everyday life	Patiënt Specifieke Klachten
Sleep	Insomnia Severity Index
Sexuality (women)	6-item Female Sexual Function Index
Sexuality (men)	5-item International Index of Erectile Function
Body image	Body image scale
Nausea and vomiting	EORTC QLQ-C30: Nausea and vomiting scale
Hearing impairment	CARON
Social life	
Social life	Jong-Gierveld Loneliness Scale: Emotional loneliness subscale
Social life	Jong-Gierveld Loneliness Scale: Social loneliness subscale
Relationship with partner	Dyadic Adjustment Scale
Relationship with children	VGK Short-form
Financial circumstances	EORTC QLQ-C30: Financial impact item
Relationship with doctor	EORTC QLQ-PATSAT32
Relationship with boss	Job Content Quenstionnaire: Supervisor scale
Relationship with coworkers	Job Content Questionnaire: Coworkers scale
Job retention and resumption	VBBA
Lifestyle	
Alcohol use	Alcohol Five Shot
Relaxation	Perceived Stress Scale
Breast cancer	
Menopausal symptoms	Endocrine Scale Functional Assessment of Cancer Therapy
Body image	Breast Impact of Treatment Scale
Lymphedema	EORTC QLQ BR23: Lymphedema item
Shoulder function	Quick Disabilities of the Arm, Shoulder, and Hand questionnaire
Breast reconstruction	BRECON-31: satisfaction, recovery, nipple reconstruction, abdomen
Breast prosthesis	Prosthesis Items of Breast Cancer patients' Needs Questionnaire
Intestinal Cancer	
Bladder issues	EORTC QLQ CR29: Urinary frequency scale
Pain (abdomen, buttocks, anus)	EORTC QLQ CR 29: Abodemen, buttocks, and anal pain items
Bloated	EORTC QLQ CR29: Bloated item
Defecation (blood, mucus)	EORTC QLQ CR29: Defecation scale
Stoma leakage	Stome quality of life questionnaire
Stoma gas	Stoma quality of life questionnaire
Stome problems being away from home	Stoma quality of life questionnaire
Stome clothing	Stoma quality of life questionnaire
Sexuality	Stoma quality of life questionnaire
Appearance	Stoma quality of life questionnaire
Social life	Stoma quality of life questionnaire
Head and neck cancer	
Speech	EORTC QLQ H&N35: Speech impedements scale

Measurement Construct	**Questionnaire**
Swallowing	EORTC QL H&N35: Swalling scale and social eating scale
Shoulder function	Shoulder Disability Questionnaire
Mouth function	EORTC QLQ H&N35: Mouth pain, teeth pain, dry mouth, sticky spit, trismus

Chapter 3

Methods

This chapter covers the methodology used in the systematic review.

3.1 Literature search strategy

A five-step cascading search methodology was used. First, a systematic search of Embase, Medline, PsycInfo, Cochrane DARE, PubMed Publisher, and Web of Science was performed for the 29 PROMs used in Oncokompas 2.0 (see Table 2.1). The search terms were the measurement names and their acronyms, combined with search terms for cancer, and search terms for review, meta-analysis or meta-review. The search was performed in February 2016.

Second, we performed a systematic search of Embase, Medline, PsycInfo, and Web of Science for the PROMs that did not provide enough literature inclusions in the previous search (see 3.2). The search terms were the measurement names and their acronyms, combined with search terms for cancer, and a precise filter for measurement properties (Terwee et al., 2009). The search was performed in July 2016.

Third, a systematic search of Embase, Medline, Cochrane DARE, and Web of Science was performed for the PROMs that did not provide enough literature inclusions in the previous search (see 3.2). The search terms were the measurement names and their acronyms, combined with search terms for review, meta-analysis or meta-review. The search was performed in November 2016.

Fourth, we performed a systematic search of Embase, Medline, and Web of Science for the PROMs that did not provide enough literature inclusions in the previous search (see 3.2). The search terms were the measurement names and their acronyms, combined with a precise filter for measurement properties (Terwee et al., 2009). The search was performed in February 2017.

Fifth, we performed a manual search of the "PROMs in care" (Dutch: Meetinstrumenten in de zorg) database as well as Google and Google Scholar, for those PROMs of which we had zero relevant article hits in the previous searches.

3.2 Inclusion criteria

Inclusion criteria for all searches were the presentation of data on a measurement property as defined by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy (Mokkink et al., 2010, 2012; Terwee et al., 2012): internal consistency, reliability, measurement error, structural validity, hypothesis testing (for construct validity), criterion validity, cross-cultural validity and responsiveness.

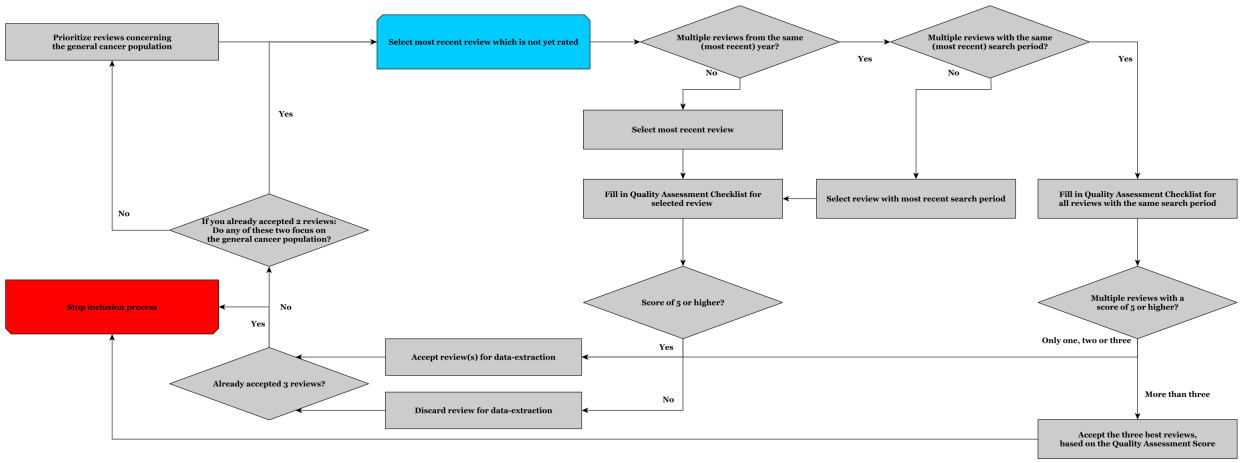

Figure 3.1: Decision tree for review selection

The first two searches excluded articles on non-cancer populations. The second and fourth search excluded reviews and meta-analyses. For all searches, we excluded articles that were only available as abstracts or conference proceedings. Non-English publications were also excluded. Titles and abstracts, and the selected full-texts were reviewed by two independent raters. Disagreements were discussed until consensus was reached.

After each search, the number of inclusions for each PROM and rough estimates of data present in these included articles were screened by two independent reviewers, and discussed until consensus was reached. Based on this information, decisions were made whether enough data would be available for individual PROMs. PROMs without - or with not enough - data, were cascaded into the next search. Cancer-specific PROMs were not cascaded past the second search, as we did not expect any hits for non-cancer populations.

3.3 Data extraction

3.3.1 Reference extraction for searches 1 and 3 (Reviews and Meta-analyses)

To reduce overlap between reviews, the three reviews of the highest quality per PROM were selected. The decision which reviews were of the highest quality was performed through a decision tree (Figure 3.1). The most recent review was selected and assessed using a quality assessment checklist (Oxman and Guyatt, 1991; Oxman et al., 1991). This checklist results in a score between 1 and 7. If the quality score was a 5 (representing 'minor flaws') or higher, it was included in the top three of reviews. This process guarantees the selection of the most recent good quality reviews. The references reported by the top three reviews were extracted, to be used in our own data extraction protocol.

References were also extracted from the reviews on PROMs that continued onto the next search phase. These references were then added to the search hits of the next search.

3.3.2 Data extraction from individual validation studies

Two independent extractors extracted information from included validation studies, on each of the measurement properties defined by the COSMIN taxonomy (Mokkink et al., 2010, 2012; Terwee et al., 2012). Relevant data included the type of measurement property, its outcome, and information on missing values. Disagreements were discussed until consensus was reached.

3.4 Data synthesis

Data was synthesized by following the criteria for good measurement properties of the COSMIN guidelines for systematic reviews of PROMSs (Prinsen et al., 2017). However, as study quality was not taken into account, I chose to make a distinction only between sufficient evidence versus indeterminate evidence. Based on these assessments, recommendations were made on two levels: (1) type of action recommended, and (2) recommended level of priority for said action.

Chapter 4

Search Results

This chapter covers the search results: the number of hits, exclusions during screening, and the number of articles used for data extraction for each PROM.

4.1 Search 1 - Reviews in cancer populations

A total of 793 articles were found in the search. After deletion of 290 duplicates, 503 articles remained. Of these 503 articles, 494 abstracts were screened. A number of abstracts could not be located. A total of 275 abstracts were excluded, and 219 full-texts were screened. A total of 154 full-texts were excluded, resulting in a total of 65 reviews fit for data extraction.

Reviews were included for:

1. Hospital Anxiety and Depression Scale (29 reviews)
2. EORTC QLQ-C30 (26 reviews)
3. SF-36 (9 reviews)
4. EORTC H&N-35 (4 reviews)
5. EORTC QLQ-BR23 (2 reviews)
6. Shoulder Disability Questionnaire (2 reviews)
7. Body Image Scale (2 reviews)
8. International Index of Erectile Functioning (1 review)
9. Female Sexual Functioning Index (1 review)
10. QuickDASH (1 review)
11. FACT-ES (1 review)

Based on breadth of data, quality of the reviews, and the recency of the reviews, it was determined enough data could be extracted for the Hospital Anxiety and Depression Questionnaire, EORTC QLQ-C30, SF-36, and QuickDASH. The rest of the PROMs continued to be part of the next search round.

A total of 15 validation studies were extracted from the top 3 reviews of the Hospital Anxiety and Depression Questionnaire; 11 validation studies were extracted from the top 3 reviews of the EORTC QLQ-C30; 21 validation studies were extracted from the top 3 reviews of the SF-36; and 13 validation studies were extracted from the review of the QuickDASH.

4.2 Search 2 - Validation studies in cancer populations

A total of 1422 articles were found in the search. After adding 123 references extracted from the previous search and the removal of 565 duplicates, a total of 980 abstracts were screened. A total of 880 abstracts

were excluded, and 100 full-texts were screened. A total of 28 full-texts were excluded, resulting in a total of 72 articles fit for data extraction.

A search update was performed for the Body Image Scale, EORTC QLQ-CR29, and EORTC IN-PATSAT32 up to July 2017. Fifteen articles were found for the Body Image Scale, which were all excluded during abstract screening. Fourteen abstracts were found for the EORTC QLQ-CR29, of which twelve were excluded during abstract screening. No articles were excluded during full-text screening, resulting in an addition of two articles for data extraction. Three abstracts were found for the EORTC IN-PATSAT32, of which one was excluded during abstract screening. No articles were excluded during full-text screening, resulting in an addition of two articles for data extraction.

Articles were included for:

1. EORTC QLQ-BR23 (20 articles)
2. Body Image Scale (8 articles)
3. EORTC QLQ-CR29 (10 articles)
4. EORTC IN-PATSAT32 (9 articles)
5. Perceived Stress Scale (5 articles)
6. Insomnia Severity Index (3 articles)
7. Female Sexual Functioning Index (3 articles)
8. International Index of Erectile Functioning (3 articles)
9. EORTC QLQ-H&N35 (2 articles)
10. FACT-ES (2 articles)
11. Breast Impact of Treatment Scale (2 articles)
12. Cancer Worry Scale (2 articles)
13. BRECON-31 (2 articles)
14. Shoulder Disability Questionnaire (1 article)

Based on breadth of data, quality of the studies, and the recency of the studies, it was determined enough data could be extracted for the EORTC QLQ-BR23, Body Image Scale, EORTC IN-PATSAT32, EORTC QLQ-CR29, Insomnia Severity Index, and Shoulder Disability Questionnaire.

The EORTC H&N-35, Cancer Worry Scale, FACT-ES, Breast Impact of Treatment Scale, BRECON-31, and Breast Cancer patients' Needs Questionnaire were also excluded from the next searches as these PROMs are cancer-specific.

4.3 Search 3 - Reviews in all populations

A total of 475 articles were found in the search. After deletion of 103 duplicates, 372 abstracts were screened. A total of 345 abstracts were excluded, and 27 full-texts were screened. A total of 19 full-texts were excluded, resulting in a total of 8 reviews fit for data extraction.

Reviews were included for:

1. International Index of Erectile Functioning (2 reviews)
2. Job Content Questionnaire (2 reviews)
3. Dyadic Adjustment Scale (2 reviews)
4. Female Sexual Functioning Index (1 review)
5. Perceived Stress Scale (1 review)

Based on breadth of data, quality of the reviews, and the recency of the reviews, it was determined enough data could be extracted only for the Perceived Stress Scale. A total of 19 validation studies were included in this particular review. The quality of reporting in the review was outstanding, and raw data was presented. Due to the quality of reporting, it was determined that seperate data extraction from the 19 validation studies was unnecessary. Instead, data was extracted directly from the review.

4.4 Search 4 - Validation studies in all populations

A total of 2680 articles were found in the search. After adding 24 references extracted from the previous search and the removal of 1303 duplicates, a total of 1401 abstracts were screened. A total of 1206 abstracts were excluded, and 195 full-texts were screened. A total of 57 full-texts were excluded. Four articles (on the International Index of Erectile Functioning, and Female Sexual Functioning Index) were added which were not excluded in the second search, but not located in the current search. Resulting in a total of 142 articles fit for data extraction. One article was excluded during data extraction as the data it presented was already presented in another paper by the same authors. One article was excluded during data extraction as the PROM studied turned out to be a different PROM, but identically named, than a PROM we were investigating (Stoma Quality of Life Questionnaire).

Articles were included for: 1. Dyadic Adjustment Scale (50 articles) 2. Job Content Questionnaire (44 articles) 3. International Index of Erectile Function (25 articles) 4. De Jong-Gierveld Loneliness Scale (14 articles) 5. Female Sexual Function Index (7 articles)

4.5 Search 5 - "PROMs in care" database & manual search

A total of 7 PROMs had no hits in all of the previous searches. These concerned three Dutch-only questionnaires, one originally French questionnaire, and three international questionnaires.

The "PROMs in care" database revealed articles for three questionnaires:

1. Patiënt Specifieke Klachten (2 articles)
2. Vragenlijst Beleving en Beoordeling van de Arbeid (1 manual)
3. Stoma Quality of Life Questionnaire (1 article)

Other manual searches revealed articles for three questionnaires:

1. Alcohol Five Shot (2 articles)
2. Breast Cancer Patients' Needs Questionnaire (1 article)
3. Vragenlijst Gezinskenmerken (1 thesis)

The development article found for the Breast Cancer Patients' Needs Questionnaire did not include any interpretable data. The CARON had no interpretable search hits.

Chapter 5

Results for Psychological Wellbeing

This chapter covers the results for the Quality of Life area of Psychological Wellbeing.

5.1 Hospital Anxiety and Depression Scale

5.1.1 Description

The Hospital Anxiety and Depression Scale (HADS) is a PROM designed to measure anxiety and depression without the involvement of physical symptoms. The HADS is not population specific, and consists of two seven-item subscales measuring depression and anxiety respectively (Zigmond and Snaith, 1983).

A total of 14 validation studies (Zigmond and Snaith, 1983; Ibbotson et al., 1994; Razavi, 1990; Smith et al., 2002; Moorey et al., 1991; Zhong et al., 2012; Holly et al., 2003; Winters et al., 2013; Potter et al., 2009; Clark et al., 2011; Gopie et al., 2014; Herrmann, 1997; Harcourt et al., 2003; Nicholson et al., 2007) were extracted from the top 3 reviews (Aktas et al., 2015; Korus et al., 2015; Wakefield et al., 2015) concerning the HADS in cancer populations. Some data was directly extracted from a review (Wakefield et al., 2015) as data was presented completely.

5.1.2 Structural validity

Two studies (concerning cancer patients) presented Principal Component Analyses: Two factors explaining >45% variance. One item loaded on both factors ("I can feel relaxed").

One review presented a Confirmatory Factor Analysis: Two factors explaining about 50% of variance. Factor structure is stable across British and German samples, as well as subsample.

5.1.3 Internal consistency

Three studies (2 of which concerned cancer patients) reported a Cronbach's Alpha:

- HADS-A: .80 - .93
- HADS-D: .79 - .90

One study presented spearman correlations between items and total subscore of remaining items:

- Anxiety items: .41 - .76.
- Depression (he authors dropped one item with a score of .11): .30 - .60

5.1.4 Test-retest reliability

One review presented an overview of Pearson correlations of measurements across time:

- HADS-A:
  - 0-2 weeks: .84
  - >2-6 weeks: .73
  - >6 weeks: .70
- HADS-D:
  - 0-2 weeks: .85
  - >2-6 weeks: .76
  - >6 weeks: .7

5.1.5 Measurement Error

No data on measurement error was reported. As the article presenting test-retest reliability did not report standard deviations, measurement error could also not be calculated manually.

5.1.6 Known-groups comparison

Two studies found no differences between surgery groups. One study found no differences between patients with postoperative morbidity vs those without postoperative morbidity. One study found that patients who underwent breast reconstruction were more depressed than those that did not, but this effect disappeared at 12 months post-surgery.

5.1.7 Convergent validity

One review presented a correlation of the HADS-A with the HADS-D: .63. One study presented correlations with Psychiatric ratings:

- HADS-A: .74
- HADS-D .70

5.1.8 Divergent validity

No data on divergent validity was reported.

5.1.9 Criterion validity

One study presented false positives and false negatives, where a score of <=7 was classified as a non-case, a score of 8-10 as a doubtful case, and a score of >=11 as a definite case:

- HADS-D: 1 false positive, 0 false negatives
- HADS-A: 1 false positive, 1 false negative.

One study (concerning cancer patients) presented criterion validity for the screening of affective disorders:

- Total sample (N=284):
  - AUC .88
  - Cut-off score of >14:
    - Sensitivity: .8
    - Specificity: .76

        * Positive Predictive Value: .41
- Disease-free population (N = 88):
  - AUC .95
  - Cut-off score of >19:
    * Sensitivity .92
    * Specificity .95
    * Positive Predictive Value: .72.
- Stable disease population (N = 113):
  - AUC .89
  - Cut-off score >15:
    * Sensitivity .83
    * Specificity .78
    * Positive Predictive Value: .42
- Under Treatment Population (n=146):
  - AUC .84
  - Cut-off score of >15:
    * Sensitivity .85
    * Specificity .77
    * Positive Predictive Value: .47.

One study (concerning cancer patients) presented criterion validity for the screening of depression: - Cut-off score of >=19: - Sensitivity .70 - Specificity .75 - Positive Predictive Value (score 26) .50 - Positive Predictive Value (score 12) .28.

One study (concerning cancer patients) presented criterion validity of the HADS-total score for the screening of adjustment disorder - Cut-off score of >=13: - Sensitivity .75 - Specificity .75 - Positive Predictive Value (score 15) .90 - Positive Predictive Value (score 4) .74.

One study (concerning cancer patients) presented criterion validity for the screening of depression and anxiety: - HADS-D: - AUC .64 - Cut-off score not reported: - Sensitivity .70 - Specificity .48 - HADS-A: - AUC .62 - Cut-off score not reported: - Sensitivity .70 - Specificity .41

One review gave recommendations for use of HADS for screening for anxiety and depression. The HADS was recommended for use in pre-treatment/diagnosis patients, active treatment patients, and post-treatment patients. The HADS was not recommended for palliative patients.

5.1.10 Responsiveness

Three studies (concerning breast cancer patients) found a lower HADS-A score after breast reconstruction. One study (concerning breast cancer patients) found an increase in HADS-D after DIEP flap reconstruction. Two studies (concerning breast cancer patients) found no difference in HADS-D score after breast reconstruction. One study (concerning breast cancer patients) found no difference in HADS-A after breast reconstruction.

5.1.11 Conclusion

The Hospital Anxiety and Depression scale is a well established PROM for use in the cancer population, with evidence for measurement properties except for *measurement error*, and *divergent validity*. This PROM is more than suitable for use in Oncokompas 2.0. The information on criterion validity can be used to inform our algorithms.

5.2 Cancer Worry Scale

5.2.1 Description

The Cancer Worry Scale measures the fear of cancer recurrence, designed to differentiate between high- and low-fearful survivors (Custers, 2014).

A total of 2 articles (Custers, 2014; Custers et al., 2016) concerning cancer populations were included for data extraction.

5.2.2 Structural validity

One study performed an Exploratory Factor Analysis and found 1 factor with eigenvalue > 1. Missing values were handled by case-wise deletion.

5.2.3 Internal consistency

Two studies reported Cronbach's Alpha: .87 - .89. One study reported missing item percentages, dealt with by case-wise deletion. Handling of missing items other study unknown.

5.2.4 Test-retest reliability

No data on test-retest reliability was reported.

5.2.5 Measurement Error

No data on measurement error was reported. As no test-retest reliability was reported, measurement error could also not be calculated manually.

5.2.6 Known-groups comparison

One study found small significant correlations between age and cancer worry (r = -0.198), and time since diagnosis and cancer worry (r = -0.213). Missing items were handled through case-wise deletion.

5.2.7 Convergent validity

One study found a correlation of .85 with the Fear of Cancer Recurrence Inventory. Another study found correlations with the Cancer Acceptance Scale (r = .75) and CIS-Fatigue scale (r = .34). The latter study reported missing percentages and dealt with them through case-wise deletion. Handling of missing items in the former is unknown.

5.2.8 Divergent validity

One study found a small negative correlation with the Cancer Empowerment Questionnaire (-.22). Handling of missing items is unknown.

5.2.9 Criterion validity

Two studies reported the AUC: .88 - .92.

One study reported screening on cut-off score 11 vs 12:

- Sensitivity .96
- Specificity .56
- Positive Predictive Value .60
- Negative Predictive Value .96

Two studies report distinguishing fearful vs non-fearful at cut-off score 13 vs 14:

- Sensitivity .77 - .86
- Specificity .81 - .87
- Positive Predictive Value .73 – .76
- Negative Predictive Value .84 – .93.

Both studies handled missing values by case-wise deletion.

5.2.10 Responsiveness

No data on responsiveness was reported.

5.2.11 Conclusion

The Cancer Worry Scale has not seen much research, but the research that has been published show indications for good measurement properties, except for *test-retest reliability*, *measurement error*, and *responsiveness*, which were not studied. More research is also needed for structural validity. The CWS is suitable for use in Oncokompas 2.0, although more research would be preferrable.

Furthermore, two different cut-off scores were investigated. This information can be used to optimize the algorithms in Oncokompas 2.0.

5.3 SF-36

5.3.1 Description

The 36-Item Short Form Survery (SF-36) consists of eight generic quality-of-life measures concerning physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions (Ware and Sherbourne, 1992; McHorney et al., 1993).

A total of 21 validation studies (Nicholson et al., 2007; Luo et al., 2005; Walker et al., 2009; Temple et al., 2009; Rowland et al., 2000; Wilkins et al., 2000; Atisha et al., 2008; Pusic et al., 1999; Manganiello et al., 2011; Girotto and Schreiber, 2003; Mullan et al., 2007; Nelson et al., 2013; Tønseth et al., 2008; Veiga et al., 2004; Brandberg et al., 2000; Edsander-Nord A et al., 2001; Elder et al., 2005; McHorney et al., 1993; Bousquet et al., 1994; Bird et al., 2010) were extracted from the top 3 reviews (Korus et al., 2015; Osborne2012; Scarpa et al., 2011) concerning the SF-36 in cancer populations. Some data was directly extracted from a review (Scarpa et al., 2011) as data was presented completely.

5.3.2 Structural validity

One study performed a Principal Component Analysis: Two overarching component of general health (physical & mental) explained 70% of variance of all subscales.

5.3.3 Internal consistency

Three studies (two of which concerning cancer populations) reported Cronbach's Alpha:

- Global Quality of Life .89 - .91
- Physical functioning .62 - .98
- Physical role .87
- Emotional role .76 - .86
- Emotional wellbeing .19 - .81
- Social functioning .83 - .90
- Fatigue .82 - .86
- Pain 75 - .87

5.3.4 Test-retest reliability

No data on test-retest reliability was reported.

5.3.5 Measurement Error

No data on measurement error was reported. As no test-retest reliability was reported, measurement error could also not be calculated manually.

5.3.6 Known-groups comparison

Two studies found no differences between breast surgery groups.

Four studies reported significant differences between breast surgery groups. Particularly higher mental health and general scores on reconstructive patients vs non-reconstructive patients, and immediate resontruction patients scored lower than delayed reconstruction patients.

One study found that chronic pain patients had higher scores on pain questions. Body pain, physical health, mental health, and overall score were lower in chronic pain patients.

One study found that patients with serious medical conditions scores significantly lower on all scales compared to patients with minor medical conditions. Psychiatric patients showed significantly lower scores on mental health, role-emotional, and social functioning.

One study found that severity of asthma as assessed by the clinical score of Aas was significantly correlated was significant for all scales. With more severe asthma scoring worse on the SF-3.

One study found that pooled scores for physical function, physical role, and social function after esophagectomy were similar to sex- and age-matched United States norms, whereas the pooled scores for physical function, vitality and general health perception were lower than the relevant norms ($p = 0.005$, $p < 0.001$ and $p = 0.006$, respectively). In contrast, scores for bodily pain and mental health in long-term survivors after esophagectomy were higher than the relevant norms ($p = 0.08$ and $p = 0.02$, respectively).

5.3.7 Convergent validity

One study found correlations with the EORTC QLQ-C30, where the strength of Spearman's correlations for eight pairs of QLQ-C30 and SF-36 scales measuring similar dimensions of HRQoL ranged from 0.35 between QLQ-C30 role functioning and SF-36 role-emotional scales to 0.67 between QLQ-C30 pain and SF-36 bodily pain scales.

One study reported correlations with the SQ-F: Functional capacity .34, emotional limitations .50, mental health .34.

5.3.8 Divergent validity

No data on divergent validity was reported.

5.3.9 Criterion validity

No data on criterion validity was reported.

5.3.10 Responsiveness

Five studies found reductions in SF-36 scales after breast reconstruction.

In particular, significant reduction on role limitation caused by emotional / physical health progrems, general health, and pain, in one study.

Another found significant increase in emotional, vitality, general mental health, and social functioning post-operation for immediate reconstruction. Significant increase in general mental health post-operation for delayed reconstruction.

Another found that the scores were significantly higher on role emotional and mental health at 3-month postoperative, on health perception and role physical at 6-month and on physical function, pain, health perception and social function at 12-month post-operation.

Another found that vitality, social functioning, emotional role, and mental health significantly increased.

The last study found significant improvement in mental health, role-emotional, vitality, social functioning and general health for both TRAM flap as Implant surgery patients.

5.3.11 Conclusion

While many studies investigated the SF-36, providing evidence for many measurement properties, a couple of important measurement properties were not investigated or not investigated thoroughly enough: *test-retest reliability*, *measurement error*, *structural validity*, *divergent validity*, and *criterion validity*.

Only the cognitive subscale is used in the Oncokompas 2.0. This subscale is most likely suitable for use, but alternatives with better defined cutoff scores should be explored.

Chapter 6

Results for Physical Wellbeing

This chapter covers the results for the Quality of Life area of Physical Wellbeing.

6.1 Patiënt Specifieke Klachten

6.1.1 Description

The Patiënt Specifieke Klachten is a Dutch-specific questionnaire designed to determine the physical functioning of patients with back issues. A patient is instructed to select three to five most important complaints concerning physical activities. The intensity of the complaint is then scores on an eleven-point numeric rating scale (Beurskens et al., 1999).

A total of 2 articles (Beurskens et al., 1999; Van der Wees et al., 2012) concerning non-cancer populations were included for data extraction.

6.1.2 Structural validity

No data on structural validity was reported.

6.1.3 Internal consistency

No data on internal consistency was reported.

6.1.4 Test-retest reliability

No data on test-retest reliability was reported.

6.1.5 Measurement Error

No data on measurement error was reported. As no test-retest reliability was reported, measurement could also not be calculated manually.

6.1.6 Known-groups comparison

No data on known-groups comparison was reported.

6.1.7 Convergent validity

One study reported correlations between .26 - .49 with the Ankle Function Score. Handling of missing items unknown.

6.1.8 Divergent validity

No data on divergent validity was reported.

6.1.9 Criterion validity

No data on criterion validity was reported.

6.1.10 Responsiveness

One study (RCT) showed that in their improved group, some patients showed more improvement on the main complaints than on the other outcome measures. This means that the main complaints were the most sensitive to change. One study (not RCT) shows a Cohen's d of 2.14 and Standardized Response Mean of 1.71 between before and after treatment.

6.1.11 Conclusion

Very little is known about the Patiënt Specifieke Klachten. No judgments regarding good or bad validity and reliability can be made, given the available information. Investigating alternative PROMs is recommended for use in Oncokompas 2.0. If no suitable alternatives can be identified, more research into the PSK is necessary to justify validity, reliability, and cut-off points.

6.2 Insomnia Severity Index

6.2.1 Description

The Insomnia Severity Index is a seven-item screening tool for insomnia (Bastien, 2001).

A total of 3 articles (Fillion et al., 2003; Kotronoulas et al., 2011; Savard et al., 2005) concerning cancer populations were included for data extraction.

6.2.2 Structural validity

One study performed a Principal Component Analysis in 785 breast, and 849 prostate cancer patients seperately. In both analyses a two-factor structure was found comprising the same items. Factor 1: severity of sleep difficulties; factor 2: impact of sleep difficulties. While not explicitly stated, missing data was most likely handled with case-wise deletion.

6.2.3 Internal consistency

Two studies reported Cronbach's Alpha:

- .86 (327 prostate cancer patients)
- .91 (785 breast cancer patients)
- .89 (849 prostate cancer patients).

While not explicitly stated, missing data was most likely handled with case-wise deletion.

6.2.4 Test-retest reliability

One study reported test-retest correlation (not ICC):

- .83 after 1 month (60 prostate or breast cancer patients)
- .77 after two months (117 prostrate or breast cancer patients)
- .73 after 3 months (83 prostrate of breast cancer patients).

While not explicitly stated, missing data was most likely handled with case-wise deletion.

6.2.5 Measurement Error

No data on measurement error was reported. As the article presenting test-retest reliability did not report standard deviations, measurement error could also not be calculated manually.

6.2.6 Known-groups comparison

No data on known-groups comparison was reported.

6.2.7 Convergent validity

One study reported correlations between .24 - .39 with subscales of Multidimensional Fatigue Inventory.

One study reported high correlations with Greek Pittburg Sleep Quality Index (-.75, and .81 for different subscales).

One study reported moderate to high correlations with sleep diary (.26 - .54), and polysomnography (.26 - .50). Moderate correlation with Dysfunctional Beliefs and Attitudes about Sleep scale (.39). Moderate correlation with ISI filled in by clinician (.35), and high correlation with ISI filled in by spouse (.53).

Not known how missing data were handled. ### Divergent validity One study reported low correlations between ISI and constructs from the EORTC QLQ-C30 that are not supposed to be directly related to insomnia, such as physical function (r=-.3, p=.03), role functioning (r=-.12, p=.38), cognitive functioning (r=-.14, p=.32), and social functioning (r=-.16, p=.4). Not known how missing data were handled.

6.2.8 Criterion validity

One study reported on criterion validity:

- AUC .86 (.81 - .91)
- Cutoff score for clinically significant is 8:
  - Sensitivity .947,
  - Specificity .474,
  - Positive Predictive Value .677,

  - Negative Predictive Value .885
- Cutoff score for clear insomnia syndrome is 15:
  - Sensitivity .513
  - Specificity .907
  - Positive Predictive Value .866
  - Negative Predictive Value .615

How missing values were handled is unknown.

6.2.9 Responsiveness

One RCT in 51 breast cancer patients was reported. Statistically significant reduction in ISI total score ($t(50)=9.1$, $p <.0001$) from pre-treatment (M=13.7, SD=4.9) to post-treatment (M=7.9, SD=4.6). Not explicitly stated, but missing data was most likely handled with case-wise deletion.

6.2.10 Conclusion

The Insomnia Severity Index is studied by a handful of studies in the cancer population, but shows evidence of good measurement properties, except for *measurement error*, and *known-groups comparison*, which were not investigated. This PROM is suitable for use in Oncokompas 2.0, and the studied cutoff points can be used to inform algorithms.

6.3 6-item Female Sexual Function Index

6.3.1 Description

The 6-item Female Sexual Function Index is a shortened version of the 19-item FSFI (Rosen et al., 2000; Isidori et al., 2010).

A total of 9 articles (Bartula and Sherman, 2015; Baser et al., 2012; Chedraui et al., 2012; DeRogatis et al., 2010; Flynn et al., 2013; Hugo et al., 2011; Isidori et al., 2010; Pérez-López et al., 2012; Rosen et al., 2000) were included, of which two (Bartula and Sherman, 2015; Baser et al., 2012) concerned cancer populations.

6.3.2 Structural validity

FSFI-19: One study (in a cancer population) performed PCA and extracted 5-factor solution (desire/arousal, lubrication, orgasm, satisfaction, and pain). Two studies performed CFA. One found 5-factor solution but extra factor was added due to theoretical considerations. Other found six factors but dropped item 14.

Handling of missing items unknown.

6.3.3 Internal consistency

A total of 7 studies reported Cronbach's Alpha, with one study each for the FSFI-19 and FSFI-6 which consisted of a cancer population. FSFI-19 (3 studies; 4 time points):

- Desire .90 - .93
- Arousal .92 - .95
- Lubrication .94 - .96
- Orgasm .90 - .94
- Satisfaction .85 - .89

- Pain .91 - .94.

Handling of missing items unknown.

FSFI-6 (4 studies): .789 - .91. One study reported missing items and handled them with case-wise deletion. Handling of missing items for other 3 studies unknown.

6.3.4 Test-retest reliability

A total of 3 studies reported test-retest reliability, of which one for the FSFI-19 consisted of a cancer population. FSFI-19 (2 studies; 2-4 weeks apart; correlations)

- Desire .83 - .86
- Arousal .82 - .85
- Lubrication .78 - .86
- Orgasm .80
- Satisfaction .76 - .83
- Pain .75 - .79
- Overall score (1 study) .88.

Handling of missing items unknown.

FSFI-6 (1 study; 18-24 days; correlation): .95 (.935 - .964). 12% missing items. Handling unknown.

6.3.5 Measurement Error

No data on measurement error was reported. None of the studies presenting test-retest reliability reported appropriate standard deviations, as such measurement error could not be calculated manually.

6.3.6 Known-groups comparison

FSFI-19: One study found no significant differences between treatment groups (chemotherapy and radiation therapy). Missing items handled by casewise deletion. One study found significant differences between patients with sexual dysfunction and controls for all domains except for lubrication and arousal. Handling of missing items unknown.

6.3.7 Convergent validity

FSFI-19: One study found that the FSFI scores tended to be correlated negatively with depression (CES-D), distress (IES), menopausal symptoms (MSCL), and reproductive concerns (RCS), and generally were correlated positively with QOL (FACT-G, FACT-Cx), functional health status (the SF-12 PCS and MCS), and relationship satisfaction (ADAS). Missing items were handled by casewise deletion.

One study found significant correlations with the Women's Sexual Interest Diagnositc Interview-SF and a Daily Log of Sexual Activities. One study found significant correlatioons with the CARES and WHOQOL-100. One study found significant correlations with the PROMIS sexual domain.

FSFI-6: One study found significant correlations with coital frequency, educational level, partner educational level, age, partner age, waist circumference, hot flush intensity. One study found significant correlations with the Menopause Rating Scale and the HADS.

Handling of missing items unknown.

6.3.8 Divergent validity

FSFI-19: One study found no correlation with the DAS. One study found no to modest correlations with the Locke-Wallace Marital Adjustment Test. One study found correlations between -.33 to .42 with the BIS, FAS, DASS-21, SF-20, and RDAS.

Handling of missing items unknown.

6.3.9 Criterion validity

FSFI-6: One study found an AUC: .984 (.951 - .997). Cutoff of 19: sensitivity 96.1; specificity 90.9.

They provided a logistic regression analysis that found an increase of one unit in a subject's FSFI-6 score decreased the chances of having FSD (against not having FSD) by about half (odds ratio = 0.41; 95% CI 0.31–0.54), with 91.87% cases classified correctly.

6.3.10 Responsiveness

No data on responsiveness was reported.

6.3.11 Conclusion

The Female Sexual Function Index 19-item has been studied in a cancer population, which showed similar results as for other populations. Given this, we can assume that the shortened 6-item version will also correspond in the cancer population. The FSFI-6 has seen some research, but information was lacking for *structural validity*, *measurement error*, *known-groups comparison*, *divergent validity*, and *responsiveness*. However, *criterion validity* was investigated, giving information related to a cutoff points for clinical diagnosis.

The FSFI-6 is most likely suitable for use in Oncokompas 2.0, but more research (particularly into *structural validity*) in the cancer population is recommended. The studied cutoff points can be used to inform algorithms.

6.4 5-item International Index of Erectile Function

6.4.1 Description

The 5-item International Index of Erectile Function is a shortened version of the 15-item IIEF (Rosen et al., 1997, 2000).

A total of 27 articles (Aslan et al., 2011; Bayraktar and Atun, 2012, 2013; Bushmakin et al., 2014; Cappelleri et al., 1999; Coyne et al., 2010; Dargis et al., 2013; Flynn et al., 2013; García-Cruz et al., 2011; Gonzáles et al., 2013; Hwang et al., 2010; Kriston et al., 2008; Lim et al., 2003; Lin et al., 2016; Mahmood et al., 2012; Mulhall et al., 2008; Parisot et al., 2014; Quek et al., 2002; Quinta Gomes and Nobre, 2012; Rosen et al., 2011, 1997; Rubio-Aurioles et al., 2009; Saffari et al., 2016; Serefoglu et al., 2008; Tang et al., 2015; Utomo et al., 2015; Wiltink et al., 2003) were included, of which two (Lin et al., 2016; Parisot et al., 2014) concerned cancer populations.

6.4.2 Structural validity

IIEF-15: PCA (6 studies): The PCA analyses were inconsistent in their findings. Two analyses found the theorized 5 components, two analyses found 4 components, and two analyses found 2 components.

CFA (3 studies): One study found acceptable fit (CFI > .92) for 5-factor structure on 3 separate time points. One study found an acceptable fit for 2-factor model (RMSEA .077; CFI .94; GFI .93; AGFI .9) as well as 5-factor model (RMSERA .067; CFI .96; GFI .95; AGFI .92).

One study found that original 5-factor model had acceptable fit (GFI = .889; TLI = .933; CFI = .949; SRMR = .045; RMSEA = .09) as did the 4-factor model (GFI = .849; TLI = .908; CFI = .926; SRMR = .049; RMSEA = .107). The 2-factor model had non-acceptable fit (CFI = .783; TLI = .854; CFI = .876; SRMR = .064; RMSEA = .134), as did the 1-factor model (GFI = .743; TLI = .812; CFI = .839; SRMR = .072; RMSEA = .152). CAIC favored the original 5-factor model (512.68).

Rasch (1 study): Monotonical increase across IIEF; one local dependency in IIEF; no substantial DIF in IIEF.

6.4.3 Internal consistency

A total of 14 studies reported Cronbach's Alpha.

IIEF-15 (9 studies; 11 measurements):

- Erectile function .76 - .931
- Orgasmic function .74 - .917
- Sexual desire .72 - .921
- Intercourse satisfaction .66 - .916
- Overall satisfaction .73 - .96.

One study imputed 1.1% missing items. Two studies removed participants with missing items.

IIEF-5 (5 studies; 6 measurements): .64 - .882. Two studies removed participants with missing items.

One study used Item Response Theory on the IIEF-15: Person seperation reliability .66; item seperation reliability .99; person seperation index >1.40; item seperation index >9.63.

6.4.4 Test-retest reliability

A total of 11 studies reported test-retest reliability

IIEF-15 (6 studies; correlations):

- Total .64 - .909
- Erectile function (3 studies) .55 - .943
- Orgasmic function (3 studies) .64 - .834
- Sexual desire (3 studies) .14 - .865
- Intercourse satisfaction (3 studies) .71 - .81
- Overall satisfaction (3 studies) .77 - .90.

IIEF-15 (3 studies; ICC):

- Total .92
- Erectile function .77 - .88
- Orgasmic function .75 - .82
- Sexual desire .82 - .87
- Intercourse satisfaction .79 - .89
- Overall satisfaction .82 -.85

IIEF-5 (2 studies; ICC): .88.

One study provided intra-rater reliability (Kappa) for the IIEF-15: .60 - .714.

Three studies excluded participants with missing data. Handling of missing items for other studies is unknown.

6.4.5 Measurement Error

IIEF-15: One study reported SEM: Erectile functio: 0.8 – 1.2; orgasmic function 0.5 – 0.6; sexual desire = 0.3; intercourse satisfaction = 0.4 – 0.6; overall satisfaction = 0.3 – 0.4. One study reported Limits of Agreement of -3.7 - 6.4.

IIEF-5: One study reported limits of agreement of 10.1.

Handling of missing items unknown.

Measurement error could be calculated for one study reporting test-retest reliability on the IIEF-15:

Standard Error of Measurement:

- Erectile function 3.59
- Orgasmic function 1.34
- Sexual desire 0.69
- Intercourse satisfaction 1.88
- Overall satisfaction 2.96

Smallest Detectable Change:

- Erectile function 9.94
- Orgasmic function 3.70
- Sexual desire 1.90
- Intercourse satisfaction 5.21
- Overall satisfaction 8.21

Seven studies presenting test-retest reliability did not report appropriate standard deviations, as such measurement error could not be calculated manually for these studies.

6.4.6 Known-groups comparison

IIEF-15 (6 studies): One study found significant differences for age on two different time points, where older men had lower scores. Four studies showed differences between ED patients and controls, with ED patients scoring lower. One study found significant differences between patients with differing ED severity.

IIEF-5 (2 studies): One study found no significant differences for age on one time point, but did find differences on another time point, where older men had lower scores. One study found differences between ED patients and control, with ED patients scoring lower.

Two studies applied case-wise deletion for missing values. Handling of missing items for other studies is unknown.

6.4.7 Convergent validity

IIEF-15 (8 studies): Three studies found significant correlations with clinician ratings of erectile function. Two studies found significant correlations with the Erection Hardness Score. Separate studies found significant correlations with the Sexual Experience Questionnaire, Female Assessment of Male Erection, Male

Fenital Self-Image Scale, and PROMIS sexual domain. Two studies reported missing item percentages, but no studies reported handling.

IIEF-5 (2 studies): One study found significant kappa (.375) with the Erection Hardness Grading Scale. One study found significant correlations with the Erection Hardness Score (.61 - .79), and with the Quality of Erection Questionnaire (.70). The former study had no missing values. Handling of missing items of latter study is unknown.

6.4.8 Divergent validity

IIEF-15 (2 studies): One study found no association with the DAS and SF-12. One study found no association with the Locke-Wallace Marital Adjustment Test or a social desirability scale. Handling of missing items unknown.

IIEF-5 (1 study): One study found no association with the DAS and SF-12.

6.4.9 Criterion validity

IIEF-15 (4 studies): Three studies reported AUC for detecting Erectile Dysfunction: .83 - .97. Two studies tested the optimal cut-off score 25:

- Sensitivity .87 - .97
- Specificity .75 - .88
- PVP .85 - .89
- PVN .97.

One study tested the cut-off score 17:

- Sensitivity 85
- Specificity 75.

IIEF-15 Erectile Functioning subscale (1 study): AUC .966. Cutoff 25:

- Sensitivity 98
- Specificity 78.9.

IIEF-5 (1 study): AUC .9707. Cut-off 22 (official cut-off):

- Sensitivity 100%
- Specificity 0.6%.

Cut-off 15.5:

- Sensitivity 97.11% (95.16-98.82)
- Specificity 85.71% (73.89; 95.07).

Missing item handling is unknown.

6.4.10 Responsiveness

IIEF-15 (3 studies): One study found significant changes after treatment on all domains. One study found significant changes after treatment for erectile function and orgasm function. One study found significange changes after treatment for ejaculation frequency and orgasm function.

IIEF-15 Erectile Functioning subscale (1 study): MCID was calculated using ANOVA-based techniques and ROC-based techniques. Anova based (development sample vs validation sample): mild 2.79 (development) & 3.45 (validation); moderate 7.21 (development) & 8.11 (validation); severe 12.38 (development) & 11.96 (validation) (table 2). ROC-based: mild 2, moderate 5, severe 7.

IIEF-5 (1 study): The change in IIEF-5 score in treated patients after 6 months was 2.2 +- 3.9 compared to -0.6 +- 2.8 in untreated patients (p = 0.007).

Missing item handling is unknown.

6.4.11 Conclusion

The International Index of Erectile Function has been studied in a cancer population, which showed similar results as for other populations. Given this, we can assumed that the shortened 5-item version will also correspond in the cancer population. The IIEF-5 has seen some research, but information was lacking for *structural validity*.

The IIEF-5 is most likely suitable for use in Oncokompas 2.0, but more research (particularly into *structural validity*) in the cancer population is recommended. The studied cutoff points can be used to inform algorithms.

6.5 Body Image Scale

6.5.1 Description

The Body Image Scale is a ten-item measurement tool designed to measure body image in cancer patients (Hopwood et al., 2001).

A total of 8 articles (Anagnostopoulos and Myrgianni, 2009; Gómez-Campelo et al., 2014; Hopwood et al., 2001; Khang et al., 2013; Moreira et al., 2010; Rhondali et al., 2014; van Verschuer et al., 2015; Whistance et al., 2010) concerning cancer populations were included for data extraction.

6.5.2 Structural validity

Five studies with eight populations found a one-factor solution using exporatory factor analyses and confirmatory factor analyses.

Three studies with three populations found a two-factor solution using exporatory factor analyses and confirmatory factor analyses.

One study used multitrait-item scaling which suggested dropping 1 item from the scale.

Handling of missing data was unknown, or by case-wise deletion. One study (1 factor solution for 3 populations, 2 factor solution for 1 population) imputed with the mean of the item.

6.5.3 Internal consistency

Seven studies with 8 populations reported the Cronbach's Alpha of the total score: .78 - .967

One study reported the Cronbach's Alpha of the attractiveness scale .92, the satisfaction scale .87 , and general body concerns .81.

Most studies did not report missing value handling. One study imputed scores with the mean of the missing items (Alpha .919 - .916).

6.5.4 Test-retest reliability

One study reported a test-retest correlation (not ICC) after 6 months of .67 (N=32; South Korea). One study reported a Rho after 2 weeks of .92 (N=209; Netherlands). One study reported an ICC after 2 weeks of .89 (N=19; United Kingdom).

Missing items were handled through case-wise deletion.

6.5.5 Measurement Error

No data on measurement error was reported. Two of the studies presenting test-retest reliability did not report appropriate standard deviations, as such measurement error could not be calculated manually for these publications.

One study presented test-retest reliability and standard deviations, as such the Standard Error of Measurement (4.00) and Smallest Detectable Change (11.08) could be calculated.

Missing items were handled through case-wise deletion.

6.5.6 Known-groups comparison

Two studies found significant relationships with age (r = -.225 - -.643). Four studies found that patients receiving Breast Conserving Treatment had significantly lower scores than patients receiving a mastectomy. One study found that patients receiving a Wide Local Excision had significantly lower scores than patients receiving a mastectomy. One study found that patients with a stoma scores significantly higher than patients without a stoma.

Four of the studies reported missing item percentages. One of the studies (regarding Wide Local Excision) imputed missing scores. Handling of missing items for the remaining four studies is unknown.

6.5.7 Convergent validity

Significant correlations were reported with the:

- General Health Questionnaire-28 (.260 - .606, 1 study)
- Rosenberg Self-esteem Scale (-1.22 [ns] - -.733, 2 studies)
- Beck Depression Inventory (.832, 1 study)
- Beck Anxiety Inventory (.564, 1 study)
- EORTC QLQ-C30 (-.632, 1 study)
- Body-esteem Scale for Adolescents and Adults (-.322, 1 study)
- HADS-A (.501 - .522, 2 studies)
- HADS-D (.422 , 2 studies)
- WHOQOL-bref (-.223 - -.51, 2 studies)
- Experience of Shame Scale (.68, 1 study)
- Derriford Appearance Scale-24 (.75, 1 study)
- Appearance Schemas Inventory-Revised (.248 - .40, 2 studies)
- Perceived Stress Scale (.371, 1 study), Physical Distress Score (.356, 1 study)
- Total Symptom Distress Score (.416, 1 study)
- Overall appearance item from Multidimensional Body-Self Relations Questionnaire (-.449, 1 study).

Handling of missing items unknown.

6.5.8 Divergent validity

1 study reported divergent validity with EORTC QLQ-C30. All scales had correlations <.40 except for emotional functioning (.45). Missing values were handled with casewise deletion.

6.5.9 Criterion validity

No data on criterion validity was reported.

6.5.10 Responsiveness

One study (not RCT, N=56), found a significant increase in total sample and two subgroups over 14 weeks.

One study (not RCT, N=17) found no significant increase before and after surgery.

Missing items were handled with casewise deletion.

6.5.11 Conclusion

The Body Image Scale has been well researched in cancer populations, only missing information on *criterion validity*. This BIS is suitable for use in Oncokompas, but research is needed to establish cutoff points.

6.6 EORTC QLQ-C30

6.6.1 Description

The European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 is a 30-item instrument to assess the quality of life of cancer patients (Bjordal et al., 2000).

A total of 11 validation studies (Winters et al., 2013; Potter et al., 2009; Harcourt et al., 2003; Conroy et al., 2004; Uwer et al., 2011; Koukouli et al., 2009; Luo et al., 2005; Wan et al., 2008; Maringwa et al., 2011a,b; Kvam et al., 2011) were extracted from the top 3 reviews (Wong et al., 2015; Korus et al., 2015; Bedard et al., 2012) concerning the EORTC QLQ-C30 in cancer populations.

6.6.2 Structural validity

One study reported a Principal Component Analysis: Six factors extracted, but only two subscales were consistent with the original instrument (emotional functioning and nausea).

One study reported a Strutual Equation Model: the structure of the QLQ-C30 can be grouped into 15 domains (is the same as theoretical structure), with acceptable goodness of fit:

- chi-square = 762.28 ($p < 0.0001$)
- root mean square error of approximation (RMSEA) = 0.054, 90% confidence interval = (0.049–0.059)
- non-normed fit index = 0.972
- comparative fit index (CFI) = 0.980
- standardized root mean square residual (SRMR) = 0.037

Handing of missing items is unknown.

6.6.3 Internal consistency

Four studies with five populations reported the Cronbach's Alpha's for:

- Physical functioning .73 - .91
- Role .86 - .90
- Emotional functioning .79 - .87
- Cognitive functioning .19 - .74
- Social functioning .74 - .80
- global QoL .87 - .92.

Three studies with four populations also reported the Cronbach's Alpha's for:

- Fatigue .78 - .96
- Nausea .68 - .91
- Pain .72 - .97

One study handled missing items with casewise deletion. Missing item handling of other studies is unknown.

6.6.4 Test-retest reliability

Two studies reported ICCs for 1-2 days and 2 weeks apart, respectively:

- Physical functioning .87 - .89
- Role .80 - .83
- Emotional functioning .73 - .84
- Cognitive functioning .79 - .82
- Social functioning .80 - .84
- Global QoL .33 - .89
- Fatigue .82 - .84
- Nausea .43 - .86
- Pain .82 - .87
- Dyspnoea .63 - .87
- Insomnia .76 - .85
- Appetite loss .74 - .77
- Constipation .54 - .80
- Diarrhea .33 - .75
- Financial impact .75 - .84.

Handling of missing items is unknown.

6.6.5 Measurement Error

Two studies reported a Standard Error of Measurement for:

- Physical functioning 7 – 7.7
- Role 14 – 14.2
- Global QoL 9 – 9.4
- Fatigue 10 – 11

One study reported a Standard Error of Measurement for:

- Social functioning 10
- Pain 12
- Cognitive functioning 11.3

Two studies reported Minimal Important Difference towards physical status for:

- Improvement:
  - Physical functioning 5.6 – 9
  - Role 14 – 14.3
  - Global QoL 7.3 – 9
  - Fatigue -12.4 – 14
- Deterioration:
  - Physical 4 – 8.5
  - Role 5 – 12.3
  - Global QoL 3.5 – 4
  - Fatigue -8.9 – 14

One study reported Minimal Important Difference towards physical status for:

-Improvement: - Social functioning 5 - Pain 16 - Cognitive functioning 7.6 - Deterioration: - Social functioning 7 - Pain 3 - Cognitive functioning 8.2

One study reported Minimal Important Difference towards self-report HRQoL improvement:

- Improvement:
  - Global QoL 7.6
- Deteriorated:
  - Global QoL -12.1

Unknown how missing items were handled.

6.6.6 Known-groups comparison

One study found no significant differences between patients with postoperative morbidity vs patients without.

One study found significant differences for nine out of fifteen domains (including all symptom domains except insomnia and diarrhea) between surgery vs chemotherapy patients. Surgery had higher QoL.

Unknown how missing items were handled.

6.6.7 Convergent validity

One study found correlations with the FACT-G, where in general, substantial when scales related to the same QoL domain, and low when they related to different domains. A high correlation was observed between FACT-G Physical Well-Being and all the function scales of QLQ-C30. The social domains of FACT-G and QLQ-C30 were poorly correlated, but the emotional subscales were highly correlated. The FACT-G global score was correlated with all QLQ-C30 domains.

One study found significant correlations with clinician's physical functioning assessment, where all six functioning sub-scales were negatively related to PS score before and after treatment. Pearson's r was highly significant ($P < 0.001$) and very strong for physical and role scales ($r > -0.7$). Substantial correlations ($r > 0.40$) were also noted between the global QoL, social and cognitive sub-scales and PS, and statisti cally significant, but relatively weak ($r < 0.40$) for the emotional scale ($r = -0.161$ before treatment and $r = -0.308$ after treatment). Also, PS was strongly associated with fatigue ($r = 0.622$ before treatment and $r = 0.704$ after treatment), but weakly with the other symptoms (subscales or items).

One study found significant correlations with the SF-36, where the strength of Spearman's correlations for eight pairs of QLQ-C30 and SF-36 scales measuring similar dimensions of HRQoL ranged from 0.35 between QLQ-C30 role functioning and SF-36 role-emotional scales to 0.67 between QLQ-C30 pain and SF-36 bodily pain scales.

Unknown how missing items were handled.

6.6.8 Divergent validity

No data on divergent validity was reported.

6.6.9 Criterion validity

No data on criterion validity was reported.

6.6.10 Responsiveness

One study found general worsening on QLQ-C30 at 3 months after treatment, but improvements at 12 months.

One study found significant improvements in the social, emotional, future perspective, and body image at 6 months and 12 months after treatment.

One study found the physical, role, emotional and cognitive function and the fatigue scale of the QLQ-C30 appeared to be responsive (after 4 weeks) with values of the indicators (SRM and ES) greater than 0.5 reflecting moderate ability to detect an effect of chemotherapy treatment. It also found the SRM for the global QoL/GHS score was 0.21, reflecting a minimal ability to detect an effect of radiotherapy treatment on clinical change, as well as the pain (SRM = 0.30) and constipation (SRM = 0.50) subscales or items.

One study found that the direction of change of QOL after treatment (21 days) differed across the cancer types; the change was negative for the group of breast and lung cancer, but positive for the group of colorectal, stomach, and head and neck cancer. It can be seen from Table 7 that for the breast and lung cancer groups, QOL score changes after treatment were of statistical significance on all domains except for QL, with SRM being >0.50 for all domains except for QL (0.10) and FI (0.28). Table 8 shows that for the colorectal, stomach, and head and neck cancer groups, all domain scores have statistically significant changes after treatment except for EF, CF, SF, AP, DI, and FI, with SRM being <0.50 for all domains except PF (0.73).

One study found that patients who improved reported significantly ($P < 0.01$) higher scores at 3 months with SRMs of .32. In patients rating themselves as unchanged, mean score changes clustered around zero. Patients who deteriorated reported lower scores at 3 months compared with T1: The global QL scale of the EORTC QLQ-C30 was the most responsive in deteriorating patients (SRM 0.57).

One study used casewise deletion for missing values. For the other studies the handling of missing items is unknown.

6.6.11 Conclusion

The EORTC QLQ-C30 has seen extensive research in cancer populations, only lacking information on *divergent validity*, and *criterion validity*. Oncokompas 2.0 uses the nausea and vomiting scale, as well as the financial impact item (in social life). Both constructs are suitable for use in Oncokompas 2.0. However, research in regard to cutoff scores is necessary.

Chapter 7

Results for Social Life

This chapter covers the results for the Quality of Life area of Social Life.

7.1 De Jong-Gierveld Loneliness Scale

7.1.1 Description

The De Jong-Gierveld Loneliness Scale is an 11-item measurement tool designed to measure loneliness and consists of two subscales: Emotional loneliness, and Social Loneliness (de Jong-Gierveld and Kamphuls, 1985). A 6-item version was developed in 2006 (Gierveld and Van Tilburg, 2006).

A total of 14 articles (Bielderman et al., 2013; Buz and Pérez-Arechaederra, 2014; Buz et al., 2014; Cramer and Barry, 1999; De Jong Gierveld and Van Tilburg, 2010; Gierveld and Van Tilburg, 2006; Grygiel et al., 2016, 2013; Iecovich, 2013; Leung et al., 2008; Penning et al., 2014; Spliethoff-Kamminga et al., 2003; Uysal-Bozkir et al., 2017; van Baarsen et al., 2001) concerning a non-cancer population were included for data extraction.

7.1.2 Structural validity

DJGLS-11 (9 studies): CFA (6 studies): 3 studies showed bad fit for unidimensional model. 4 studies showed acceptable fit for two-factor model, but 1 study showed poor fit for two-factor model. 2 studies showed best fit for bi-factor model (two-hierarchical).

Rasch (2 studies): 2 studies showed good fit of rasch model, as well that there are two distincts construct (emotional and social loneliness).

PCA (1 study): Three component solution (66.13% variance explained): one component for emotional loneliness (3 items); one for social loneliness (5 items); one with 3 items from emotional loneliness.

DJGLS-6 (1 study): CFA showed good model fit, indicating that the emotional and social subscales were two dimensions of the overarching loneliness concept.

CFA studies used maximum likelihood to deal with missing items. Two studies handled missing items through casewise deletion. Handling of missing items of remaining studies unknown.

7.1.3 Internal consistency

DJGLS-11: Cronbach's Alpha:

- Total (4 studies; 8 populations) .85 - .92
- Emotional (3 studies; 7 populations) .73 - .92
- Social (3 studies; 7 populations) .77 - .87

Omega Total scale (1 study; 2 measurements): .93 - .94.

DJGLS-6: Cronbach's Alpha:

- Total (2 studies) .70 - .76
- Emotional (1 study) .67 - .74
- Social (1 study) .69-.73

Four studies removed participants with missing values. Handling of remaining 3 studies is unknown.

7.1.4 Test-retest reliability

DJGLS-11: Test-retest correlation after 1 year of .65 (1 study). Person reliability (1 study; IRT) of .65. Item reliability (1 study; IRT) of .99. Both studies used casewise deletion for missing items.

DJGLS-6: Inter-rater (self-rated vs physician-assisted) ICC .98 – 1.00. Handling of missing items unknown.

7.1.5 Measurement Error

DJGLS-11: One study reported person standard error (IRT): .12 – 1.10. Missing items were handled through casewise deletion. None of the studies presenting test-retest reliability reported appropriate standard deviations, as such measurement error could not be calculated manually.

7.1.6 Known-groups comparison

DJGLS-11: One study found that older and younger adults in less than optimal health are significantly more at risk of emotional and of social loneliness. Differences were also found in households that had difficulty making ends meet vs those that didn't. Handling of missing items unknown.

One study found that participants who gave good ratings to their health situation or who did not have any limitations in daily functioning exhibited lower levels of loneliness than those who did. Participants with a partner scored lower than those without. Missing items were handled through casewise deletion.

7.1.7 Convergent validity

DJGLS-11 (9 studies): Single studies found significant correlations with the Groningen Frailty Indicator, self-rater loneliness, GDS-8, NA, SWLS, PA, satisfaction with life (2 studies), social self-efficacy, identification with class, UCLA loneliness scale, Lubben Social Network, Berliner Social Support subscale, Rosenberg Self-esteem Scale, BDI, Center for Epidemiologic Studies Depression Scale, BELA-A-k, quality of life, emotional state, physical health, marital status, social skills, self-esteem, need for affiliation, network size, and network support received.

One study had no missing values. Four studies handled missing items through casewise deletion. Handling of missing items for remaining four studies is unknown.

DJGLS-6 (2 studies): 1 study found high correlations between the 11-item and 6-item scales. One study found small ($r < .30$) correlations with the Cornell Scale for Depression in Dementia, current smoking status, category verbal fluency test, and visual forward span, but a high correlation with a direct question pertaining loneliness. Handling of missing items unknown.

7.1.8 Divergent validity

DJGLS-11: One study found no significant correlation with the need for help in taking medications, and need for help with finances. Handling of missing items is unknown.

7.1.9 Criterion validity

No data on criterion validity was reported.

7.1.10 Responsiveness

No data on responsiveness was reported.

7.1.11 Conclusion

The De Jong-Gierveld Loneliness Scale is well-established in non-cancer populations, with only information missing on *criterion validity*, and *responsiveness*. Establishing criterion validity for a questionnaire assessing loneliness is nigh imposibble, this does leave the issue of establishing cut-off points.

The De Jong-Gierveld Loneliness Scale is suitable for use in Oncokompas 2.0. Although no validation exists for cancer populations, given the construct, it is unlikely measurement properties will differ for cancer patients.

7.2 Dyadic Adjustment Scale

7.2.1 Description

The Dyadic Adjustment scale is a 32-item measurement tool to assess relationship adjustment of both partners in a relationship. It consists of four subscales: Dyadic concensus, dyadic cohesion, dyadic satisfaction, and affectional expression (Spanier, 1976). A number of shortened versions have been developed: DAS-14 (Busby et al., 1995), DAS-7 (Sharpley and Cross, 1982), and DAS-4 (Sabourin et al., 2005).

A total of 50 articles (Ahlborg et al., 2005; Anderson et al., 2014; Assari et al., 2009; Bailey et al., 2012; Baser et al., 2012; Busby et al., 1995; Busonera et al., 2014; Buyukbayraktar et al., 2015; Cano-Prous et al., 2014; Crane et al., 1991; Cuenca Montesino et al., 2013; Dargis et al., 2013; DeRogatis et al., 2010; Dinkel and Balck, 2005; Donarelli et al., 2016; Eddy et al., 1991; Fisiloglu and Demir, 2000; Funk and Rogge, 2007; Grice, 1997; Hjemboe and Butcher, 1991; Hollist et al., 2012; Hunsley et al., 2001, 1995; Isanezhad et al., 2012; Karakurt et al., 2009; Kazak et al., 1988; Lau et al., 2010; Lesch and Engelbrecht, 2008; Lim and Ivey, 2000; Manne and Schnoll, 2001; Pascoe and French, 1990; Pascoe et al., 1988; Rossier et al., 2006; Sabourin et al., 2005; Sanford, 1998; Sharpley and Cross, 1982; Shek, 1994, 1998; Shek and Cheung, 2008; South et al., 2009; Spanier, 1976; Stevenson-Hinde and Akister, 1995; Tantillo and Sanftner, 2010; Vaughn and Baier, 1999; Villeneuve et al., 2015; Wang et al., 2009; Wood et al., 1988; Woody and D'Souza, 1994; Youngblut et al., 2006) were included for data extraction. Two papers concerned cancer populations (Baser et al., 2012; Manne and Schnoll, 2001).

7.2.2 Structural validity

DAS-32: CFA (7 studies): Studies indicated that simple factor models (unifactorial or 4-factor models) did not fit adequately. Hierarchical models (second-order factor over 4 first-order factors) fit adequately in 5 studies).

EFA (2 studies): One study found that the Dyadic Satisfaction Subscale could not be supported by the data. One study found a 2-factor model.

PCA (2 studies): One study found a good component representation of the 4-factor model. One study found that for women the 4-factor model was supported, but not for men.

DAS-14: CFA (2 studies): Both studies found a good fit for a hierarchical model with 3 first-order factors and 1 second-order factor.

Three studies provided missing percentages. Handling for remaining studies is unknown.

7.2.3 Internal consistency

A total of 24 studies reported Cronbach's Alpha:

DAS-32 (16 studies; 18 samples):

- Total scale (15 studies; 16 samples) .70 - .95
- Concensus (8 studies) .78 - .90
- Cohesion (8 studies) .72 - .91
- Satisfaction (8 studies) .27 (outlier) - .94
- Affectional expression (7 studies) .13 (outlier) - .80

DAS-14 (2 studies): .822 - .86.

DAS -7 (4 studies): .76 - .84.

DAS-4 (2 studies): .84.

Seven studies reported missing item percentages.

7.2.4 Test-retest reliability

DAS-32: Test-retest correlation (2 studies):

- Consensus .82 - .85
- Cohesion .73 - .89
- Satisfaction .83 - .88
- Affectional expression .72 - .79
- Total (1 study) .70.

Intra-couple agreement (2 studies; lower limit is ICC, upper limit is correlation):

- Consensus .48 - .61
- Cohesion .53 - .68
- Satisfaction .49 - .68
- Affectional expression .29 - .66
- Total .52 - .72.

DAS-14: Test-retest correlation (2 studies; 5 samples): .71 - .98.

DAS-4: Test-retest correlation (1 study; 2 samples): .82 - .87.

Handling of missing items unknown for all studies.

7.2.5 Measurement Error

No data on measurement error was reported. None of the studies presenting test-retest reliability reported appropriate standard deviations, as such measurement error could not be calculated manually for these studies.

7.2.6 Known-groups comparison

DAS-32 (4 studies): One study showed significant differences between clinical and community samples. One study found significant gender differences. One study found no significant gender differences. One study found significant differences between maritally adapted and maladapted samples. One study found signicant differences between married and divorced samples.

DAS-7 (2 studies): One study showed significant differences between clinical and community samples, as well as between genders, and between maritally distressed and nondistressed samples. One study showed significant differences between married, divorced, separated, and co-habiting samples.

Missing item handling is unknown for all studies.

7.2.7 Convergent validity

DAS-32 (29 studies): Many significant associations found with related constructs. Most notably with other measures of Marital Adjustment and Marital Satisfaction.

DAS-14 (2 studies): One study found significant correlation (.52) with likert question on marital quality. One study found significant relation with Marital Happiness Scale and ENRICH.

DAS-7 & DAS-4 (3 studies): Significant correlations were found with many measurement instruments of Marital Adjustment and Marital Satisfaction.

seven studies provided missing percentages. Missing item handling of remaining studies is unknown.

7.2.8 Divergent validity

DAS-32 (5 studies): One study showed very low correlations with the Revised Conflict Tactics Scales. One study showed very low correlations with the IIEF-15. One study showed very low correlations with the FSFI. One study showed no significant correlation with 10 random unrelated questions. One study showed no significant correlation with years of cohabitation.

Two studies handled missing items through case-wise deletion. Handling of missing items of the remaining three studies is unknown.

7.2.9 Criterion validity

DAS-32 (3 studies): One study showed AUCs for Dyadic consensus .807 (.779-.836); dyadic satisfaction .857 (.833-.881); affectional expression .802 (.773-.831).

One study showed an area under the curve was 0.89 (95% CI=0.83–0.93). A cut-off of 8 for those satisfied and 7 for those who were unsatisfied was felt best able to maximize sensitivity and specificity. Patients who were lassified this way had a sensitivity and specificity of 86%. This study used casewise deletion for missing values.

One study tested multiple cut-off scores: Cut-off score 100: Sensitivity: .74 (men), .86 (women); specificity: .92 (men), .92 (women); PVP: .70 (men), .73 (women). Cut-off score 97: Sensitivity: .65 (men), .77 (women); specificity: .95 (men), .95 (women); PVP: .76 (men); .78 (women)

DAS-14 (2 studies): One study showed 86% accuracy for identifying nondistressed respondents, 74% accuracy for distressed respondents. 81% accuracy overall.

One study showed correct clasification of 94.2%.

DAS-7 (1 study): One study showed correct clasification of 90.7%.

DAS-10 (1 study): One study showed correct clasification of 91.7%.

DAS-4 (1 study): One study showed correct clasification of 90.2%.

Handling of missing items unknown.

7.2.10 Responsiveness

No data on responsiveness was reported.

7.2.11 Conclusion

The Dyadic Adjustment Scale is well-researched in the population of patients seeking marital therapy, but not in the cancer population. Given the strain cancer can put on relationships, it is questionable whether the validity and reliability is comparable in the cancer population.

Given the DAS is the most researched PROM for the measured construct, it is most likely suitable for use in Oncokompas 2.0, but research in the cancer population is recommended. The information on criterion validity can be used to inform our algorithms.

7.3 VGK Short-form

7.3.1 Description

The Vragenlijst Gezinskenmerken (Questionnaire Family Characteristics) is a Dutch-specific 23-item measurement tool consisting of two subscales: Waardering (appreciation) and samenwerking (cooperation). The VGK aims to screen for family dysfunction (Klijn, 2013).

A total of one thesis (Klijn, 2013) concerning noncancer populations was included for data extraction.

7.3.2 Structural validity

One study performed a CFA, which showed an acceptable fit for 2-factor model with a number of changes to which questions relate to which factors from the theoretical model:

- CGI .93
- TFI .93
- RMSEA .05
- SRMR .04.

Unknown how missing items were handled.

7.3.3 Internal consistency

One study presented Cronbach's Alphas:

- Waardering (appreciation) .88
- Samenwerking (cooperation) .92
- Total score .94

Unknown how missing items were handled.

7.3.4 Test-retest reliability

One study (N=60) presented a correlation, Green's Alpha, and Coefficient of Equivalence and Stability (CES) of two measurements 4 weeks apart:

- Waardering (appreciation):
  - r .77
  - alpha .74
  - CES .79
- Samenwerking (cooperation):
  - r .82
  - alpha .72
  - CES .81
- Total score:
  - r .85
  - alpha .81
  - CES .80

Unknown how missing items were handled.

7.3.5 Measurement Error

No data on measurement error was reported. One study presented test-retest reliability and standard deviations, as such the Standard Error of Measurement and Smallest Detectable Change could be calculated (based on CES):

- SEM:
  - Waardering (appreciation): 2.38
  - Samenwerking (cooperation): 3.14
  - Total score: 4.83
- SDC:
  - Waardering (appreciation): 6.61
  - Samenwerking (cooperation): 8.70
  - Total score: 13.39

Unknown how missing items were handled.

7.3.6 Known-groups comparison

All VGK scales significantly differentiate between psychiatric and non-psychiatric samples (psychiatric samples score lower). The total and waardering (appreciation) subscale, but not the samenwerking (cooperation) subscale, differentiate between samples with relationship problems and no relationship problems.

Unknown how missing items were handled.

7.3.7 Convergent validity

One study examined correlations with Family of Origin Scale (FOS); Interactionele Probleem Oplossings Vragenlijst (IPOV), and Level of Expressed Emotionscale (LEE):

- Waardering (appreciation):
  - FOS .82
  - IPOV .67 - .74
  - LEE -.52 - -.59
- Samenwerking (cooperation):
  - FOS .90
  - IPOV .63 - .77
  - LEE -.49 - .60
- Total score:
  - FOS .88
  - IPOV .68 - .76
  - LEE -.51 - -.61

Unknown how missing values were handled.

7.3.8 Divergent validity

One study found small correlation between -.20 - -.30 (ns) with unrelated constructs of Level of Expressed Emotion Scale. Unknown how missing values were handled.

7.3.9 Criterion validity

No data on criterion validity was reported.

7.3.10 Responsiveness

No data on responsiveness was reported.

7.3.11 Conclusion

The Vragenlijst Gezinskenmerken (Questionnaire Family Characteristics), has not been researched extensively. However, the thesis included reported on everything except for *criterion validity* and *responsiveness*. However, this validation was not performed in the cancer population. Given the strain cancer can put on the home situation, it is questionable whether the validity and reliability is comparable in the cancer population.

Given that the VGK is one of very few PROMs measuring child-parent relationships, it is most likely suitable for use in Oncokompas 2.0, but research in the cancer population is recommended.

7.4 EORTC IN-PATSAT32

7.4.1 Description

The European Organisation for Research and Treatment of Cancer (EORTC) IN-PATSAT32 is a 32-item tool to measure inpatient health care experience and appraisal (Brédart et al., 2005).

A total of nine articles (Arraras et al., 2010, 2009; Asadi-lari et al., 2015; Pishkuhi et al., 2014; Zhang et al., 2014, 2015; Aboshaiqah et al., 2016; Obtel et al., 2017) concerning a cancer population were included for data extraction.

7.4.2 Structural validity

Five studies performed multitrait item scaling and found that all item-scale correlations were >.40, and all item-other scale correlations were <.40. However, three studies had items (14% - 50%) that correlated higher with items in other scales than items in their own scales.

One Principal Component Analysis extracted five components with eigenvalue >1: Satisfaction of nurses, satisfaction of services and care, satisfaction of doctors, satisfaction of information provided by doctors, satisfaction of information provided by nurses.

Unknown how missing data was handled.

7.4.3 Internal consistency

Five studies reported Cronbach's Alphas for:

- Doctor technical skills .873 - .91
- Doctor interpersonal skills .91 - .96
- Doctor information provision .873 - .95
- Doctor availability .789 - .932
- Nurse technical skills .87 - .97
- Nurse interpersonal skills .835 - .93
- Nurse information provision .87 - .98
- Nurse availability .77 - .92
- Other staff skills .79 - .883
- Waiting times .668 - .87
- Hospital access .36 - .851

One study reported a Cronbach's Alpha for the Total score: .959.

One study reported missing data percentages. Unknown how missing items were handled.

7.4.4 Test-retest reliability

One study (N=70) tested test-retest after 2 weeks with correlations:

- Doctor technical skills .877
- Doctor interpersonal skills .87
- Doctor information provision .916
- Doctor availability .87
- Nurse technical skills .91
- Nurse interpersonal skills .936
- Nurse information provision .958
- Nurse availability .913
- Other staff skills .926
- Waiting times .903
- Hospital access .859

One study (N=133) tested test-retest reliability with ICCs after 30 minutes:

- Doctor technical skills .88

- Doctor interpersonal skills .91
- Doctor information provision .91
- Doctor availability .64
- Nurse technical skills .89
- Nurse interpersonal skills .73
- Nurse information provision .86
- Nurse availability .87
- Other staff skills .82
- Waiting times .70
- Hospital access .75
- Information exchange .84
- Hospital comfort .73
- Overall satisfaction .67

Unknown how missing items were handled.

7.4.5 Measurement Error

No data on measurement error was reported. Two studies presented test-retest reliability and standard deviations, as such the Standard Error of Measurement and Smallest Detectable Change could be calculated:

- SEM:
  - Doctor technical skills: 7.83 - 9.53
  - Doctor interpersonal skills: 8.09 - 10.82
  - Doctor information provision: 7.99 - 8.38
  - Doctor availability: 10.53 - 16.81
  - Nurse technical skills: 7.84 - 7.88
  - Nurse interpersonal skills: 6.47 - 12.08
  - Nurse information provision: 6.05 - 8.78
  - Nurse availability: 8.87 - 9.10
  - Other staff skills: 6.80 - 10.92
  - Waiting times: 9.03 - 16.13
  - Hospital access: 10.60 - 12.44
  - Information exchange (1 study): 10.70
  - Hospital comfort (1 study): 14.40
  - Overall satisfaction (1 study): 14.88
- SDC:
  - Doctor technical skills: 21.69 - 26.42
  - Doctor interpersonal skills: 22.41 - 29.98
  - Doctor information provision: 23.23 - 22.13
  - Doctor availability: 29.17 - 46.60
  - Nurse technical skills: 21.74 - 21.85
  - Nurse interpersonal skills: 17.93 - 33.49
  - Nurse information provision: 16.77 - 24.38
  - Nurse availability: 24.58 - 25.22
  - Other staff skills: 18.85 - 30.26
  - Waiting times: 25.04 - 44.70
  - Hospital access: 29.39 - 34.48
  - Information exchange (1 study): 29.66
  - Hospital comfort (1 study): 39.93
  - Overall satisfaction (1 study): 41.24

Unknown how missing items were handled.

7.4.6 Known-groups comparison

One study found patients younger than 59 to score significantly higher than those aged 59 or older. Another study did not find differences among age groups. This study did find that higher educated patients had significantly higher scores than lower educated patients. Longer diagnoses time was associated with higher scores in the nurse, other staff, and information provision scales. One study found that patients with higher scores on Oberst "patients' perception of care quality and satisfaction scale" had higher scores on the IN-PATSAT32.

One study reported missing item percentages. Handling of missing items for the remaining two studies is unknown.

7.4.7 Convergent validity

One study found correlations of .30 - .61 with the EORTC QLQ-INFO25 (N=509). One study found Rho correlations between .152 - .407 with EORTC QLQ-INFO25 (N=173). One study found correlation between relevant Oberst items between .60 - .70 (N=80). One study (N=130) found correlations with the QLQ-C15-PAL between .21 - .32.

One study (N=509) did not report handling of missing items. One study (N=80) reported missing item rates, but not how they were handled. One study (N=173) reported case-wise deletion.

7.4.8 Divergent validity

Four studies found either no significant, or correlations $< .40$ with the EORTC QLQ-C30.

One study did not report handling of missing items. Three studies mentioned missing item rates, but not how they were handled.

7.4.9 Criterion validity

No data on criterion validity was reported.

7.4.10 Responsiveness

No data on responsiveness was reported.

7.4.11 Conclusion

The EORTC IN-PATSAT32 has been researched in the cancer population, reporting on all measurement properties except for *criterion validity* and *responsiveness*. Doubt is shed on the *structural validity* of the EORTC IN-PATSAT32, which requires further research.

Taking into account that the EORTC IN-PATSAT32 is one of the more well-established PROMs measuring cancer patient satisfaction with care, it is likely suitable for Oncokompas 2.0, but investigation into structural validity is recommended.

7.5 Job Content Questionnaire

7.5.1 Description

The Job Content Questionnaire is a 49-item measurement tool to assess psychosocial job characteristics. It consists of eight scales: Skill discretion, decision authority, decision latitude, psychological demands, physical demands, job insecurity, supervisor support, and coworker support (Karasek et al., 1998).

A total of 44 articles were included for data extraction, however only 16 articles (Alexopoulos et al., 2015; Amin et al., 2015; Brisson et al., 1998; Chauvin et al., 2014; Cheng et al., 2003; Chien et al., 2011a,b; Choi et al., 2009, 2014; de Araújo and Karasek, 2008; Eum et al., 2007; Idrovo et al., 2012; Irniza et al., 2014; Karasek et al., 1998; Niedhammer, 2002; Pelfrene et al., 2001) which used the entire or subscales of the 49-item Job Content Questionnaire were used for this report. The remaining 28 articles used singularly adapted versions of the Job Content Questionnaire, of which the results were not generalisable.

7.5.2 Structural validity

Two studies performed a PCA and found a 4-factor structure consistent with the original model. Six studies performed an EFA, and found differing factor structures: 4-factor (2 studies), 5-factor (2 studies), 6-factor, and 8-factor structures. Four studies performed CFA of which only one found a good fit of the original model, with the others finding good fits after removing items or moving them to other subscales, or adding more factors.

Two studies performed a Rasch analysis, which fit well and discriminated well between individuals.

Two studies reported no missing. One study reported case-wise exclusion. Handling of missing items of other studies is unknown.

7.5.3 Internal consistency

Seven studies reported Cronbach's Alpha:

- Decision latitude (7 studies): .66 - .83
- Decision authority (6 studies) .45 - .72
- Psychological demand (11 studies) .51 - .84
- Social support (composite of supervisor and coworker support; 5 studies) .71 - .83
- Skill discretion (7 studies) .59 - .78
- Supervisor support (8 studies) .86 - .92
- Coworker support (8 studies) .69 - .86

Four studies used case-wise deletion for missing items. One study imputed 2.9%. Handling of missing items remaining studies unknown.

7.5.4 Test-retest reliability

Two studies reported test-retest correlations:

- Skill discretion .73 -.76
- Decision authority .59 - .64
- Psychological demands .62 -.73
- Coworker support .62 (1 study)
- Supervisor support .36 (1 study).

One study reported test-retest ICCs: .44 - .62.

Handling of missing items unknown

7.5.5 Measurement Error

One study reported Person standard error (1.99) and SEM (0.68). Handling of missing items unknown.

Measurement error could be calculated for one study reporting test-retest reliability:

Standerd Error of Measurement:

- Decision authority 0.52
- Psychological demand 0.72
- Coworker support 0.28
- Supervisor support 0.18

Smallest Detectable Change:

- Decision authority 1.43
- Psychological demand 1.99
- Coworker support 0.77
- Supervisor support 0.51

Two studies presenting test-retest reliability did not report appropriate standard deviations, as such measurement error could not be calculated manually for these studies.

7.5.6 Known-groups comparison

One study found multiple differences between formal and informal jobs. One study found that skill discretion and decision authority increased with employment grades, and low-skilled women had lower job control than low-skilled men. Handling of missing items is unknown.

7.5.7 Convergent validity

Two studies found significant correlations or regression coefficients with a job satisfaction question.Six individual studies found significant correlations with the PSS-14, self-reported general health score, education level, Social Capital at work Scale, Police Stress Questionnaire, and a job stress question. Two studies reported missing item percentages. Handling of missing items is unknown.

7.5.8 Divergent validity

No data on divergent validity was reported.

7.5.9 Criterion validity

One study found predictive value of the JCQ for psychological stress, of which the psychological demands subscale was the largest predictor. Organizational difficulties, and supervisor support added extra predictive value. Handlign of missing items is unknown.

7.5.10 Responsiveness

No data on responsiveness was reported.

7.5.11 Conclusion

The Job Content Questionnaire has seen a vast amount of research in noncancer populations, only missing information on *divergent validity*, *criterion validity*, and *responsiveness*.

Given that the JCQ is one of more well-established PROMs measuring supervisor and coworker support, it is likely suitable for use in Oncokompas 2.0. Investigations into suitable cut-off points is recommended.

7.6 VBBA

7.6.1 Description

The Vragenlijst Beleving en Beoordeling van de Arbeid (Questionnaire Experience and Rating of Labour) is a 243-item measurement tool of 27 subscales measuring different aspects of psychosocial workload in the workplace (Veldhoven et al., 2002).

A total of one manual (Veldhoven et al., 2002) concerning noncancer populations was included for data extraction.

7.6.2 Structural validity

One study performed a factor analysis which showed a six factor structure. These factors are assumed to be the second-level constructs of which subscales are a part of. Handling of missing items unknown.

7.6.3 Internal consistency

One study reported Cronbach's Alpha's:

- Werktempo (working pace) .89
- Emotionele belasting (emotional stress) .85
- Lichamelijke inspanning (physical effort) .90
- Geestelijke belasting (mental stress) .87
- Afwisseling (alternation) .82
- Leermogelijkheden (learning opportunities) .84
- Zelfstandigheid (independence) .90
- Inspraak (participation) .85
- Contactmogelijkheden (contact opportunities) .77
- Relatie collega's (relationship coworkers) .87
- Relatie leiding (relationship supervisors) .90
- Problemen met taak (difficulties with tasks) .80
- Onduidelijkheid taak (unclearness tasks) .81
- Verandering taak (change of tasks) .86
- Informatie (information) .83
- Communicatie (communication) .79
- Toekomstonzekerheid (future uncertainty) .95
- Beloning (rewards) .80
- Loopbaanmogelijkheden (career opportunities) .77
- Plezier (pleasure) .79
- Betrokkenheid (involvement) .79
- Verandering (change) .90
- Herstelbehoefte (recovery requirement) .87
- Piekeren (mulling) .80

- Emotionele reacties (emotional reactions) .89

Handling of missing items unknown.

7.6.4 Test-retest reliability

No data on test-retest reliability was reported.

7.6.5 Measurement Error

No data on measurement error was reported. As no test-retest reliability was reported, measurement could also not be calculated manually.

7.6.6 Known-groups comparison

No data on known-groups comparison was reported.

7.6.7 Convergent validity

No data on convergent validity was reported.

7.6.8 Divergent validity

No data on divergent validity was reported.

7.6.9 Criterion validity

Absence frequency could be predicted with 6% variance explained. Absence duration could be predicted with 8% variance explained. Handling of missing items unknown.

7.6.10 Responsiveness

No data on responsiveness was reported.

7.6.11 Conclusion

The Vragenlijst Beleving en Beoordeling van de Arbeid (Questionnaire Experience and Rating of Labour) lacks information on most measurement properties. No judgments regarding good or bad validity and reliability can be made, given the available information. Investigating alternative PROMs is recommended for use in Oncokompas 2.0. If no suitable alternatives can be identified, more research into the VBBA is necessary to justify validity, reliability, and cut-off points.

Chapter 8

Results for Lifestyle

This chapter covers the results for the Quality of Life area of Lifestyle.

8.1 Alcohol Five Shot

8.1.1 Description

The Alcohol Five Shot is 5-item measurement tool designed to screen for alcohol abuse and alcohol dependence (Seppä and Lepistö, 1998).

A total of two articles (Seppä and Lepistö, 1998; Aertgeerts et al., 2001) concerning noncancer populations was included for data extraction.

8.1.2 Structural validity

No data on structural validity was reported.

8.1.3 Internal consistency

No data on internal consistency was reported.

8.1.4 Test-retest reliability

No data on test-retest reliability was reported.

8.1.5 Measurement Error

No data on measurement error was reported. As no test-retest reliability was reported, measurement could also not be calculated manually.

8.1.6 Known-groups comparison

No data on known-groups comparison was reported.

8.1.7 Convergent validity

One study found a low correlation of .29 with alcohol consumption and a high correlation of .81 with the CAGE questionnaire. Unknown how missing values were handled.

8.1.8 Divergent validity

No data on divergent validity was reported.

8.1.9 Criterion validity

Two studies with three populations reported sensitivity and specificity for a cut-off point of 2.5 for screening alcohol abuse:

- Sensitivity .63 - .96
- Specificity .58 - .95

One study with two populations reported Positive Predictive Value and Negative Predictive Value for cut-off point of 2.5 for screening alcohol abuse:

- PPV .36 - .38
- NPV .92 - .95

One study with two populations reported AUC: .84 - .88.

One study reported sensitivity and specificity for cut-off point of 3 for screening alcohol abuse:

- Sensitivity .77
- Specificity .83

Handling of missing items is unknown.

8.1.10 Responsiveness

No data on responsiveness was reported.

8.1.11 Conclusion

Not much is known of the measurement properties of the Alcohol Five Shot, except for *criterion validity*. No judgments regarding good or bad validity and reliability can be made, given the available information. Investigating alternative PROMs is recommended for use in Oncokompas 2.0. If no suitable alternatives can be identified, more research into the A5S is necessary to justify validity, and reliability.

8.2 Perceived Stress Scale

8.2.1 Description

The Perceived Stress Scale is a screening tool for stress, by measuring the perception of stress in one's life. The Perceived Stress Scale comes in a 14-item (Cohen et al., 1983), 10-item (Taylor, 2015), and 4-item variant (Cohen, 1988).

A total of one review (based on 19 validation studies) (Lee, 2013) concerning noncancer populations was included for data extraction.

8.2.2 Structural validity

Exploratory factor analysis for the PSS-14 and PSS-10 indicated that a two-factor structure was more dominant than a one-factor structure. This was confirmed by the findings of confirmatory factor analysis. However, in many of the studies, the two-factor structure for the PSS-14 accounted for less than 50% of the total variance, which is the minimum percentage of cumulative variance extracted by successive factors.

8.2.3 Internal consistency

Eleven studies reported Cronbach's Alpha for the 14-item PSS: .75 - .89. Twelve studies reported Cronbach's Alpha for the 10-item PSS: .74 - .91. Six studies reported Cronbach's Alpha for the 4-item PSS: .60 - .82.

8.2.4 Test-retest reliability

Test-retest reliability was reported for the 14-item, and 10-item PSS:

- PSS-14:
  - ICC (2 weeks; 1 study) .90
  - Correlation (2 days – 6 weeks; 2 studies) .55 - .85.
- PSS-10:
  - ICC (1 – 4 weeks; 2 studies) .72-.88
  - Correlation (1 - 2 weeks; 2 studies) .74-.77.

8.2.5 Measurement Error

No data on measurement error was reported. As the article presenting test-retest reliability did not report standard deviations, measurement error could also not be calculated manually.

8.2.6 Known-groups comparison

No data on known-groups comparison was reported.

8.2.7 Convergent validity

The PSS was either moderately or strongly correlated with the hypothesized emotional variables, such as depression or anxiety, as measured using the Center for Epidemiologic Studies Depression Scale, Inventory to Diagnose Depression, Beck Depression Inventory, Hospital Anxiety and Depression Scale, State-Trait Anxiety Inventory, Escala de Cansancio Emo- cional (Scale of Emotional Exhaustion), General Health Questionnaire, Edinburgh Postnatal Depression Scale, Thai Depression Inventory, and Depression Anxiety Stress Scale-21.

8.2.8 Divergent validity

No data on divergent validity was reported.

8.2.9 Criterion validity

The criterion validity of PSS was evaluated in a few studies, of which the PSS was strongly correlated with only the mental component of health status as measured by the Medical Outcomes Study Short Form 36 (r = -.65 - -.70).

8.2.10 Responsiveness

No data on responsiveness was reported.

8.2.11 Conclusion

The Perceived Stress Scale has been well researched in noncancer populations, except for *measurement error*, *known-groups comparison*, *divergent validity*, *criterion validity*, and *responsiveness*. Particularly the lack of *criterion validity* research is an issue for establishing cut-off points. Furthermore, it could be argued that stress in the cancer population may be represented by other aspects than noncancer populations.

Given that the PSS is one of the most well-established PROMs measuring stress, it is likely suitable for use in Oncokompas 2.0. Research in the cancer population is recommended, with particular attention towards criterion validity.

Chapter 9

Results for Breast Cancer

This chapter covers the results for the Quality of Life areas particular for Breast Cancer patients.

9.1 Functional Assessment of Cancer Treatment - Endocrine Scale

9.1.1 Description

The Functional Assessment of Cancer Treatment Endocrine Scale is a 19-item tool designed to measure side effects of hormonal treatments given to breast cancer patients (Fallowfield et al., 1999).

A total of two articles (Fallowfield et al., 1999; Lester et al., 2012) concerning cancer populations were included for data extraction.

9.1.2 Structural validity

No data on structural validity was reported.

9.1.3 Internal consistency

One study reported a Cronbach's Alpha of .79 (N=306). Unknown how missing items were handled.

9.1.4 Test-retest reliability

One study reported test-retest reliability (N=56). Correlation between first measurement and second measurement (3-5 days later) was .86. Unknown how missing items were handled.

9.1.5 Measurement Error

No data on measurement error was reported. As the article presenting test-retest reliability did not report appropriate standard deviations, measurement error could also not be calculated manually.

9.1.6 Known-groups comparison

One study found no significant differences between treatment groups on the overall FACT-ES score. Significant differences were found between treatment groups on hot flushes and vaginal dryness items. Unknown how missing values were handled.

9.1.7 Convergent validity

One study found that the correlation between FACT-B and FACT-ES items measuring weight gain was .70. One study found that the correlation between FACT-B item measuring side effects and total FACT-ES was .41. One study found that the correlation between FACT-ES total and Urogenital Atrophy Questionnaire genital symptoms subscale was .549. Unknown how missing items were handled.

9.1.8 Divergent validity

No data on divergent validity was reported.

9.1.9 Criterion validity

No data on criterion validity was reported.

9.1.10 Responsiveness

One prospective treatment study on the effect of tamoxifen and anastrozole, found significant difference in mean FACT-ES score between baseline and 8 weeks, as well as between baseline and 12 weeks. Unknown how missing values were handled.

9.1.11 Conclusion

Relatively little is known of the measurement properties of the Functional Assessment of Cancer Treatment - Endocrine Scale. In particular, information is missing on *measurement error*, *structural validity*, *divergent validity*, and *criterion validity*.

Given that the FACT-ES is one of few PROMs measuring endocrine symptoms for breast cancer patients, it is likely suitable for use in Oncokompas 2.0. However, more research is recommended, particularly into *structural validity*.

9.2 Breast Impact of Treatment Scale

9.2.1 Description

The Breast Impact of Treatment Scale is a 13-item tool designed to measure body change stress of breast cancer patients as a result of treatment (Frierson et al., 2006).

A total of two articles (Frierson et al., 2006; Zainal et al., 2013) concerning cancer populations were included for data extraction.

9.2.2 Structural validity

Two studies performed exploratory factor analyses. One found one factor with 13 items (2 items deleted) with acceptable fit (RMSEA = .093). The other analysis extracted two factors with eigenvalue >1: Factor 1 was called Intrusion with 8 items and explaining 61% variance. Factor 2 was called Avoidance with 5 items and explaining 9.3% variance. The first study had no missing values. It is unknown how the second study handled missing values.

9.2.3 Internal consistency

Two studies reported a Cronbach's Alpha: .91 (N=194), and .945 (N=70). In the former, there were no missing values. In the latter it is unknown how missing values were handled.

9.2.4 Test-retest reliability

Two studies reported test-retest reliability. The ICC after 3 weeks was reported (N=70): .84 (.741 - .901). A correlation and mean change after 12 months was reported (N=62). The correlation was .70, and the within-person average change was -4.67 (p =.002). This mean change was small in relation to the SDs of the measures (SDs = 14.74 at initial and 14.55 at 12 months). Unknown how missing values were handled.

9.2.5 Measurement Error

No data on measurement error was reported. As the articles presenting test-retest reliability did not report standard deviations, measurement error could also not be calculated manually.

9.2.6 Known-groups comparison

One study found significant differences between treatment, with patients receiving Breast Conserving Therapy reporting lower levels of body change stress than patients receiving Modified Radical Mastectomy. Another study did not find such a difference. This latter study found scores to be significantly higher for patients with anxiety.

One study reported no missing values. Handling of missing items is unknown in the second study.

9.2.7 Convergent validity

One study found significant correlations with:

- Sexual responding .51
- Sexual unhappiness .51
- Sexual activity -.21
- Impact of Event Scales, Profile of Mood States, CES-D .43-.50

Another study found significant correlations with:

- HADS-A .584
- HADS-D .341
- HADS-Total .527
- No significant correlation with Rosenberg Self-Esteem Scale

Unknown how missing values were handled.

9.2.8 Divergent validity

No data on divergent validity was reported.

9.2.9 Criterion validity

One study performed a regression with the BITS predicting psychological and sexual outcomes (N=194). BITS explained 21-26% variance of psychological outcomes, and 18-44% variance of sexual outcomes. Unknown how missing values were handled.

9.2.10 Responsiveness

No data on responsiveness was reported.

9.2.11 Conclusion

The Breast Impact of Treatment Scale has seen little research, yet reporting information on most measurement properties except for *measurement error*, *divergent validity*, *criterion validity* and *responsiveness*. Notably, the two studies were contradicting on *structural validity*.

Given the lack of research on *criterion validity* and inconsistencies on *structural validity*, it is recommended to investigate alternative PROMs for use. If no suitable alternatives can be identified, more research into the BITS is necessary to justify validity, and cut-off points.

9.3 EORTC QLQ-BR23

9.3.1 Description

The European Organization for Research and Treatment of Cancer (EORTC) QLQ-BR23 is a 23-item tool designed to measure quality-of-life of breast cancer patients, and is designed to be complimentary to the EORTC QLQ-C30 (Sprangers et al., 1996).

A total of twenty articles (Alawadhi and Ohaeri, 2010; Alawadi and Ohaeri, 2009; Awad et al., 2008; Cerezo et al., 2012; Chie et al., 2003; Demirci et al., 2011; Den Oudsten et al., 2009; Escobar et al., 2015; Evangelista and Santos, 2012; Galalae et al., 2005; Glangkarn et al., 2011; Heil et al., 2010; Jayasekara et al., 2008; Kontodimopoulos et al., 2011; Michels et al., 2013; Montazeri et al., 2000; Parmar et al., 2005; Sprangers et al., 1996; Van Esch et al., 2011; Yun et al., 2004) concerning cancer populations were included for data extraction.

9.3.2 Structural validity

One study performed a Confirmatory Factor Analysis. While an acceptable fit was found for the theoretical factor structure, each subscale had a number of items that did not load sufficiently on their theorized factor. One study performed a Principal Component Analysis, and despite the fact that some items were not extracted exactly as in the original scales, they tended to be loaded into factors that consisted of similar items.

Six studies performed Multitrait Item Scaling. Almost all analyses found that body image, breast symptoms, and arm symptom items had $>.40$ correlation with other items in their scale, and lower correlations $<.40$ with items from other scales. One study found sexual functioning to have $<.40$ correlations with other items

in their scale. Two studies found systemic side effects to have <.40 correlations with other items in their scale.

Four studies (all Multitrait Item Scaling) reported their missing item percentages. The other four studies did not report how missing items were handled.

9.3.3 Internal consistency

A large number of studies reported Cronbach's Alphas:

- Total score .71 - .92 (4 studies)
- Breast symptoms .37 - .88 (11 studies with 15 populations)
- Arm symptoms .48 - .85 (11 studies with 15 populations)
- Body image .71 - .94 (10 studies with 14 populations)
- Sexual functioning .73 - .97 (10 studies with 14 populations)
- Systemic side effects .56 - .78 (10 studies with 14 populations)

7 out of 11 studies reported missing item percentages. The remaining 4 studies unknown how missing values were handled.

9.3.4 Test-retest reliability

Two studies reported ICCs. The first study reported ICCs of .70 for functional scales, and .77 for symptom scales after 14 days.

The second study reported ICCs after 1-2 weeks:

- Body image .93
- Sexual functioning .92
- Arm symptoms .92
- Breast symptoms .72
- Systemic side effects .92

The second study also reported Kappa agreement for the individual items: - Upset by hair loss .82 - Sexual enjoyment .75 - Future perspectives .45.

Unknown how missing values were handled.

9.3.5 Measurement Error

No data on measurement error was reported. One of the articles presenting test-retest reliability did not report appropriate standard deviations, as such measurement error could not be calculated manually for this study. One study reported standard deviations for functional and symptom scales for which Standard Error of Measurement and Smallest Detectable Change could be calculated:

- SEM:
  - Functional: 9.89
  - Symptom: 7.89
- SDC:
  - Functional: 26.58
  - Symptom: 21.21

Unknown how missing values were handled.

9.3.6 Known-groups comparison

Five studies reported significant differences between treatment groups, with mastectomy patients and patients undergoing radiotherapy combined with adjuvant chemotherapy having worse scores.

Five studies reported significant differences between different disease stages, with advanced stages reporting worse scores.

One study reports worse scores in patients with anxiety or depression. One study reports patients with lymphedema reporting worse scores than patients without lymphedema.

Six studies reported missing item percentages. Handling of missing items of the remaining four studies is unknown.

9.3.7 Convergent validity

Two studies found significant correlations between QLQ-BR23 functioning scales correlated with QLQ-C30 functioning scales and between most BR23 symptom scales correlated with QLQ-C30 symptom scales. Furthermore, three studies found significant correlations between conceptually related scales between the QLQ-C30 and QLQ-BR23.

Two studies found significant correlations with conceptually related scales between the WHOQOL-100 and QLQ-BR23. One study found significant correlations with conceptually related scales between the FACT-B and QLQ-BR23. One study found significant correlations with conceptually related scales between the Breast Cancer Treatment Outcome Scale and QLQ-BR23.

One study reported a correlation between the QLACS distress over recurrence scale and the QLQ-BR23 future perspective scale of .60.

One study found significant correlations between the POMS humor states and QLQ-BR23 functional scales.

One study found significant correlations between SF-36 physical scales and QLQ-BR23 functional and symptom scales.

Six out of eleven studies reported missing item percentages. Handling of missing items of the remaining four studies is unknown.

9.3.8 Divergent validity

One study found no significant correlations for unrelated subscales of the C30 and BR23. Unknown how missing values were handled.

9.3.9 Criterion validity

One study found that systemic side effects, future perspective, and upset by hair loss predicted Global Health Status (unknown how missing values were handled).

One study found that the BR23 distinguished between breast cancer patients and patients who subsequently were found to have benign lesions at follow-up after treatment (missing items handled with casewise deletion).

One study found that patients vs non-patients had significantly different scores on most scales (0% missing items).

9.3.10 Responsiveness

Five studies reported responsiveness not based on a RCT. Two studies reported before vs after treatment. Differences were found particularly in breast symptoms, upset by hair loss, and future perspectives.

One study found that the mean QLQ-BR23 scores indicated that those patients whose performance status had deteriorated over time reported decreased levels of functioning and increased levels of symptoms, and those patients whose performance status had improved over time reported increased levels of functioning and decreased levels of symptoms.

One study found differences on almost all scales for different treatment groups after 2-8 months, and after 6-18 months.

Three studies reported missing value percentages. For one study it is unknown how missing values were handled.

9.3.11 Conclusion

The EORTC QLQ-BR23 has been extensively researched, and is suitable for use in Oncokompas 2.0. Notably, in Oncokompas 2.0 the QLQ-BR23 is only used for measuring Lymphedema, of which no detailed information is given in the included studies. However, the symptom scales (taken together) were reportedly valid and reliable. The matter of cut-off points however, is not well-established and needs consideration.

9.4 Quick Disabilities of the Arm, Shoulder, and Hand Questionnaire

9.4.1 Description

The Quick Disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) is a shortened 11-item version of the DASH Outcome Measure, designed to measure physical function and symptoms in patients with any or multiple musculoskeletal disorders of the upper limb (Beaton et al., 2005; Gummesson et al., 2006).

A total of 13 validation studies (Angst et al., 2011; Mehta et al., 2010; Gabel et al., 2009, 2010; Polson et al., 2010; Roy et al., 2011; Stover et al., 2007; Mintken et al., 2009; Beaton et al., 2005; Fan et al., 2011; Quatman-Yates et al., 2013; Angst et al., 2009; Kennedy et al., 2013) were extracted from one review (Eden et al., 2014) concerning the QuickDASH in noncancer populations.

9.4.2 Structural validity

One study performed a Principal Component Analysis and found a bidimensional structure.

Handling if missing items is unknown.

9.4.3 Internal consistency

Seven studies reported the Cronbach's Alpha of the total QuickDASH score: .87 - .95.

One study reported the Cronbach's Alpha for the two seperate dimensions they found in a Principal Component Analysis: Symptoms: .75, functioning: .90"

9.4.4 Test-retest reliability

Five studies reported an ICC: .90 - .94. Only two of the studies were clear on the time between measurements (72 hours and 2-4 weeks respectively).

One study reported missing item percentages. Handling of missing items is unknown.

9.4.5 Measurement Error

Five studies reported Standard Error of Measurement: 0.06 – 6.37.

Three studies reported Minimal Detectable Change: 11.00 – 17.20.

One study reported Limits of Agreement: 2.77 (-7.23 - 12.77).

The Standard Error of Measurement (7.97) and Smallest Detectable Change (21.41) could be manually calculated for one study which reported test-retest reliability and standard deviations.

Handling of missing items is unknown.

9.4.6 Known-groups comparison

One study found higher QuickDASH scores for workers unable to do all works vs workers able to do all work. They also found higher scores for workers unable to work at all vs workers able to work.

9.4.7 Convergent validity

Three studies reported correlations with the SPADI (.69 - .84) and SF-36 (.68 - .73). Two studies reported correlations with the PREE (.69 - .81) and PRWE (.67 - .80). One study reported correlations with:

- DASH .97
- NDI .82
- CSOQ .57 - .68
- Symptom VAS .64

One study reported correlations with:

- Overall problem .70
- Pain .73
- Ability to function .80
- Ability to work .76

Handling of missing items is unknown.

9.4.8 Divergent validity

No data on divergent validity was reported.

9.4.9 Criterion validity

Four studies reported criterion validity for shoulder patients:

- AUC .59 - .82-
- Sensitivity .70 - .80
- Specificity .54 - .77

One study reported criterion validity for elbow patients:

- AUC .76
- Sensitivity .82
- Specificity .62

One study reported criterion validity for wrist patients:

- AUC .62
- Sensitivity .40
- Specificity .80

Handling of missing items is unknown.

9.4.10 Responsiveness

Two studies report significant changes after treatment (2 – 13 weeks), with a Standardized Response Mean of 1.1 – 1.25.

Two studies report Standardized Response Means for differing cases. Patients with observed change: 0.79, patients that rate themselves as improved 1.03, self-reported incident cases 0.6, confirmed incident cases 1.0, self-reported recovered cases -1.1, confirmed recovered cases -1.1.

Two studies report Minimally Clinically Important Differences of 4.5 – 8.

One study reports a Minimal Detectable Change of 13.3.

Handling of missing items is unknown.

9.4.11 Conclusion

The Quick Disabilities of the Arm, Shoulder, and Hand Questionnaire has been well researched, only lacking information on *divergent validity*. While not researched in cancer populations, the construct of shoulder function is physical and unlikely to be different for cancer populations.

The QuickDASH is suitable for use in Oncokompas 2.0. Information on criterion validity can be used to further inform algorithms.

9.5 BRECON-31

9.5.1 Description

The Breast Reconstruction Satisfaction Questionnaire (BRECON-31) is a 31-item tool to measure satisfaction with breast reconstruction for patients who have undergone breast reconstructive surgery (Temple-Oberle et al., 2013).

A total of two articles (Temple-Oberle et al., 2012, 2013) concerning cancer populations were included for data extraction.

9.5.2 Structural validity

Two studies report a Principal Component Analysis. Both studies found the exact same eight subscales. Missing item percentages reported by one study. Missing item handling of other study unknown.

9.5.3 Internal consistency

Two studies reported Cronbach's Alphas for:

- Self-image .86 - .91
- Arm concerns .91-.92
- Intimacy .83 - .85
- Satisfaction .90
- Recovery .76 - .84
- Self-consciousness .80 - .88
- Expectations .78 - .81
- Appearance .67
- Nipple .34 - .49

One study reported Cronbach's Alphas for:

- Abdomen .90
- Abdomen-strength .85
- Abdomen appearance .72

Missing item percentage reported by one study. Item missing handling of other study unknown.

9.5.4 Test-retest reliability

One study reported ICCs after 3 weeks (N=43):

- Self-image .66
- Arm concerns .73
- Intimacy .64
- Satisfaction .79
- Recovery .76
- Self-conscious .78
- Expectations .85
- Appearance .82
- Nipple .55
- Abdomen .83

Missing item handling unknown.

9.5.5 Measurement Error

No data on measurement error was reported. One study reported test-retest reliability and appropriate standard deviations. As such the Standard Error of Measurement and Smallest Detectable Change could be calculated manually:

- SEM:
  - Self-image: 8.05
  - Arm concerns: 8.31
  - Intimacy: 9.00
  - Satisfaction: 7.75
  - Recovery: 8.18
  - Self-conscious: 7.97
  - Expectations: 5.38
  - Appearance: 7.09
  - Nipple: 7.92
  - Abdomen: 5.40

- SDC:
  - Self-image: 21.62
  - Arm concerns: 22.34
  - Intimacy: 24.18
  - Satisfaction: 20.81
  - Recovery: 21.98
  - Self-conscious: 21.43
  - Expectations: 14.47
  - Appearance: 19.04
  - Nipple: 21.27
  - Abdomen: 14.51

Missing item handling unknown.

9.5.6 Known-groups comparison

No data on known-groups comparison was reported.

9.5.7 Convergent validity

One study found significant correlations of .34 - .76 between related constructs with the BREAST-Q. They also compared a summary score of BRECON-31 with the EQ-5D. The EQ-5D health thermometer correlated .50, and EQ-5D utility ratings correlated .42. Handling of missing items is unknown.

9.5.8 Divergent validity

No data on divergent validity was reported.

9.5.9 Criterion validity

No data on criterion validity was reported.

9.5.10 Responsiveness

No data on responsiveness was reported.

9.5.11 Conclusion

The Breast Reconstruction Satisfaction Questionnaire has not seen much research, but was only missing information on *known-groups comparison*, *divergent validity*, *criterion validity*, and *responsiveness*. The lack of information on *criterion validity* is an issue for establishing cut-off points. However, *criterion validity* is nigh impossible to establish for this construct.

Given that the BRECON-31 one of few PROMs measuring satisfaction with breast reconstruction and shows good evidence for most measurement properties (except for criterion validity which is nigh impossible to test), it is suitable for use in Oncokompas 2.0. However, without information on criterion validity, cut-off points are of consideration.

Chapter 10

Results for Intestinal Cancer

This chapter covers the results for the Quality of Life areas particular for Colorectal Cancer patients.

10.1 EORTC QLQ-CR29

10.1.1 Description

The European Organisation for Research and Treatment of Cancer (EORTC) QLQ-CR29 is a 29-item tool designed to measure quality-of-life of colorectal cancer patients, and is designed to be complimentary to the EORTC QLQ-C30 (Gujral et al., 2007).

A total of ten articles (Arraras et al., 2010; Costa et al., 2016; Hou et al., 2015; Ihn et al., 2015; Magaji et al., 2015; Nowak et al., 2011; Stiggelbout et al., 2016; Whistance et al., 2009; Lin et al., 2017; Sanna et al., 2017) concerning a cancer population were included for data extraction.

10.1.2 Structural validity

One study performed an Exploratory Factor Analysis and found 7 factors, of which 4 of the original scales. One combined factor for all bowel and stoma problems instead of seperate factors. The remaining factors were uninterpretable. Missing item percentages reported.

Seven studies performed multitrait item analysis. All analyses find most items correlating >.40 with items in their own scale, and correlating <.40 with items in their own scale. Exceptions are items in the blood and mucus scale (problems with convergent [2 studies] and discriminant [1 study]), urination frequency (problems with discriminant [1 study]), stool incontinence (problems with discriminant [1 study]), body image (problems with discriminant [1 study]).

Three of these studies reported missing item percentages. Missing value handling is unknown in the remaining two studies.

10.1.3 Internal consistency

Differing numbers of studies reported Cronbach's Alphas for certain subscales:

- Body image .55 - .87 (6 studies with 12 populations)
- Urinary frequency .363 - .80 (8 studies with 16 populations)
- Blood and mucus in stool .27 - .76 (7 studies, 12 populations)

- Stool frequency .46 - .87 (6 studies, 12 populations)
- Defaecation problems .84 - .85 (2 studies, 3 populations)
- Stoma problems .80 - .85 (2 studies, 2 populations)
- Abdominal pain .27 - .49 (1 study, 3 populations)
- Anxiety .00 - .68 (1 study, 3 populations).

Four studies reported missing value percentages. Three studies unknown how missing data was handled.

10.1.4 Test-retest reliability

Three studies reported ICCs for all scales and items after 7-14 (N=35) days, after a mean of 19 days (range 4-46; N=27), and after 2 weeks (N=40):

- Urinary frequency .33 -.64
- Stool frequency .27 - .89
- Body image .41 - .91
- Blood and mucus .88 (1 study)
- Urinary incontinence .20 - 1.00
- Dysuria .36 - .92
- Abdominal pain .79 - .85
- Buttock pain .74 - .94
- Bloated feeling .55 - .70
- Dry mouth .76 - .93
- Hair loss .74 - .95
- Trouble with taste .75 - .78
- Anxiety .54 - .83
- Weight .71 - .96
- Flatulence .64 - .82
- Fecal incontinence .75 - .80
- Sore skin .82 - .93
- Embarrassment .65 - .81
- Stoma care problems .86 (1 study)
- Defecation problems .89 (1 study)
- Impotence (men) .67 - .80
- Sexual function (men) .79 - .85
- Dyspareunia (women) .77 (1 study)
- Sexual function (women) .81 (1 study)

One study (N=70) reported that ICCs for all scales were $> .68$. One study (N=unclear) reported that test-retest correlation coefficients were moderate to high (range: r=0.51 to 1.00) for most of the scales except anxiety (r=0.30), urinary frequency (r=0.45), buttock pain (r=0.49), hair loss (r=0.14), stoma care related problems (r=0.41) and dyspareunia (r=0.33).

Missing item percentages known for the two former studies. Handling of missing items unknown.

10.1.5 Measurement Error

No data on measurement error was reported. Three studies reported test-retest reliability and standard deviations, due to which Standard Error of Measurement and Smallest Detectable Change could be calculated:

- SEM:
- Urinary frequency: 10.38 - 13.00 (2 studies)
- Stool frequency: 6.37 - 8.13 (2 studies)
- Body image: 5.12 - 7.44 (2 studies)
- Blood and mucus: 4.54 (1 study)

- Urinary incontinence: 0 - 16.36
- Dysuria: 3.39 - 10.96
- Abdominal pain: 6.58 - 10.17
- Buttock pain: 4.04 - 12.65
- Bloated feeling: 9.49 - 15.23
- Dry mouth: 6.80 - 9.98
- Hair loss: 3.82 - 8.87
- Trouble with taste: 8.35 - 12.45
- Anxiety: 10.60 - 17.91
- Weight: 3.82 - 14.33
- Flatulence: 9.16 - 16.44
- Fecal incontinence: 8.20 - 11.60
- Sore skin: 6.85 - 11.37
- Embarrassment: 10.16 - 17.75
- Stoma care problems: 7.41 (1 study)
- Impotence (men): 8.27 - 19.56
- Sexual function (men): 8.68 - 9.99
- Dyspareunia (women): 11.85 (1 study)
- Sexual function (women): 8.89 (1 study)
- SDC:
- Urinary frequency: 27.89 - 34.93 (2 studies)
- Stool frequency: 17.11 - 21.84 (2 studies)
- Body image: 13.76 - 19.99 (2 studies)
- Blood and mucus: 12.19 (1 study)
- Urinary incontinence: 0 - 43.96
- Dysuria: 9.12 - 29.45
- Abdominal pain: 17.69 - 27.34
- Buttock pain: 10.86 - 33.98
- Bloated feeling: 25.58 - 40.92
- Dry mouth: 18.27 - 26.82
- Hair loss: 10.27 - 23.83
- Trouble with taste: 22.43 - 33.45
- Anxiety: 28.47 - 48.11
- Weight: 10.26 - 38.49
- Flatulence: 24.62 - 44.17
- Fecal incontinence: 22.04 - 31.17
- Sore skin: 18.41 - 30.55
- Embarrassment: 27.29 - 47.69
- Stoma care problems: 19.91 (1 study)
- Impotence (men): 22.23 - 52.56
- Sexual function (men): 23.31 - 26.84
- Dyspareunia (women): 31.83 (1 study)
- Sexual function (women): 23.89 (1 study)

Missing item percentages known for two studies. Handling of missing items unknown.

10.1.6 Known-groups comparison

Six studies report significant worse scores for stoma patients. One study reported no significant different scores for stoma patients. Four studies report worse scores for patients with worse Karnofsky performance scores. Three studies report conflicting differences between curative and palliative patients, with one study reporting more blood/mucus and defaecation problems for the palliative group, and two other studies vice versa. Three studies report conflicting differences between older and younger patients, with one study

reporting lower sexual functioning for older patients, and two others study vice versa. Numerous small effects are noted between different cancer groups across two studies.

Four studies report missing item percentages. Missign item handling is unknown for the remaining two studies.

10.1.7 Convergent validity

Five studies report correlations with QLQ-C30. Noted are that the convergent correlations are relatively small, but do follow the expected pattern where the functional scales correlate with eachother, and symptom scales correlate with eachother.

One study notes some correlations between related scales and items of the QLQ-C30:

- C30 social & CR29 body image: .57
- C30 emotional & CR29 body image: .51
- C30 pain & CR29 abdominal pain: .51
- C30 constipation & CR29 buttock pain: .52
- C30 diarrhea & CR29 faecal incontinence: .51

Handling of missing values is unknown for one study. The remaining three studies report missing item percentage.

One study reports positive correlations of all scales with the Low Anterior Resection Syndrome score, although a number of them are $< .30$ (anxiety, weight, urinary incontinence, dysuria, abdominal pain, bloated feeling, blood and mucus, dry mouth). Handling of missing values is unknown.

10.1.8 Divergent validity

Two studies report correlations between irrelevant constructs of the QLQ-C30 being between .02 - .40. Missing item percentages are reported.

10.1.9 Criterion validity

No data on citerion validity was reported.

10.1.10 Responsiveness

One study (not RCT) reports significant mean reduction after palliative chemotherapy in anxiety about weight loss, physical function, and pain. The same study (not RCT) reports significant improval of social function after stoma closure. Handing of missing values is unknown.

10.1.11 Conclusion

The EORTC QLQ-CR29 is well studied, with only information missing on *criterion validity*. However, *structural validity* is questionable.

Oncokompas 2.0 uses the bladder issues, abdomnal pain, buttocks pain, anal pain, bloated feeling, and blood and mucus in defecation subscales / items. Particularly the abdominal pain subscale shows bad internal consistency, which raises doubts into its' suitability for use in Oncokompas 2.0.

The EORTC QLQ-CR29 (except for abdominal pain) is likely suitable for use in Oncokompas 2.0, but more research into structural validity is recommended. Furthermore, the lack of research on criterion validity raises issues with defining cut-off points.

10.2 Stoma Quality of Life Questionnaire

10.2.1 Description

The Stoma Quality of Life Questionnaire is a 20-item tool designed to measure the quality of life of patients with a stoma (Prieto et al., 2005).

A total of one article (Prieto et al., 2005) concerning a noncancer population was included for data extraction.

10.2.2 Structural validity

One study performed a Rasch analysis and Principal Component Analysis. Rasch: The 20 items fit to define a unidimensional variable according to initial Rasch specifications (Infit MNSQ < 1.3). Item parameters by country also fitted to the Rasch model (Infit MNSQ$<$1.3) and had very similar item calibrations: ICC of the item calibrations by country was 0.81 (0.67–0.91 95% CI). PCA: Single component accounting for 38% variance.

Missing items handled by casewise deletion.

10.2.3 Internal consistency

One study reported a Index of Person Seperation of 2.92, and a Cronbach's Alpha of .92. Missing items handled by casewise deletion.

10.2.4 Test-retest reliability

One study reported correlations between measurements:

- 1st and 2nd measurements: .913
- 1st and 3rd measurements: .881
- 2nd and 3rd measurements: .946

Time between measurements unknown. Missing items handled by casewise deletion.

10.2.5 Measurement Error

No data on measurement error was reported. As the article presenting test-retest reliability did not report standard deviations, measurement error could not be calculated manually.

10.2.6 Known-groups comparison

No data on known-groups comparison was reported.

10.2.7 Convergent validity

No data on convergent validity was reported.

10.2.8 Divergent validity

No data on divergent validity was reported.

10.2.9 Criterion validity

No data on citerion validity was reported.

10.2.10 Responsiveness

No data on responsiveness was reported.

10.2.11 Conclusion

The Stoma Quality of Life Questionnaire has not seen much research, notably lacking information on *measurement error*, *construct validity* (*known-groups comparison*, *convergent validity*, and *divergent validity*), *criterion validity*, and *responsiveness*.

The S-QoL does show good evidence of the measurement properties that were investigated. As such, it is likely suitable for Oncokompas 2.0, but more research into construct validity is recommended. The lack of research on criterion validity raises issues with defining cut-off points.

Chapter 11

Results for Head and Neck Cancer

This chapter covers the results for the Quality of Life areas particular for Head and Neck Cancer patients.

11.1 EORTC QLQ-H&N35

11.1.1 Description

The European Organization for Research and Treatment of Cancer (EORTC) QLQ-H&N35 is a 35-item tool designed to measure quality-of-life of head and neck cancer patients, and is designed to be complimentary to the EORTC QLQ-C30 (Singer et al., 2015).

A total of two articles (Bjordal et al., 2000, 1999) concerning cancer populations were included for data extraction.

11.1.2 Structural validity

Two studies report multitrait item scaling. Both studies found correlations >.40 with items in their own scale, except for item 4, 8, 18, 19, 20. Missing item percentages were reported.

11.1.3 Internal consistency

Two studies reported Cronbach's Alphas for:

- Pain .78 - .81
- Swallowing .78 - .82
- Social contact .82 - .83
- Senses .10 - .72 (5 populations)
- Speech .10 - .87 (3 populations)
- Social eating .79 - .87

One study reported the Cronbach's Alpha for sexuality: .95.

Missing item percentages were reported.

11.1.4 Test-retest reliability

No data on test-retest reliability was reported.

11.1.5 Measurement Error

No data on measurement error was reported. As no test-retest reliability was reported, measurement could also not be calculated manually.

11.1.6 Known-groups comparison

Two studies report worse scores for recurrent patients. One study reports numerous differences between Oral Cavity, Pharynx, and Larynx patients. The same study reports worse scores for patients with worse Karnofsky performance scores. Both studies report missing item percentages.

11.1.7 Convergent validity

One study reports correlations with the QLQ-C30:

- Pain scales of both instruments .63
- Social functioning scale of QLQ-C30 and social contact scale of QLQ-H&N35 -.58

Missing item percentages were reported.

11.1.8 Divergent validity

Two studies report correlations with irrelevant constructs of the QLQ-C30. One study reportes that most scales had correlations <.70, while the other reports correlations <.55. Missing item percentages are reported.

11.1.9 Criterion validity

No data on criterion validity was reported.

11.1.10 Responsiveness

One study reports significant deterioration of almost all functional and symptom scales and items after treatment. One study reports a deterioration of more than 10 units in mean scores from before treatment to 2 months after the start of treatment, except for the social contact scale, and single items of “problems opening mouth”, “dry mouth”, “sticky saliva”, and “felt ill”. Missing item percentages were reported.

11.1.11 Conclusion

The EORTC QLQ-H&N35 has not seen much research, lacking information on *test-retest reliability*, *measurement error*, and *criterion validity*. Furthermore, the investigations into *structural validity* were of poor quality.

Given that the EORTC QLQ-H&N35 is one of few PROMs measuring quality of life for head and neck cancer patients, it is likely the most suitable PROM for use in Oncokompas 2.0 for the constructs it measures. However, more research is highly recommended. The lack of research into criterion validity raises issues with defining cut-off points.

11.2 Shoulder Disability Questionnaire

11.2.1 Description

The Shoulder Disability Questionnaire is a 16-item tool designed to screen for Shoulder Disability symptoms (van der Heijden et al., 2000).

A total of one article (Stuiver et al., 2016) concerning a cancer population was included for data extraction.

11.2.2 Structural validity

No data on structural validity was reported.

11.2.3 Internal consistency

One study reported a Cronbach's Alpha of .91. Missing items were handled through case-wise deletion.

11.2.4 Test-retest reliability

One study reported an ICC (N=58) after 7 days of .84 (.74 - .90). Missing items were handled through case-wise deletion.

11.2.5 Measurement Error

No data on measurement error was reported. One study reported test-retest reliability and standard deviations, due to which Standard Error of Measurement (11.36) and Smallest Detectable Change (30.52) could be calculated. Missing items were handled through case-wise deletion.

11.2.6 Known-groups comparison

One study reports significant differences between patients with <90 degree vs >90 degree AROM abduction, with <90 degree reporting worse scores. The same study reports significant differences between patients who report pain vs no pain at passive external rotation of the shoulder, with the pain reporters having worse scores. The study reported missing score percentages.

11.2.7 Convergent validity

One study reports correlations with:

- Shoulder Pain and Disability Index .78
- NDII -.76
- Clinician-rated abduction -.56
- SF-36 physical -.45
- SF-36 social -.41
- SF-36 physical role -.49
- SF-36 pain -.59

Handling of missing items unknown.

11.2.8 Divergent validity

One study reported low correlations with theoretically unrelated constructs:

- SF-36 emotional role -.28
- SF-36 mental health -.28
- SF-36 vitality -.41
- SF-36 health perception -.19
- SF-36 health change -.33

Handling of missing items unknown.

11.2.9 Criterion validity

One study reports the ability to disciminate between patients reporting need for treatment vs patients reporting no need for treatment, after surgery. AUC 3 months after surgerys was .85 (.78 - .94). AUC 6 months after surgerys was .77 (.63 – 91). Handling of missing items unknown.

11.2.10 Responsiveness

One study reports a visual comparison of mean differences after surgery showed a decrease in shoulder disability from baseline to 3 months and 6 months after surgery. Handling of missing values is unknown.

11.2.11 Conclusion

The Shoulder Disability Questionnaire has not seen much investigation in the cancer population. Information was lacking on *structural validity*. The other measurement properties showed good evidence.

The Shoulder Disability Questionnaire is likely suitable for use in Oncokompas 2.0. Research into structural validity is recommended.

Chapter 12

Recommendations

Table 12.1 contains a summary of recommendations of the PROMs used in Oncokompas 2.0. These recommendations are based on the conclusions previously discussed. All conclusions are collated again in chapter 13 for an easy overview.

Five PROMs were indicated as high priority (Patiënt Specifieke Klachten, CARON, VBBA, Alcohol Five Shot, Breast Cancer Patients' Needs Questionnaire), with the main recommendation of exploring alternative PROMs. If no alternative PROMs are available, the recommendation is a validation study.

Six PROMs were indicated as medium priority (Dyadic Adjustment Scale, VGK Short-form, Functional Assessment of Cancer Therapy - Endocrine Scale, Breast Impact of Treatment Scale, EORTC QLQ-CR29, EORTC QLQ-H&N35), with the main recommendation of additional research into either missing information of important measurement properties, or into replication of the validity in a cancer population.

The remaining PROMs were either classified as no priority (7 PROMS: definitely suitable for use) or low priority (11 PROMS: likely suitable, but with some missing information). The low priority PROMs could use extra validation studies for certain measurement properties (mostly structural or criterion validity), or validition studies for a cancer population.

Table 12.1: Recommendations for PROMs in Oncokompas 2.0

Measurement Construct	**Questionnaire**	**Validated in cancer population?**	**Recommendation for Oncokompas 2.0**	**Priority**
Psychological wellbeing				
Depression & Anxiety	Hospital Anxiety and Depression Scale	Yes	Suitable for use	None
Fear of recurrence	Cancer Worry Scale	Yes	Suitable for use	None
Subjective cognitive functioning	SF-36: Cognitive function scale	Yes	Explore alternatives	Low
Physical generic wellbeing				
General everyday life	Patiënt Specifieke Klachten	No	Explore alternatives	High
Sleep	Insomnia Severity Index	Yes	Suitable for use	None

Measurement Construct	Questionnaire	Validated in cancer population?	Recommendation for Oncokompas 2.0	Priority
Sexuality (women)	6-item Female Sexual Function Index	Yes (FSFI-19: 2 studies: internal consistency, test-retest, structural validity, convergent validity, divergent validity)	Likely suitable for use / Research recommended for structural validity	Low
Sexuality (men)	5-item International Index of Erectile Function	Yes (IIEF-15: 2 studies: structural validity [Rasch analysis], known-groups comparison, convergent validity, responsiveness).	Likely suitable for use / Research recommended for structural validity	Low
Body image	Body Image Scale	Yes	Likely suitable for use / Research recommended for criterion validity	Low
Nausea and vomiting	EORTC QLQ-C30: Nausea and vomiting scale	Yes	Suitable for use / Research recommended for criterion validity	Low
Hearing Impairment	CARON	No	Explore alternatives	High
Social life				
Social life	De Jong-Gierveld Loneliness Scale	No	Suitable for use	None
Relationship	Dyadic Adjustment Scale	Yes (DAS-14 [1 study]: internal consistency; structural validity; known-groups comparison; convergent validity)	Suitable for use / Research recommended for cancer population	Medium
Relationship with doctor	EORTC IN-PATSAT32	Yes	Likely suitable for use / Research recommended for structural validity	Low
Relationship with children	VGK Short-form	No	Likely suitable for use / Research recommended for cancer population	Medium
Financial circumstances	EORTC QLQ-C30: Financial impact item	Yes	Suitable for use / Research recommended for criterion validity	Low
Relationship with boss & coworkers	Job Content Questionnaire: Supervisor & coworkers scales	No	Likely suitable for use / Research recommended for criterion validity	Low

Measurement Construct	Questionnaire	Validated in cancer population?	Recommendation for Oncokompas 2.0	Priority
Job retention and resumption	VBBA	No	Explore alternatives	High
Lifestyle				
Alcohol Use	Alcohol Five Shot	No	Explore alternatives	High
Relaxation	Perceived Stress Scale	No	Likely suitable for use / Research recommended for criterion validity	Low
Breast cancer				
Menopausal symptoms	Functional Assessment of Cancer Therapy - Endocrine Scale	Yes	Resarch for structural and criterion validity	Medium
Body image	Breast Impact of Treatment Scale	Yes	Explore alternatives	Medium
Lymphedema	EORTC QLQ BR23	Yes	Suitable for use	None
Shoulder function	Quick Disabilities of the Arm, Shoulder, and Hand questionnaire	No	Suitable for use	None
Breast reconstruction	BRECON-31	Yes	Likely suitable for use	None
Breast Prosthesis	Breast Cancer Patients' Needs Questionnaire: Prosthesis items	No	Explore alternatives	High
Intestinal Cancer				
Bladder issues, Pain (abdomen, buttocks, anus), Bloated, Defecation (blood, mucus)	EORTC QLQ CR29	Yes	Likely suitable for use / Research recommended for structural and criterion validity	Medium
Stoma leakage, gas, being away from home, clothing, sexuality, appearance, social life	Stoma Quality of Life Questionnaire	No	Likely suitable for use / Research recommended for construct and criterion validity	Low
Head and neck cancer				
Speech & swallowing, & mouth function	EORTC QLQ H&N35	Yes	Likely suitable for use / Research recommended for structural and criterion validity	Medium
Shoulder function	Shoulder Disability Questionnaire	Yes	Likely suitable for use / Research recommended for structural validity	Low

Chapter 13

All conclusions

In this chapter, all conclusions have been put together for a quick and easy overview of the results.

13.1 Psychological

13.1.1 Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression scale is a well established PROM for use in the cancer population, with evidence for measurement properties except for *measurement error*, and *divergent validity*. This PROM is more than suitable for use in Oncokompas 2.0. The information on criterion validity can be used to inform our algorithms.

13.1.2 Cancer Worry Scale

The Cancer Worry Scale has not seen much research, but the research that has been published show indications for good measurement properties, except for *test-retest reliability*, *measurement error*, and *responsiveness*, which were not studied. More research is also needed for structural validity. The CWS is suitable for use in Oncokompas 2.0, although more research would be preferrable.

Furthermore, two different cut-off scores were investigated. This information can be used to optimize the algorithms in Oncokompas 2.0.

13.1.3 SF-36

While many studies investigated the SF-36, providing evidence for many measurement properties, a couple of important measurement properties were not investigated or not investigated thoroughly enough: *test-retest reliability*, *measurement error*, *structural validity*, *divergent validity*, and *criterion validity*.

Only the cognitive subscale is used in the Oncokompas 2.0. This subscale is most likely suitable for use, but alternatives with better defined cutoff scores should be explored.

13.2 Physical

13.2.1 Patiënt Specifieke Klachten

Very little is known about the Patiënt Specifieke Klachten. No judgments regarding good or bad validity and reliability can be made, given the available information. Investigating alternative PROMs is recommended for use in Oncokompas 2.0. If no suitable alternatives can be identified, more research into the PSK is necessary to justify validity, reliability, and cut-off points.

13.2.2 Insomnia Severity Index

The Insomnia Severity Index is studied by a handful of studies in the cancer population, but shows evidence of good measurement properties, except for *measurement error*, and *known-groups comparison*, which were not investigated. This PROM is suitable for use in Oncokompas 2.0, and the studied cutoff points can be used to inform algorithms.

13.2.3 6-item Female Sexual Function Index

The Female Sexual Function Index 19-item has been studied in a cancer population, which showed similar results as for other populations. Given this, we can assume that the shortened 6-item version will also correspond in the cancer population. The FSFI-6 has seen some research, but information was lacking for *structural validity*, *measurement error*, *known-groups comparison*, *divergent validity*, and *responsiveness*. However, *criterion validity* was investigated, giving information related to a cutoff points for clinical diagnosis.

The FSFI-6 is most likely suitable for use in Oncokompas 2.0, but more research (particularly into *structural validity*) in the cancer population is recommended. The studied cutoff points can be used to inform algorithms.

13.2.4 5-item International Index of Erectile Function

The International Index of Erectile Function has been studied in a cancer population, which showed similar results as for other populations. Given this, we can assumed that the shortened 5-item version will also correspond in the cancer population. The IIEF-5 has seen some research, but information was lacking for *structural validity*.

The IIEF-5 is most likely suitable for use in Oncokompas 2.0, but more research (particularly into *structural validity*) in the cancer population is recommended. The studied cutoff points can be used to inform algorithms.

13.2.5 Body Image Scale

The Body Image Scale has been well researched in cancer populations, only missing information on *criterion validity*. This BIS is suitable for use in Oncokompas, but research is needed to establish cutoff points.

13.2.6 EORTC QLQ-C30

The EORTC QLQ-C30 has seen extensive research in cancer populations, only lacking information on *divergent validity*, and *criterion validity*. Oncokompas 2.0 uses the nausea and vomiting scale, as well as the financial impact item (in social life). Both constructs are suitable for use in Oncokompas 2.0. However, research in regard to cutoff scores is necessary.

13.3 Social

13.3.1 De Jong-Gierveld Loneliness Scale

The De Jong-Gierveld Loneliness Scale is well-established in non-cancer populations, with only information missing on *criterion validity*, and *responsiveness*. Establishing criterion validity for a questionnaire assessing loneliness is nigh imposibble, this does leave the issue of establishing cut-off points.

The De Jong-Gierveld Loneliness Scale is suitable for use in Oncokompas 2.0. Although no validation exists for cancer populations, given the construct, it is unlikely measurement properties will differ for cancer patients.

13.3.2 Dyadic Adjustment Scale

The Dyadic Adjustment Scale is well-researched in the population of patients seeking marital therapy, but not in the cancer population. Given the strain cancer can put on relationships, it is questionable whether the validity and reliability is comparable in the cancer population.

Given the DAS is the most researched PROM for the measured construct, it is most likely suitable for use in Oncokompas 2.0, but research in the cancer population is recommended. The information on criterion validity can be used to inform our algorithms.

13.3.3 VGK Short-form

The Vragenlijst Gezinskenmerken (Questionnaire Family Characteristics), has not been researched extensively. However, the thesis included reported on everything except for *criterion validity* and *responsiveness*. However, this validation was not performed in the cancer population. Given the strain cancer can put on the home situation, it is questionable whether the validity and reliability is comparable in the cancer population.

Given that the VGK is one of very few PROMs measuring child-parent relationships, it is most likely suitable for use in Oncokompas 2.0, but research in the cancer population is recommended.

13.3.4 EORTC IN-PATSAT32

The EORTC IN-PATSAT32 has been researched in the cancer population, reporting on all measurement properties except for *criterion validity* and *responsiveness*. Doubt is shed on the *structural validity* of the EORTC IN-PATSAT32, which requires further research.

Taking into account that the EORTC IN-PATSAT32 is one of the more well-established PROMs measuring cancer patient satisfaction with care, it is likely suitable for Oncokompas 2.0, but investigation into structural validity is recommended.

13.3.5 Job Content Questionnaire

The Job Content Questionnaire has seen a vast amount of research in noncancer populations, only missing information on *divergent validity*, *criterion validity*, and *responsiveness*.

Given that the JCQ is one of more well-established PROMs measuring supervisor and coworker support, it is likely suitable for use in Oncokompas 2.0. Investigations into suitable cut-off points is recommended.

13.3.6 VBBA

The Vragenlijst Beleving en Beoordeling van de Arbeid (Questionnaire Experience and Rating of Labour) lacks information on most measurement properties. No judgments regarding good or bad validity and reliability can be made, given the available information. Investigating alternative PROMs is recommended for use in Oncokompas 2.0. If no suitable alternatives can be identified, more research into the VBBA is necessary to justify validity, reliability, and cut-off points.

13.4 Lifestyle

13.4.1 Alcohol Five Shot

Not much is known of the measurement properties of the Alcohol Five Shot, except for *criterion validity*. No judgments regarding good or bad validity and reliability can be made, given the available information. Investigating alternative PROMs is recommended for use in Oncokompas 2.0. If no suitable alternatives can be identified, more research into the A5S is necessary to justify validity, and reliability.

13.4.2 Perceived Stress Scale

The Perceived Stress Scale has been well researched in noncancer populations, except for *measurement error*, *known-groups comparison*, *divergent validity*, *criterion validity*, and *responsiveness*. Particularly the lack of *criterion validity* research is an issue for establishing cut-off points. Furthermore, it could be argued that stress in the cancer population may be represented by other aspects than noncancer populations.

Given that the PSS is one of the most well-established PROMs measuring stress, it is likely suitable for use in Oncokompas 2.0. Research in the cancer population is recommended, with particular attention towards criterion validity.

13.5 Breast cancer

13.5.1 Functional Assessment of Cancer Treatment - Endocrine Scale

Relatively little is known of the measurement properties of the Functional Assessment of Cancer Treatment - Endocrine Scale. In particular, information is missing on *measurement error*, *structural validity*, *divergent validity*, and *criterion validity*.

Given that the FACT-ES is one of few PROMs measuring endocrine symptoms for breast cancer patients, it is likely suitable for use in Oncokompas 2.0. However, more research is recommended, particularly into *structural validity*.

13.5.2 Breast Impact of Treatment Scale

The Breast Impact of Treatment Scale has seen little research, yet reporting information on most measurement properties except for *measurement error*, *divergent validity*, *criterion validity* and *responsiveness*. Notably, the two studies were contradicting on *structural validity*.

Given the lack of research on *criterion validity* and inconsistencies on *structural validity*, it is recommended to investigate alternative PROMs for use. If no suitable alternatives can be identified, more research into the BITS is necessary to justify validity, and cut-off points.

13.5.3 EORTC QLQ-BR23

The EORTC QLQ-BR23 has been extensively researched, and is suitable for use in Oncokompas 2.0. Notably, in Oncokompas 2.0 the QLQ-BR23 is only used for measuring Lymphedema, of which no detailed information is given in the included studies. However, the symptom scales (taken together) were reportedly valid and reliable. The matter of cut-off points however, is not well-established and needs consideration.

13.5.4 Quick Disabilities of the Arm, Shoulder, and Hand Questionnaire

The Quick Disabilities of the Arm, Shoulder, and Hand Questionnaire has been well researched, only lacking information on *divergent validity*. While not researched in cancer populations, the construct of shoulder function is physical and unlikely to be different for cancer populations.

The QuickDASH is suitable for use in Oncokompas 2.0. Information on criterion validity can be used to further inform algorithms.

13.5.5 BRECON-31

The Breast Reconstruction Satisfaction Questionnaire has not seen much research, but was only missing information on *known-groups comparison*, *divergent validity*, *criterion validity*, and *responsiveness*. The lack of information on *criterion validity* is an issue for establishing cut-off points. However, *criterion validity* is nigh impossible to establish for this construct.

Given that the BRECON-31 one of few PROMs measuring satisfaction with breast reconstruction and shows good evidence for most measurement properties (except for criterion validity which is nigh impossible to test), it is suitable for use in Oncokompas 2.0. However, without information on criterion validity, cut-off points are of consideration.

13.6 Intestinal cancer

13.6.1 EORTC QLQ-CR29

The EORTC QLQ-CR29 is well studied, with only information missing on *criterion validity*. However, *structural validity* is questionable.

Oncokompas 2.0 uses the bladder issues, abdomnal pain, buttocks pain, anal pain, bloated feeling, and blood and mucus in defecation subscales / items. Particularly the abdominal pain subscale shows bad internal consistency, which raises doubts into its' suitability for use in Oncokompas 2.0.

The EORTC QLQ-CR29 (except for abdominal pain) is likely suitable for use in Oncokompas 2.0, but more research into structural validity is recommended. Furthermore, the lack of research on criterion validity raises issues with defining cut-off points.

13.6.2 Stoma Quality of Life Questionnaire

The Stoma Quality of Life Questionnaire has not seen much research, notably lacking information on *measurement error*, *construct validity* (*known-groups comparison*, *convergent validity*, and *divergent validity*), *criterion validity*, and *responsiveness*.

The S-QoL does show good evidence of the measurement properties that were investigated. As such, it is likely suitable for Oncokompas 2.0, but more research into construct validity is recommended. The lack of research on criterion validity raises issues with defining cut-off points.

13.7 Head and neck cancer

13.7.1 EORTC QLQ-H&N35

The EORTC QLQ-H&N35 has not seen much research, lacking information on *test-retest reliability*, *measurement error*, and *criterion validity*. Furthermore, the investigations into *structural validity* were of poor quality.

Given that the EORTC QLQ-H&N35 is one of few PROMs measuring quality of life for head and neck cancer patients, it is likely the most suitable PROM for use in Oncokompas 2.0 for the constructs it measures. However, more research is highly recommended. The lack of research into criterion validity raises issues with defining cut-off points.

13.7.2 Shoulder Disability Questionnaire

The Shoulder Disability Questionnaire has not seen much investigation in the cancer population. Information was lacking on *structural validity*. The other measurement properties showed good evidence.

The Shoulder Disability Questionnaire is likely suitable for use in Oncokompas 2.0. Research into structural validity is recommended.

Bibliography

Aboshaiqah, A., Al-Saedi, T. S. B., Abu-Al-Ruyhaylah, M. M. M., Aloufi, A. A., Alharbi, M. O., Alharbi, S. S. R., Al-Saedi, A. S., and Al-Erwi, A. F. (2016). Quality of life and satisfaction with care among palliative cancer patients in Saudi Arabia. *Palliative and Supportive Care*, 14(06):621–627.

Aertgeerts, B., Buntinx, F., Ansoms, S., and Fevery, J. (2001). Screening properties of questionnaires and laboratory tests for the detection of alcohol abuse or dependence in a general practice population. *Br J Gen Pract*.

Ahlborg, T., Persson, L. O., and Hallberg, L. R. (2005). Assessing the quality of the dyadic relationship in first-time parents: development of a new instrument. *J Fam Nurs*, 11:19–37.

Aktas, A., Walsh, D., and Kirkova, J. (2015). The psychometric properties of cancer multisymptom assessment instruments: a clinical review. *Supportive Care Cancer*, 23:2189–2202.

Alawadhi, S. A. and Ohaeri, J. U. (2010). Validity and reliability of the European Organization for Research and Treatment in Cancer Quality of Life Questionnaire (EORTC QLQ): experience from Kuwait using a sample of women with breast cancer. *Annals of Saudi Medicine*, 30:390–396.

Alawadi, S. A. and Ohaeri, J. U. (2009). Health - related quality of life of Kuwaiti women with breast cancer: a comparative study using the EORTC Quality of Life Questionnaire. *Bmc Cancer*, 9.

Alexopoulos, E. C., Argyriou, E., Bourna, V., and Bakoyannis, G. (2015). Reliability and validity of the Greek version of the job content questionnaire in Greek health care workers. *Saf Health Work*, 6:233–239.

Amin, N. A., Quek, K. F., Oxley, J. A., Noah, R. M., and Nordin, R. (2015). Validity and reliability of Malay version of the job content questionnaire among public hospital female nurses in Malaysia. *Int J Occup Environ Med*, 6:232–242.

Anagnostopoulos, F. and Myrgianni, S. (2009). Body image of Greek breast cancer patients treated with mastectomy or breast conserving surgery. *J Clin Psychol Med Settings*, 16:311–321.

Anderson, S. R., Tambling, R. B., Huff, S. C., Heafner, J., Johnson, L. N., and Ketring, S. A. (2014). The development of a reliable change index and cutoff for the Revised Dyadic Adjustment Scale. *J Marital Fam Ther*, 40:525–534.

Angst, F., Goldhahn, J., Drerup, S., Flury, M., Schwyzer, H.-K., and Simmen, B. R. (2009). How sharp is the short QuickDASH? A refined content and validity analysis of the short form of the disabilities of the shoulder, arm and hand questionnaire in the strata of symptoms and function and specific joint conditions. *Quality of Life Research*, 18(8):1043–1051.

Angst, F., Schwyzer, H., and Aeschlimann, A. (2011). Measures of adult shoulder function: disabilities of the arm, shoulder, and hand questionnaire (DASH) and its short version (QuickDASH), shoulder pain and. *Arthritis care &*.

Arraras, J. I., Greimel, E., Sezer, O., Chie, W. C., Bergenmar, M., Costantini, A., Young, T., Vlasic, K. K., and Velikova, G. (2010). An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. *European Journal of Cancer*, 46(15):2726–2738.

Arraras, J. I., Vera, R., Martínez, M., Hernández, B., Laínez, N., Rico, M., Vila, M., Chicata, V., and Asín, G. (2009). The EORTC cancer in-patient satisfaction with care questionnaire: EORTC IN-PATSAT32. *Clinical and Translational Oncology*, 11(4):237–242.

Asadi-lari, M., Ahmadi Pishkuhi, M., Almasi-Hashiani, A., Safiri, S., and Sepidarkish, M. (2015). Validation study of the EORTC information questionnaire (EORTC QLQ-INFO25) in Iranian cancer patients. *Supportive Care Cancer*, 23(7):1875–1882.

Aslan, Y., Tuncel, A., Aydin, O., Balci, M., Karabulut, E., and Atan, A. (2011). The association between erection hardness grading scale and international index of erectile function in men with erectile dysfunction treated with sildenafil citrate. *Urol Int*, 86:434–438.

Assari, S., Lankarani, M. M., and Tavallaii, S. A. (2009). Revised dyadic adjustment scale as a reliable tool for assessment of quality of marital relationship in patients on long-term hemodialysis. *Iran J Kidney Dis*, 3:242–245.

Atisha, D., Alderman, A. K., Lowery, J. C., Kuhn, L. E., Davis, J., and Wilkins, E. G. (2008). Prospective Analysis of Long-term Psychosocial Outcomes in Breast Reconstruction. *Annals of Surgery*, 247(6):1019–1028.

Awad, M. A., Denic, S., and El Taji, H. (2008). Validation of the European Organization for Research and Treatment of Cancer quality of life questionnaires for Arabic-speaking populations.

Bailey, J., Kerley, S., and Kibelstis, T. (2012). A brief marital satisfaction screening tool for use in primary care medicine. *Fam Med*, 44:105–109.

Bartula, I. and Sherman, K. A. (2015). The Female Sexual Functioning Index (FSFI): evaluation of acceptability, reliability, and validity in women with breast cancer. *Supportive Care in Cancer*, 23:2633–2641.

Baser, R. E., Li, Y. L., and Carter, J. (2012). Psychometric validation of the female sexual function index (FSFI) in cancer survivors. *Cancer*, 118:4606–4618.

Bastien, C. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, 2(4):297–307.

Bayraktar, Z. and Atun, A. I. (2012). Despite some comprehension problems the international index of erectile function is a reliable questionnaire in erectile dysfunction. *Urol Int*, 88:170–176.

Bayraktar, Z. and Atun, I. (2013). Impact of physician assistance on the reliability of the International Index of Erectile Function. *Andrologia*, 45:73–77.

Beaton, D., Wright, J., and Katz, J. (2005). Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.*

Bedard, G., Zeng, L., Lam, H., Cella, D., Zhang, L. Y., Lauzon, N., and Chow, E. (2012). Meaningful change in oncology quality-of-life instruments: a systematic literature review. *Expert Review of Pharmacoeconomics & Outcomes Research*, 12:475–483.

Beurskens, A. J., de Vet, H. C., Kökeb, A. J., Lindeman, E., van der Heijden, G. J., Regtope, W., and Knipschild, P. G. (1999). A patient-specific approach for measuring functional status in low back pain. *Journal of Manipulative and Physiological Therapeutics*, 22(3):144–148.

Bielderman, A., van der Schans, C. P., van Lieshout, M. R. J., de Greef, M. H., Boersma, F., Krijnen, W. P., and Steverink, N. (2013). Multidimensional structure of the Groningen Frailty Indicator in community-dwelling older people. *BMC geriatr.*, 13:86.

Bird, L., Arthur, A., Niblock, T., Stone, R., Watson, L., and Cox, K. (2010). Rehabilitation programme after stem cell transplantation: randomized controlled trial. *Journal of Advanced Nursing*, 66(3):607–615.

Bjordal, K., De Graeff, A., Fayers, P. M., Hammerlid, E., Van Pottelsberghe, C., Curran, D., Ahlner-Elmqvist, M., Maher, E. J., Meyza, J. W., Br??dart, A., S??derholm, A. L., Arraras, J. J., Feine, J. S., Abendstein, H., Morton, R. P., Pignon, T., Huguenin, P., Bottomly, A., and Kaasa, S. (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 and N35) in head and neck patients. *European Journal of Cancer*, 36(14):1796–1807.

Bjordal, K., Hammerlid, E., Ahlner-Elmqvist, M., de Graeff, A., Boysen, M., Evensen, J. F., Biorklund, A., de Leeuw, J. R., Fayers, P. M., Jannert, M., Westin, T., and Kaasa, S. (1999). Quality of life in head and neck cancer patients: validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-H&N35. *J Clin Oncol*, 17(3):1008–1019.

Bousquet, J., Knani, J., Dhivert, H., Richard, A., Chicoye, A., Ware, J. E., and Michel, F. B. (1994). Quality of life in asthma. I. Internal consistency and validity of the SF-36 questionnaire. *American journal of respiratory and critical care medicine*, 149(2 Pt 1):371–5.

Brandberg, Y., Malm, M., and Blomqvist, L. (2000). A prospective and randomized study, "SVEA," comparing effects of three methods for delayed breast reconstruction on quality of life, patient-defined problem areas of life, and cosmetic result. *Plastic and reconstructive surgery*, 105(1):66–74; discussion 75–6.

Brédart, A., Bottomley, A., Blazeby, J. M., Conroy, T., Coens, C., D'Haese, S., Chie, W. C., Hammerlid, E., Arraras, J. I., Efficace, F., Rodary, C., Schraub, S., Costantini, M., Costantini, A., Joly, F., Sezer, O., Razavi, D., Mehlitz, M., Bielska-Lasota, M., and Aaronson, N. K. (2005). An international prospective study of the EORTC cancer in-patient satisfaction with care measure (EORTC IN-PATSAT32). *European Journal of Cancer*, 41(14):2120–2131.

Brisson, C., Blanchette, C., Guimont, C., Dion, G., Moisan, J., Vézina, M., Dagenais, G. R., and Masse, L. (1998). Reliability and validity of the French version of the 18-item Karasek Job Content Questionnaire. *Work Stress*, 12:322–336.

Busby, D., Christensen, C., and Crane, D. (1995). A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: Construct hierarchy and multidimensional scales. *Journal of Marital and*.

Bushmakin, A. G., Cappelleri, J. C., Symonds, T., and Stecher, V. J. (2014). Further Understanding of the International Index of Erectile Function at 15+ Years: Confirmatory Factor Analysis and Multidimensional Scaling. *Ther Innov Regul Sci*, 48:246–254.

Busonera, A., Martini, P. S., Zavattini, G. C., and Santona, A. (2014). Psychometric properties of an italian version of the experiences in close relationships-revised (ECR-R) scale. *Psychol Rep*, 114:785–801.

Buyukbayraktar, C. G., Ozteke, H. I., and Kesici, S. (2015). The Validity and Reliability Study for the Turkish Version of the Commitment Scale. *Anthropologist*, 20:523–534.

Buz, J. and Pérez-Arechaederra, D. (2014). Psychometric properties and measurement invariance of the Spanish version of the 11-item De Jong Gierveld loneliness scale. *Int Psychogeriatr*, 26:1553–1564.

Buz, J., Urchaga, D., and Polo, M. E. (2014). Factor structure of the de Jong Gierveld loneliness scale in Spanish elderly adults. *Anales De Psicologia*, 30:588–596.

Cano-Prous, A., Martin-Lanas, R., Moya-Querejeta, J., Beunza-Nuin, M. I., Lahortiga-Ramos, F., and Garcia-Granero, M. (2014). Psychometric properties of a Spanish version of the Dyadic Adjustment Scale. *International Journal of Clinical and Health Psychology*, 14:137–144.

Cappelleri, J., Rosen, R., Smith, M., and Quirk, F. (1999). Some developments on the international index of erectile function (IIEF). *Drug information*.

Cerezo, O., Oñate-Ocaña, L. F., Arrieta-Joffe, P., González-Lara, F., García-Pasquel, M. J., Bargalló-Rocha, E., and Vilar-Compte, D. (2012). Validation of the Mexican-Spanish version of the EORTC QLQ-C30 and BR23 questionnaires to assess health-related quality of life in Mexican women with breast cancer. *Eur J Cancer Care*, 21:684–691.

Chauvin, B., Rohmer, O., Spitzenstetter, F., Raffin, D., Schimchowitsch, S., and Louvet, E. (2014). Assessment of job stress factors in a context of organizational change. *European Review of Applied Psychology-Revue Europeenne De Psychologie Appliquee*, 64:299–306.

Chedraui, P., Pérez-López, F. R., Sánchez, H., Aguirre, W., Martínez, N., Miranda, O., Plaza, M. S., Schwager, G., Narváez, J., Quintero, J. C., and Zambrano, B. (2012). Assessment of sexual function of mid-aged Ecuadorian women with the 6-item Female Sexual Function Index. *Maturitas*, 71:407–412.

Cheng, Y., Luh, W. M., and Guo, Y. L. (2003). Reliability and validity of the Chinese Version of the Job Content Questionnaire in Taiwanese workers. *Int J Behav Med*, 10:15–30.

Chie, W. C., Chang, K. J., Huang, C. S., and Kuo, W. H. (2003). Quality of life of breast cancer patients in Taiwan: Validation of the Taiwan Chinese version of the EORTC QLQ-C30 and EORTC QLQ-BR23. *Psycho-Oncology*, 12:729–735.

Chien, T. W., Lai, W. P., Lu, C. W., Wang, W. C., Chen, S. C., Wang, H. Y., and Su, S. B. (2011a). Web-based computer adaptive assessment of individual perceptions of job satisfaction for hospital workplace employees. *BMC Med Res Methodol*, 11:47.

Chien, T. W., Lai, W. P., Wang, H. Y., Hsu, S. Y., Castillo, R. V., Guo, H. R., Chen, S. C., and Su, S. B. (2011b). Applying the revised Chinese Job Content Questionnaire to assess psychosocial work conditions among Taiwan's hospital workers. *BMC Public Health*, 11:478.

Choi, B., Bjorner, J. B., Ostergren, P. O., Clays, E., Houtman, I., Punnett, L., Rosengren, A., De Bacquer, D., Ferrario, M., Bilau, M., and Karasek, R. (2009). Cross-language differential item functioning of the job content questionnaire among European Countries: The JACE study. *Int J Behav Med*, 16:136–147.

Choi, B., Ko, S., Dobson, M., Schnall, P. L., Garcia-Rivas, J., Israel, L., and Baker, D. (2014). Short-term test-retest reliability of the Job Content Questionnaire and Effort-Reward Imbalance Questionnaire items and scales among professional firefighters. *Ergonomics*, 57:897–911.

Clark, L., Holcombe, C., Hill, J., Krespi-Boothby, M. R., Fisher, J., Seward, J., and Salmon, P. (2011). Sexual abuse in childhood and postoperative depression in women with breast cancer who opt for immediate reconstruction after mastectomy. *The Annals of The Royal College of Surgeons of England*, 93(2):106–110.

Cohen, S. (1988). Perceived stress in a probability sample of the United States. In *S. Spacapan, S. Oskamp (Eds.), The social psychology of health: Claremont symposium on applied social psychology, Sage, Newbury Park, CA*.

Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, 24(4):385.

Conroy, T., Mercier, M., Bonneterre, J., Luporsi, E., Lefebvre, J., Lapeyre, M., Puyraveau, M., and Schraub, S. (2004). French version of FACT-G: validation and comparison with other cancer-specific instruments. *European Journal of Cancer*, 40(15):2243–2252.

Costa, A. L. S., Heitkemper, M. M., Alencar, G. P., Damiani, L. P., Silva, R. M. D., and Jarrett, M. E. (2016). Social Support Is a Predictor of Lower Stress and Higher Quality of Life and Resilience in Brazilian Patients With Colorectal Cancer. *Cancer Nurs*.

Coyne, K., Mandalia, S., McCullough, S., Catalan, J., Noestlinger, C., Colebunders, R., and Asboe, D. (2010). The international index of erectile function: Development of an adapted tool for use in HIV-positive men who have sex with men. *J Sex Med*, 7:769–774.

Cramer, K. M. and Barry, J. E. (1999). Conceptualizations and measures of loneliness: a comparison of subscales. *Personality and Individual Differences*, 27:491–502.

Crane, D. R., Busby, D. M., and Larson, J. H. (1991). A Factor-Analysis of the Dyadic Adjustment Scale with Distressed and Nondistressed Couples. *American Journal of Family Therapy*, 19:60–66.

Cuenca Montesino, M. L., Luis Gómez, J., Peña Fernández, M. E., and Andreu Rodríguez, J. M. (2013). Psychometric properties of the Dyadic Adjustment Scale (DAS) in a community sample of couples. *Psicothema*, 25:536–541.

Custers, J. (2014). The cancer worry scale: detecting fear of recurrence in breast cancer survivors. *Cancer …*.

Custers, J. A. E., Gielissen, M. F. M., Janssen, S. H. V., de Wilt, J. H. W., and Prins, J. B. (2016). Fear of cancer recurrence in colorectal cancer survivors. *Supportive Care in Cancer*, 24:555–562.

Dargis, L., Trudel, G., Cadieux, J., Villeneuve, L., Préville, M., and Boyer, R. (2013). Validation of the International Index of Erectile Function (IIEF) and presentation of norms in older men. *Sexologies*, 22:e20–e26.

de Araújo, T. M. and Karasek, R. (2008). Validity and reliability of the job content questionnaire in formal and informal jobs in Brazil. *Scand J Work Environ Health Suppl*, pages 52–59.

de Jong-Gierveld, J. and Kamphuls, F. (1985). The Development of a Rasch-Type Loneliness Scale. *Applied Psychological Measurement*, 9(3):289–299.

De Jong Gierveld, J. and Van Tilburg, T. (2010). The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. *Eur. j. ageing*, 7:121–130.

Demirci, S., Eser, E., Ozsaran, Z., Tankisi, D., Aras, A. B., Ozaydemir, G., and Anacak, Y. (2011). Validation of the Turkish versions of EORTC QLQ-C30 and BR23 modules in breast cancer patients. *Asian Pac J Cancer Preven*, 12:1283–1287.

Den Oudsten, B. L., Van Heck, G. L., Van der Steeg, A. F. W., Roukema, J. A., and De Vries, J. (2009). The WHOQOL-100 has good psychometric properties in breast cancer patients. *J Clin Epidemiol*, 62:195–205.

DeRogatis, L. R., Allgood, A., Auerbach, P., Eubank, D., Greist, J., Bharmal, M., Zipfel, L., and Guo, C. Y. (2010). Validation of a Women's Sexual Interest Diagnostic Interview - Short Form (WSID-SF) and a Daily Log of Sexual Activities (DLSA) in postmenopausal women with hypoactive sexual desire disorder. *J Sex Med*, 7:917–927.

Dinkel, A. and Balck, F. (2005). An evaluation of the German Relationship Assessment Scale. *Swiss Journal of Psychology*, 64:259–263.

Donarelli, Z., Lo Coco, G., Gullo, S., Salerno, L., Marino, A., Sammartano, F., and Allegra, A. (2016). The fertility quality of life questionnaire (FertiQoL) relational subscale: Psychometric properties and discriminant validity across gender. *Hum Reprod*, 31:2061–2071.

Duman-Lubberding, S., van Uden-Kraan, C. F., Jansen, F., Witte, B. I., van der Velden, L. A., Lacko, M., Cuijpers, P., Leemans, C. R., and Verdonck-de Leeuw, I. M. (2016). Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. *Supportive Care in Cancer*, 24(5):2163–2171.

Eddy, J. M., Heyman, R. E., and Weiss, R. L. (1991). An Empirical-Evaluation of the Dyadic Adjustment Scale - Exploring the Differences between Marital Satisfaction and Adjustment. *Behavioral Assessment*, 13:199–220.

Eden, M. M., Flores, A. M., Galantino, M. L., and Spinelli, B. A. (2014). Recommendations for patient-reported outcome measures for head and neck cancer-related shoulder dysfunction: A systematic review. *Rehabil Oncol*, 32:6–19.

Edsander-Nord A, E.-N., Brandberg, Y., and Wickman, M. (2001). Quality of life, patients' satisfaction, and aesthetic outcome after pedicled or free TRAM flap breast surgery. *Plastic and reconstructive surgery*, 107(5):1142–53; discussion 1154–5.

Elder, E. E., Brandberg, Y., Björklund, T., Rylander, R., Lagergren, J., Jurell, G., Wickman, M., and Sandelin, K. (2005). Quality of life and patient satisfaction in breast cancer patients after immediate breast reconstruction: a prospective study. *The Breast*, 14(3):201–208.

EORTC Quality of Life Department. Modules that have completed phase III.

EORTC Quality of Life Department. Validated modules specific diseases.

Escobar, A., Trujillo-Martín, M. M., Rueda, A., Pérez-Ruiz, E., Avis, N. E., and Bilbao, A. (2015). Cross-cultural adaptation, reliability and validity of the Spanish version of the Quality of Life in Adult Cancer Survivors (QLACS) questionnaire: Application in a sample of short-term survivors. *Health Qual Life Outcomes*, 13.

Eum, K. D., Li, J., Jhun, H. J., Park, J. T., Tak, S. W., Karasek, R., and Cho, S. I. (2007). Psychometric properties of the Korean version of the job content questionnaire: Data from health care workers. *Int Arch Occup Environ Health*, 80:497–504.

Evangelista, A. L. and Santos, E. M. M. (2012). Cluster of symptoms in women with breast cancer treated with curative intent. *Supportive Care Cancer*, 20:1499–1506.

Fallowfield, L. J., Leaity, S. K., Howell, A., Benson, S., and Cella, D. (1999). Assessment of quality of life in women undergoing hormonal therapy for breast cancer: validation of an endocrine symptom subscale for the FACT-B. *Breast Cancer Research and Treatment*, 55(2):189–199.

Fan, Z. J., Smith, C. K., and Silverstein, B. A. (2011). Responsiveness of the QuickDASH and SF-12 in Workers with Neck or Upper Extremity Musculoskeletal Disorders: One-Year Follow-Up. *Journal of Occupational Rehabilitation*, 21(2):234–243.

Fillion, L., Gélinas, C., Simard, S., Savard, J., and Gagnon, P. (2003). Validation evidence for the French Canadian adaptation of the Multidimensional Fatigue Inventory as a measure of cancer-related fatigue. *Cancer Nurs*, 26:143–154.

Fisiloglu, H. and Demir, A. (2000). Applicability of the dyadic adjustment scale for measurement of marital quality with Turkish couples. *European Journal of Psychological Assessment*, 16:214–218.

Flynn, K. E., Reeve, B. B., Lin, L., Cyranowski, J. M., Bruner, D. W., and Weinfurt, K. P. (2013). Construct validity of the PROMIS® sexual function and satisfaction measures in patients with cancer. *Health Qual Life Outcomes*, 11.

Frierson, G. M., Thiel, D. L., and Andersen, B. L. (2006). Body change stress for women with breast cancer: The breast-impact of treatment scale. *Ann Behav Med*, 32:77–81.

Funk, J. L. and Rogge, R. D. (2007). Testing the Ruler With Item Response Theory: Increasing Precision of Measurement for Relationship Satisfaction With the Couples Satisfaction Index. *J Fam Psychol*, 21:572–583.

Gabel, C. P., Michener, L. A., Melloh, M., and Burkett, B. (2010). Modification of the Upper Limb Functional Index to a Three-point Response Improves Clinimetric Properties. *Journal of Hand Therapy*, 23(1):41–52.

Gabel, C. P., Yelland, M., Melloh, M., Burkett, B., Angst, F., Goldhahn, J., Drerup, S., Aeschlimann, A., Schwyzer, H., R, S., Doward, L., McKenna, S., Cleland, J., Childs, J., Whitman, J., Beaton, D., Boers, M., Wells, G., Beaton, D., Katz, N., Fossel, A., Wright, J., Tarasuk, V., Bombardier, C., Michener, L., Leggins, B., Gabel, C., Michener, L., Burkett, B., Neller, A., Hudak, P., Amadio, P., Bombardier, C., Davis, A., Beaton, D., Hudak, P., Amadio, P., Bombardier, C., Cole, C., Hawker, G., Katz, J., Makela, M., Marx, R., MacDermid, J., Tottenham, V., Bialocerkowski, A., Bot, S., Terwee, C., van der Windt, D., Bouter, L., Dekker, J., de Vet, H., Mousavi, S., Parnianpour, M., Abedi, M., Askary-Ashtiani, A., Karimi, A., Khorsandi, A., Mehdian, H., Beaton, D., Wright, J., Katz, J., Group, U., Wyatt, M., Veale, G., Atroshi, I., Lyrén, P., Gummesson, C., Mintken, P., Glynn, P., Cleland, J., Imaeda, T., Toh, S., Wada, T., Uchiyama, S., Okinaga, S., Kusunose, K., Sawaizumi, T., Fayad, F., Lefevre-Colau, M., Gautheron,

V., Macé, Y., Fermanian, J., Mayoux-Benhamou, A., Roren, A., Rannou, F., Roby-Brami, A., Revel, M., Meads, D., Doward, L., McKenna, S., Fisk, J., Twiss, J., Eckert, B., Matheson, L., Melhorn, J., Mayer, T., Theodore, B., Gatchel, R., Wu, A., Edgar, D., Wood, F., Stover, B., Silverstein, B., Wickizer, T., Martin, D., Kaufman, J., Abramo, A., Kopylov, P., Tagil, M., Gabel, C., Michener, L., Melloh, M., Burkett, B., Gabel, C., Burkett, B., Yelland, M., McConnell, S., Beaton, D., Bombardier, C., Gabel, P., Barden, L., Burkett, B., Neller, L., O'Halloran, J., Miller, G., Britt, H., Shrout, P., Fleiss, J., Liang, M., Liang, M., Jacobson, N., Roberts, L., Berns, S., McGlinchey, J., Cronbach, L., Nunnally, J., Bernstein, I., Fields, A., Stevens, J., Cattell, R., Kaiser, H., Dawson, B., Trapp, R., Doak, C., Doak, L., Root, J., Paasche-Orlow, M., Taylor, H., Brancati, F., Vet, H. D., Ostelo, R., Terwee, C., van der Roer, N., Knol, D., Beckerman, H., Boers, M., Bouter, L., Lee, E., Chung, M., Li, A., Lo, S., MacDermid, J., Wessel, J., Humphrey, R., Ross, D., Roth, J., Themistocleous, G., Goudelis, G., Kyrou, I., Chloros, G., Krokos, A., Galanos, A., Gerostathopoulos, N., Soucacos, P., Reeve, B., Hays, R., Bjorner, J., Cook, K., Crane, P., Teresi, J., Thissen, D., Revicki, D., Weiss, D., Hambleton, R., Coste, J., Guillemin, F., Pouchot, J., Fermanian, J., Young, S., Aprill, C., Braswell, J., Ogard, W., Richards, J., McCarthy, J., van der Velde, G., Beaton, D., Hogg-Johnston, S., Hurwitz, E., and Tennant, A. (2009). A modified QuickDASH-9 provides a valid outcome instrument for upper limb function. *BMC Musculoskeletal Disorders*, 10(1):161.

Galalae, R. M., Michel, J., Siebmann, J. U., Küchler, T., Eilf, K., and Kimmig, B. (2005). Significant negative impact of adjuvant chemotherapy on health-related quality of life (HR-QoL) in women with breast cancer treated by conserving surgery and postoperative 3-D radiotherapy: A prospective measurement. *Strahlenther Onkol*, 181:645–651.

García-Cruz, E., Romero Otero, J., Martínez Salamanca, J. I., Leibar Tamayo, A., Rodríguez Antolín, A., Astobieta Odriozola, A., and Alcaraz, A. (2011). Linguistic and Psychometric Validation of the Erection Hardness Score to Spanish. *J Sex Med*, 8:470–474.

Garratt, A., Schmidt, L., Mackintosh, A., and Fitzpatrick, R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *Bmj*, 324(7351):1417.

Gierveld, J. D. J. and Van Tilburg, T. (2006). A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. *Res Aging*, 28:582–598.

Girotto, J. and Schreiber, J. (2003). Breast reconstruction in the elderly: preserving excellent quality of life. *Annals of plastic surgery*.

Glangkarn, S., Promasatayaprot, V., Porock, D., and Edgley, A. (2011). Measuring quality of life in Thai women with breast cancer. *Asian Pac J Cancer Preven*, 12:637–644.

Gómez-Campelo, P., Bragado-Álvarez, C., Hernández-Lloreda, M. J., and Sánchez-Bernardos, M. L. (2014). The Spanish version of the Body Image Scale (S-BIS): psychometric properties in a sample of breast and gynaecological cancer patients. *Supportive Care Cancer*, 23:473–481.

Gonzáles, A. I., Sties, S. W., Wittkopf, P. G., de Mara, L. S., Ulbrich, A. Z., Cardoso, F. L., and de Carvalho, T. (2013). Validation of the International Index of Erectile Function (IIFE) for use in Brazil. *Arq Bras Cardiol*, 101:176–181.

Gopie, J. P., ter Kuile, M. M., Timman, R., Mureau, M. A. M., and Tibben, A. (2014). Impact of delayed implant and DIEP flap breast reconstruction on body image and sexual satisfaction: a prospective follow-up study. *Psycho-Oncology*, 23:100–107.

Grice, J. W. (1997). On the validity of the perceived understanding instrument. *Psychological Reports*, 80:1007–1010.

Grygiel, P., Humenny, G., and Rebisz, S. (2016). Using the De Jong Gierveld Loneliness Scale With Early Adolescents: Factor Structure, Reliability, Stability, and External Validity. *Assessment*.

Grygiel, P., Humenny, G., Rebisz, S., Switaj, P., and Sikorska, J. (2013). Validating the Polish Adaptation of the 11-Item De Jong Gierveld Loneliness Scale. *European Journal of Psychological Assessment*, 29:129–139.

Gujral, S., Conroy, T., Fleissner, C., Sezer, O., King, P. M., Avery, K. N. L., Sylvester, P., Koller, M., Sprangers, M. A. G., and Blazeby, J. M. (2007). Assessing quality of life in patients with colorectal cancer: An update of the EORTC quality of life questionnaire. *Eur J Cancer*, 43:1564–1573.

Gummesson, C., Ward, M. M., and Atroshi, I. (2006). The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. *BMC musculoskeletal disorders*, 7:44.

Harcourt, D., Rumsey, N., and Ambler, N. (2003). The psychological effect of mastectomy with or without breast reconstruction: a prospective, multicenter study. *Plastic and*.

Heil, J., Holl, S., Golatta, M., Rauch, G., Rom, J., Marm??, F., Gebauer, G., and Sohn, C. (2010). Aesthetic and functional results after breast conserving surgery as correlates of quality of life measured by a German version of the Breast Cancer Treatment Outcome Scale (BCTOS). *Breast*, 19(6):470–474.

Herrmann, C. (1997). International experiences with the Hospital Anxiety and Depression Scale-A review of validation data and clinical results. *Journal of Psychosomatic Research*, 42(1):17–41.

Hjemboe, S. and Butcher, J. N. (1991). Couples in marital distress: a study of personality factors as measured by the MMPI-2. *J Pers Assess*, 57:216–237.

Hollist, C. S., Falceto, O. G., Ferreira, L. M., Miller, R. B., Springer, P. R., Fernandes, C. L. C., and Nunes, N. A. (2012). Portuguese translation and validation of the revised dyadic adjustment scale. *J Marital Fam Ther*, 38:348–358.

Holly, P., Kennedy, P., Taylor, A., and Beedie, A. (2003). Immediate breast reconstruction and psychological adjustment in women who have undergone surgery for breast cancer: a preliminary study. *Psychology, Health & Medicine*, 8(4):441–452.

Hopwood, P., Fletcher, I., Lee, A., and Al Ghazal, S. (2001). A body image scale for use with cancer patients. *European Journal of Cancer*, 37:189–197.

Hou, X. t., Pang, D., Lu, Q., Yang, P., Jin, S. l., Zhou, Y. j., and Tian, S. h. (2015). Validation of the Chinese version of the low anterior resection syndrome score for measuring bowel dysfunction after sphincter-preserving surgery among rectal cancer patients. *Eur J Oncol Nurs*, 19:495–501.

Hugo, S., Chedraui, P., Pérez-López, F. R., Aguirre, W., Martínez, N., Miranda, O., Schwager, G., Narváez, J., Quintero, J. C., and Zambrano, B. (2011). Sexual function assessment in mid-aged women: The short 6-item female sexual function index. *Climacteric*, 14:183–184.

Hunsley, J., Best, M., and Lefebvre, M. (2001). The seven-item short form of the Dyadic Adjustment Scale: Further evidence for construct validity. *American Journal of Family*.

Hunsley, J., Pinsent, C., Lefebvre, M., Jamestanner, S., and Vito, D. (1995). Construct-Validity of the Short Forms of the Dyadic Adjustment Scale. *Family Relations*, 44:231–237.

Hwang, T. I. S., Tsai, T. E., Lin, Y. I. C., Chiang, H. S., and Chang, L. S. (2010). A survey of erectile dysfunction in Taiwan: Use of the erection hardness score and quality of erection questionnaire. *J Sex Med*, 7:174.

Ibbotson, T., Maguire, P., Selby, P., and Priestman, T. (1994). Screening for anxiety and depression in cancer patients: the effects of disease and treatment. *European Journal of*.

Idrovo, A. J., Camacho-Avila, A., García-Rivas, J., and Juárez-García, A. (2012). Social capital at work: Psychometric analysis of a short scale in Spanish among Mexican health workers. *Rev Bras Epidemiol*, 15:536–547.

Iecovich, E. (2013). Psychometric Properties of the Hebrew Version of the de Jong Gierveld Loneliness Scale. *Educational Gerontology*, 39:12–27.

Ihn, M. H., Lee, S. M., Son, I. T., Park, J. T., Oh, H. K., Kim, D. W., and Kang, S. B. (2015). Cultural adaptation and validation of the Korean version of the EORTC QLQ-CR29 in patients with colorectal cancer. *Supportive Care Cancer*, 23:3493–3501.

Irniza, R., Emilia, Z. A., Muhammad Saliluddin, S., and Nizam Isha, A. S. (2014). A psychometric properties of the Malay-version police stress questionnaire. *Malays J Med Sci*, 21:42–50.

Isanezhad, O., Ahmadi, S. A., Bahrami, F., Baghban-Cichani, I., Farajzadegan, Z., and Etemadi, O. (2012). Factor structure and reliability of the revised dyadic adjustment scale (RDAS) in Iranian population. *Iran J Psychiatr Behav Sci*, 6:55–61.

Isidori, A. M., Pozza, C., Esposito, K., Giugliano, D., Morano, S., Vignozzi, L., Corona, G., Lenzi, A., and Jannini, E. A. (2010). Development and validation of a 6-item version of the female sexual function index (FSFI) as a diagnostic tool for female sexual dysfunction. *J Sex Med*, 7:1139–1146.

Jansen, F., van Uden-Kraan, C. F., Van Zwieten, V., Witte, B. I., Leemans, C. R., and Verdonck-de Leeuw, I. M. (2015). Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. *Support Care Cancer*, 23(6):1679–1688.

Jayasekara, H., Rajapaksa, L. C., and Brandberg, Y. (2008). Measuring breast cancer-specific health-related quality of life in South Asia: Psychometric properties of the Sinhala version of the EORTC QLQ-BR23. *Quality of Life Research*, 17(6):927–932.

Karakurt, G., Erguner-Tekinalp, B., and Terzi, S. (2009). The Psychometric Qualities of the Turkish Version of Emotional Abuse Questionnaire. *Egitim Arastirmalari-Eurasian Journal of Educational Research*, 9:7–22.

Karasek, R., Brisson, C., and Kawakami, N. (1998). The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. *of occupational health ….*

Kazak, A., Jarmas, A., and Snitzer, L. (1988). The assessment of marital satisfaction: An evaluation of the Dyadic Adjustment Scale. *Journal of Family Psychology*.

Kennedy, C. A., Beaton, D. E., Smith, P., Van Eerd, D., Tang, K., Inrig, T., Hogg-Johnson, S., Linton, D., and Couban, R. (2013). Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Quality of Life Research*, 22(9):2509–2547.

Khang, D., Rim, H. D., and 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.

Klijn, W. (2013). *Gezinskenmerken: De constructie van de Vragenlijst Gezinskenmerken (VGK)*. PhD thesis, University of Amsterdam.

Kontodimopoulos, N., Ntinoulis, K., and Niakas, D. (2011). Validity of the Greek EORTC QLQ-C30 and QLQ-BR23 for measuring health-related quality of life in breast cancer patients. *European Journal of Cancer Care*, 20(3):354–361.

Korus, L. J., Cypel, T., Zhong, T., and Wu, A. W. (2015). Patient-Reported Outcome Measures in Reconstructive Breast Surgery: Is There a Role for Generic Measures? *Plastic and Reconstructive Surgery*, 135:479E–490E.

Kotronoulas, G. C., Papadopoulou, C. N., Papapetrou, A., and Patiraki, E. (2011). Psychometric evaluation and feasibility of the Greek Pittsburgh sleep quality index (GR-PSQI) in patients with cancer receiving chemotherapy. *Supportive Care Cancer*, 19:1831–1840.

Koukouli, S., STAMOU, A., ALEGAKIS, A., GEORGOULIAS, V., and SAMONIS, G. (2009). Psychometric properties of the QLQ-C30 (version 3.0) in a sample of ambulatory Cretan cancer patients. *European Journal of Cancer Care*, 18(5):447–456.

Kriston, L., Günzler, C., Harms, A., and Berner, M. (2008). Confirmatory factor analysis of the German version of the International Index of Erectile Function (IIEF): A comparison of four models. *J Sex Med*, 5:92–99.

Kvam, A. K., Fayers, P. M., and Wisloff, F. (2011). Responsiveness and minimal important score differences in quality-of-life questionnaires: a comparison of the EORTC QLQ-C30 cancer-specific questionnaire to the generic utility questionnaires EQ-5D and 15D in patients with multiple myeloma. *European Journal of Haematology*, 87(4):330–337.

Lau, Y., Wang, Y., Yin, L., Chan, K. S., and Guo, X. (2010). Validation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale in Chengdu mothers. *Int J Nurs Stud*, 47:1139–1151.

Lee, E. H. (2013). Review of the Psychometric Evidence of the Perceived Stress Scale. *Asian Nursing Research*, 7:160.

Lesch, E. and Engelbrecht, S. K. (2008). The reliability of the Dyadic Satisfaction (DS) subscale in a low-income semi-rural South African community. *Journal of Psychology in Africa*, 18:245–248.

Lester, J., Bernhard, L., and Ryan-Wenger, N. (2012). A self-report instrument that describes urogenital atrophy symptoms in breast cancer survivors. *West J Nurs Res*, 34:72–96.

Leung, G. T. Y., De Jong Gierveld, J., and Lam, L. C. W. (2008). Validation of the Chinese translation of the 6-item De Jong Gierveld Loneliness Scale in elderly Chinese. *Int Psychogeriatr*, 20:1262–1272.

Lim, B. K. and Ivey, D. (2000). The assessment of marital adjustment with Chinese populations: A study of the psychometric properties of the dyadic adjustment scale. *Contemporary Family Therapy*, 22:453–465.

Lim, T. O., Das, A., Rampal, S., Zaki, M., Sahabudin, R. M., Rohan, M. J., and Isaacs, S. (2003). Cross-cultural adaptation and validation of the English version of the International Index of Erectile Function (IIEF) for use in Malaysia. *Int J Impotence Res*, 15:329–336.

Lin, C. Y., Pakpour, A. H., Burri, A., and Montazeri, A. (2016). Rasch Analysis of the Premature Ejaculation Diagnostic Tool (PEDT) and the International Index of Erectile Function (IIEF) in an Iranian Sample of Prostate Cancer Patients. *PLoS ONE*, 11:e0157460.

Lin, J., L., Z., D.-W., W., Z.-H., X., X.-J., W., Y.-S., L., W., F., J.-R., T., M., W., P., P., A., G., Z., Y., L., L., L.-Y., J., Q.-Q., L., X.-Y., Z., Y.-F., Z., H.-W., X., B., Y., X.-L., L., Y.-X., L., Lin, J.-B., Zhang, L., Wu, D.-W., Xi, Z.-H., Wang, X.-J., Lin, Y.-S., Fujiwara, W., Tian, J.-R., Wang, M., Peng, P., Guo, A., Yang, Z., Luo, L., Jiang, L.-Y., Li, Q.-Q., Zhang, X.-Y., Zhang, Y.-F., Xu, H.-W., Yang, B., Li, X.-L., and Lei, Y.-X. (2017). Validation of the Chinese version of the EORTC QLQ-CR29 in patients with colorectal cancer. *World Journal of Gastroenterology*, 23(10):1891–1898.

Lubberding, S., van Uden-Kraan, C. F., Te Velde, E. A., Cuijpers, P., Leemans, C. R., and Verdonck-de Leeuw, I. M. (2015). Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. *Journal of clinical nursing*, 24(9-10):1367–1379.

Luo, N., Fones, C. S. L., Lim, S. E., Xie, F., Thumboo, J., and Li, S. C. (2005). The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30): Validation of English version in Singapore. *Quality of Life Research*, 14(4):1181–1186.

Magaji, B. A., Moy, F. M., Roslani, A. C., Law, C. W., Raduan, F., and Sagap, I. (2015). Psychometric Validation of the Bahasa Malaysia Version of the EORTC QLQ-CR29. *Asian Pac J Cancer Prev*, 16:8101–8105.

Mahmood, M. A., Ur Rehman, K., Khan, M. A., and Sultan, T. (2012). Translation, Cross-Cultural Adaptation, and Psychometric Validation of the 5-Item International Index of Erectile Function (IIEF-5) into Urdu. *J Sex Med*, 9:1883–1886.

Manganiello, A., Hoga, L. A. K., Reberte, L. M., Miranda, C. M., and Rocha, C. A. M. (2011). Sexuality and quality of life of breast cancer patients post mastectomy. *European Journal of Oncology Nursing*, 15(2):167–172.

Manne, S. and Schnoll, R. (2001). Measuring cancer patients' psychological distress and well-being: A factor analytic assessment of the mental health inventory. *Psychol Assess*, 13:99–109.

Maringwa, J., Quinten, C., King, M., Ringash, J., Osoba, D., Coens, C., Martinelli, F., Reeve, B. B., Gotay, C., Greimel, E., Flechtner, H., Cleeland, C. S., Schmucker-Von Koch, J., Weis, J., Van Den Bent, M. J., Stupp, R., Taphoorn, M. J., Bottomley, A., EORTC PROBE Project and Brain Cancer Group, o. b. o. t. E. P. P., and Cancer, B. (2011a). Minimal clinically meaningful differences for the EORTC QLQ-C30 and EORTC QLQ-BN20 scales in brain cancer patients. *Annals of oncology : official journal of the European Society for Medical Oncology*, 22(9):2107–12.

Maringwa, J. T., Quinten, C., King, M., Ringash, J., Osoba, D., Coens, C., Martinelli, F., Vercauteren, J., Cleeland, C. S., Flechtner, H., Gotay, C., Greimel, E., Taphoorn, M. J., Reeve, B. B., Koch, J. S.-V., Weis, J., Smit, E. F., van Meerbeeck, J. P., Bottomley, A., Group, o. b. o. t. E. P. p., and the Lung Cancer (2011b). Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. *Supportive Care in Cancer*, 19(11):1753–1760.

McHorney, C., Jr, J. W., and Raczek, A. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical care*.

Mehta, S., MacDermid, J. C., Carlesso, L. C., and McPhee, C. (2010). Concurrent Validation of the DASH and the QuickDASH in Comparison to Neck-Specific Scales in Patients With Neck Pain. *Spine*, 35(24):2150–2156.

Michels, F. A. S., Latorre, M. R. D. O., and Maciel, M. D. S. (2013). Validity, reliability and understanding of the EORTC-C30 and EORTC-BR23, quality of life questionnaires specific for breast cancer. *Rev Bras Epidemiol*, 16:352–363.

Mintken, P. E., Glynn, P., and Cleland, J. A. (2009). Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. *Journal of Shoulder and Elbow Surgery*, 18(6):920–926.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., and de Vet, H. C. (2012). *COSMIN checklist manual*. Manual, VU University Medical Center.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., and De Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, 19(4):539–549.

Montazeri, A., Harirchi, I., Vahdani, M., Khaleghi, F., Jarvandi, S., Ebrahimi, M., and Haji-Mahmoodi, M. (2000). The EORTC breast cancer-specific quality of life questionnaire (EORTC QLQ-BR23): Translation and validation study of the Iranian version. *Qual Life Res*, 9:177–184.

Moorey, S., Greer, S., Watson, M., and Gorman, C. (1991). The factor structure and factor stability of the hospital anxiety and depression scale in patients with cancer. *The British Journal of*.

Moreira, H., Silva, S., Marques, A., and Canavarro, M. C. (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.

Mulhall, J. P., King, R., Kirby, M., Hvidsten, K., Symonds, T., Bushmakin, A. G., and Cappelleri, J. C. (2008). Evaluating the sexual experience in men: Validation of the sexual experience questionnaire. *J Sex Med*, 5:365–376.

Mullan, M., Wilkins, E., Goldfarb, S., Lowery, J., Smith, D., Wickman, M., and Sandelin, K. (2007). Prospective analysis of psychosocial outcomes after breast reconstruction: Cross-cultural comparisons of 1-year postoperative results. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 60(5):503–508.

Nelson, J. a., Fischer, J. P., Pasick, C., Nelson, P., Chen, A. J., Fosnot, J., Selber, J. C., Wu, L. C., and Serletti, J. M. (2013). Chronic Pain Following Abdominal Free Flap Breast Reconstruction. *Annals of Plastic Surgery*, 71(3):278–282.

Nicholson, R., Leinster, S., and Sassoon, E. (2007). A comparison of the cosmetic and psychological outcome of breast reconstruction, breast conserving surgery and mastectomy without reconstruction. *The Breast*, 16(4):396–410.

Niedhammer, I. (2002). Psychometric properties of the French version of the Karasek Job Content Questionnaire: A study of the scales of decision latitude, psychological demands, social support, and physical demands in the GAZEL cohort. *Int Arch Occup Environ Health*, 75:129–144.

Nowak, W., Tobiasz-Adamczyk, B., Brzyski, P., Sałówka, J., Kuliś, D., and Richter, P. (2011). Adaptation of quality of life module EORTC QLQ-CR29 for polish patients with rectal cancer - Initial assessment of validity and raeliability. *Pol Prz Chir Pol J Surg*, 83:502–510.

Obtel, M., Serhier, Z., and Bendahhou, K. (2017). Validation of EORTC IN-PATSAT 32 in Morocco: Methods and Processes. *Asian Pacific journal*, 18(5):1403–1409.

Oxman, A. D. and Guyatt, G. H. (1991). Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology*, 44(11):1271–1278.

Oxman, A. D., Guyatt, G. H., Singer, J., Goldsmith, C. H., Hutchison, B. G., Milner, R. A., and Streiner, D. L. (1991). Agreement among reviewers of review articles. *Journal of Clinical Epidemiology*, 44(1):91–98.

Parisot, J., Yiou, R., Salomon, L., de la Taille, A., Lingombet, O., and Audureau, E. (2014). Erection hardness score for the evaluation of erectile dysfunction: Further psychometric assessment in patients treated by intracavernous prostaglandins injections after radical prostatectomy. *J Sex Med*, 11:2109–2118.

Parmar, V., Badwe, R. A., Hawaldar, R., Rayabhattanavar, S., Varghese, A., Sharma, R., and Mittra, I. (2005). Validation of EORTC quality-of-life questionnaire in Indian women with operable breast cancer. *National Medical Journal of India*, 18(4):172–177.

Pascoe, J. M. and French, J. (1990). The reliability and validity of the maternal social support index for primiparous mothers: A brief report. *FAM MED*, 22:228–230.

Pascoe, J. M., Ialongo, N. S., Horn, W. F., Reinhart, M. A., and Perradatto, D. (1988). The reliability and validity of the Maternal Social Support Index. *FAM MED*, 20:271–276.

Pelfrene, E., Vlerick, P., Mak, R. P., De Smet, P., Kornitzer, M., and De Backer, G. (2001). Scale reliability and validity of the Karasek 'Job Demand-Control-Support model in the belstress study. *Work Stress*, 15:297–313.

Penning, M. J., Liu, G. P., and Chou, P. H. B. (2014). Measuring Loneliness Among Middle-Aged and Older Adults: The UCLA and de Jong Gierveld Loneliness Scales. *Social Indicators Research*, 118:1147–1166.

Pérez-López, F. R., Fernández-Alonso, A. M., Trabalón-Pastor, M., Vara, C., and Chedraui, P. (2012). Assessment of sexual function and related factors in mid-aged sexually active Spanish women with the six-item Female Sex Function Index. *Menopause*, 19:1224–1230.

Pishkuhi, M. A., Salmaniyan, S., Nedjat, S., Zendedel, K., and Lari, M. A. (2014). Psychometric properties of the Persian version of satisfaction with care EORTC-in-patsat32 questionnaire among Iranian cancer patients. *Asian Pacific Journal of Cancer Prevention*, 15(23):10121–10128.

Polson, K., Reid, D., McNair, P. J., and Larmer, P. (2010). Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. *Manual Therapy*, 15(4):404–407.

Potter, S., Thomson, H. J., Greenwood, R. J., Hopwood, P., and Winters, Z. E. (2009). Health-related quality of life assessment after breast reconstruction. *British Journal of Surgery*, 96(6):613–620.

Prieto, L., Thorsen, H., and Juul, K. (2005). Development and validation of a quality of life questionnaire for patients with colostomy or ileostomy. *Health and quality of life*.

Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., and Terwee, C. (2017). COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. *Manuscript submitted for publication.*

Pusic, a., Thompson, T. a., Kerrigan, C. L., Sargeant, R., Slezak, S., Chang, B. W., Kelzlsouer, K. J., and Manson, P. N. (1999). Surgical options for the early-stage breast cancer: factors associated with patient choice and postoperative quality of life.

Quatman-Yates, C. C., Gupta, R., Paterno, M. V., Schmitt, L. C., Quatman, C. E., and Ittenbach, R. F. (2013). Internal Consistency and Validity of the QuickDASH Instrument for Upper Extremity Injuries in Older Children. *Journal of Pediatric Orthopaedics*, 33(8):838–842.

Quek, K. F., Low, W. Y., Razack, A. H., Chua, C. B., and Loh, C. S. (2002). Reliability and validity of the Malay version of the International Index of Erectile Function (IIEF-15) in the Malaysian population. *Int J Impotence Res*, 14:310–315.

Quinta Gomes, A. L. and Nobre, P. (2012). The International Index of Erectile Function (IIEF-15): Psychometric Properties of the Portuguese Version. *J Sex Med*, 9:180–187.

Razavi, D. (1990). Screening for Adjustment Disorders and Major Depressive Disorders in Cancer In-Patients. *Br J Psychiatry*, 156:79–83.

Rhondali, W., Chisholm, G., Filbet, M., Kang, D. H., Hui, D., Cororve Fingeret, M., and Bruera, E. (2014). Screening for body image dissatisfaction in advanced cancer patients. *Palliative Med*, 28:572.

Rosen, R. C., Allen, K. R., Ni, X., and Araujo, A. B. (2011). Minimal clinically important differences (MCID) in the erectile function (EF) domain of the international index of erectile function (IIEF). *J Urol*, 185:e615.

Rosen, R. C., Cappelleri, J. C., Smith, M. D., Lipsky, J., and Peña, B. M. (2000). Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. *International Journal of Impotence Research*, 11(6):319–326.

Rosen, R. C., Riley, A., Wagner, G., Osterloh, I. H., Kirkpatrick, J., and Mishra, A. (1997). The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology*, 49(6):822–830.

Rossier, J., Rigozzi, C., Charvoz, L., and Bodenmann, G. (2006). Marital satisfaction: Psychometric properties of the PFB and comparison with the DAS. *Swiss Journal of Psychology*, 65:55–63.

Rowland, J. H., Desmond, K. A., Meyerowitz, B. E., Belin, T. R., Wyatt, G. E., and Ganz, P. A. (2000). Role of breast reconstructive surgery in physical and emotional outcomes among breast cancer survivors. *Journal of the National Cancer Institute*, 92(17):1422–9.

Roy, J.-S., MacDermid, J. C., Amick, B. C., Shannon, H. S., McMurtry, R., Roth, J. H., Grewal, R., Tang, K., and Beaton, D. (2011). Validity and Responsiveness of Presenteeism Scales in Chronic Work-Related Upper-Extremity Disorders. *Physical Therapy*, 91(2).

Rubio-Aurioles, E., Sand, M., Terrein-Roccatti, N., Dean, J., Longworth, J., Eardley, I., Brock, G., Lee, J., Arango De Montis, I., and Rampazzo-Bonaldo, C. (2009). Female assessment of male erectile dysfunction detection scale (FAME): Development and validation. *J Sex Med*, 6:2255–2270.

Sabourin, S., Valois, P., and Lussier, Y. (2005). Development and validation of a brief version of the dyadic adjustment scale with a nonparametric item analysis model. *Psychol Assess*, 17:15–27.

Saffari, M., Pakpour, A. H., and Burri, A. (2016). Cross-Cultural Adaptation of the Male Genital Self-Image Scale in Iranian Men. *Sex Med*, 4:e34–e42.

Sanford, K. (1998). Memories of feeling misunderstood, and a schema of partner empathic responding: New scales for marital research. *Journal of Social and Personal Relationships*, 15:490–501.

Sanna, B., Bereza, K., Paradowska, D., Kucharska, E., Tomaszewska, I., Dudkiewicz, Z., Golec, J., Bottomley, A., and Tomaszewski, K. (2017). A large scale prospective clinical and psychometric validation of the EORTC colorectal (QLQ-CR29) module in Polish patients with colorectal cancer. *European Journal of Cancer Care*, page e12713.

Savard, M. H., Savard, J., Simard, S., and Ivers, H. (2005). Empirical validation of the insomnia severity index in cancer patients. *Psycho-Oncology*, 14:429–441.

Scarpa, M., Valente, S., Alfieri, R., Cagol, M., Diamantis, G., Ancona, E., and Castoro, C. (2011). Systematic review of health-related quality of life after esophagectomy for esophageal cancer. *World J Gastroenterol*, 17:4660–4674.

Seppä, K. and Lepistö, J. (1998). Five-Shot Questionnaire on Heavy Drinking. *Experimental Research*, 22(8):1788–1791.

Serefoglu, E. C., Atmaca, A. F., Dogan, B., Altinova, S., Akbulut, Z., and Balbay, M. D. (2008). Problems in understanding the Turkish translation of the international index of Erectile Function. *J Androl*, 29:369–373.

Sharpley, C. and Cross, D. (1982). A psychometric evaluation of the Spanier Dyadic Adjustment Scale. *Journal of Marriage and the Family*.

Shek, D. T. L. (1994). Psychometric Properties of the Chinese Version of the Dyadic Adjustment Scale. *Psychologia*, 37:7–17.

Shek, D. T. L. (1998). Reliability and validity of the Kansas martial satisfaction scale for Chinese parents. *Psychological Reports*, 83:81–82.

Shek, D. T. L. and Cheung, C. K. (2008). Dimensionality of the Chinese Dyadic Adjustment Scale based on confirmatory factor analyses. *Social Indicators Research*, 86:201–212.

Singer, S., Araújo, C., Arraras, J. I., Baumann, I., Boehm, A., Brokstad Herlofson, B., Castro Silva, J., Chie, W.-C., Fisher, S., Guntinas-Lichius, O., Hammerlid, E., Irarrázaval, M. E., Jensen Hjermstad, M., Jensen, K., Kiyota, N., Licitra, L., Nicolatou-Galitis, O., Pinto, M., Santos, M., Schmalz, C., Sherman, A. C., Tomaszewska, I. M., Verdonck de Leeuw, I., Yarom, N., Zotti, P., and Hofmeister, D. (2015). Measuring quality of life in patients with head and neck cancer: Update of the EORTC QLQ-H{&}N Module, Phase III. *Head {&} Neck*, 37(9):1358–1367.

Smith, A. B., Selby, P. J., Velikova, G., Stark, D., Wright, E. P., Gould, A., and Cull, A. (2002). Factor analysis of the Hospital Anxiety and Depression Scale from a large cancer population. *Psychology and Psychotherapy: Theory, Research and Practice*, 75(2):165–176.

South, S. C., Krueger, R. F., and Iacono, W. G. (2009). Factorial Invariance of the Dyadic Adjustment Scale Across Gender. *Psychol Assess*, 21:622–628.

Spanier, G. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*.

Spliethoff-Kamminga, N. G. A., Zwinderman, A. H., Springer, M. P., and Roos, R. A. C. (2003). A disease-specific psychosocial questionnaire for Parkinson's disease caregivers. *J Neurol*, 250:1162–1168.

Sprangers, M. A., Groenvold, M., Arraras, J. I., Franklin, J., te Velde, A., Muller, M., Franzini, L., Williams, A., de Haes, H. C., Hopwood, P., Cull, A., and Aaronson, N. K. (1996). The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: First results from a three-country field study. *Journal of Clinical Oncology*, 14(10):2756–2768.

Stevenson-Hinde, J. and Akister, J. (1995). The McMaster Model of Family Functioning: observer and parental ratings in a nonclinical sample. *Fam Process*, 34:337–347.

Stiggelbout, A. M., Kunneman, M., Baas-Thijssen, M. C. M., Neijenhuis, P. A., Loor, A. K., Jägers, S., Vree, R., Marijnen, C. A. M., and Pieterse, A. H. (2016). The EORTC QLQ-CR29 quality of life questionnaire for colorectal cancer: validation of the Dutch version. *Qual Life Res*, 25:1853–1858.

Stover, B., Silverstein, B., Wickizer, T., Martin, D. P., and Kaufman, J. (2007). Accuracy of a Disability Instrument to Identify Workers Likely to Develop Upper Extremity Musculoskeletal Disorders. *Journal of Occupational Rehabilitation*, 17(2):227–245.

Stuiver, M. M., Tusscher, M. R. T., Van Opzeeland, A., Brendeke, W., Lindeboom, R., Dijkstra, P. U., and Aaronson, N. K. (2016). Psychometric properties of 3 patient-reported outcome measures for the assessment of shoulder disability after neck dissection. *Head Neck*, 38:102–110.

Tang, Y., Wang, Y., Zhu, H., Jiang, X., Gan, Y., and Yang, J. (2015). Bias in Evaluating Erectile Function in Lifelong Premature Ejaculation Patients with the International Index of Erectile Function-5. *J Sex Med*, 12:2061–2069.

Tantillo, M. and Sanftner, J. L. (2010). Measuring perceived mutuality in women with eating disorders: The development of the connection-disconnection scale. *J Nurs Meas*, 18:100–119.

Taylor, J. M. (2015). Psychometric analysis of the ten-item perceived stress scale. *Psychol Assess*, 27:90–101.

Temple, C. L. F., Ross, D. C., Kim, S., Tse, R., Bettger-Hahn, M., Gan, B. S., and MacDermid, J. (2009). Sensibility following innervated free TRAM flap for breast reconstruction: Part II. Innervation improves patient-rated quality of life. *Plastic and reconstructive surgery*, 124(5):1419–1425.

Temple-Oberle, C. F., Ayeni, O., Cook, E. F., Bettger-Hahn, M., Mychailyshyn, N., and Macdermid, J. (2013). The breast reconstruction satisfaction questionnaire (BRECON-31): An affirmative analysis. *Journal of Surgical Oncology*, 107:451–455.

Temple-Oberle, C. F., Cook, E. F., Bettger-Hahn, M., Mychailyshyn, N., Naeem, H., and Macdermid, J. (2012). Development of a breast reconstruction satisfaction questionnaire (BRECON-31): Principal components analysis and clinimetric properties. *Journal of Surgical Oncology*, 106:799–806.

Terwee, C. B., Jansma, E. P., Riphagen, I. I., and De Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, 18(8):1115–1123.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W. J. G., Bouter, L. M., and De Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, 21(4):651–657.

Tønseth, K., Hokland, B., Tindholdt, T., Åbyholm, F., and Stavem, K. (2008). Quality of life, patient satisfaction and cosmetic outcome after breast reconstruction using DIEP flap or expandable breast implant. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 61(10):1188–1194.

Utomo, E., Blok, B. F., Pastoor, H., Bangma, C. H., and Korfage, I. J. (2015). The measurement properties of the five-item International Index of Erectile Function (IIEF-5): A Dutch validation study. *Andrology*, 3:1154–1159.

Uwer, L., Rotonda, C., Guillemin, F., Miny, J., Kaminsky, M.-C., Mercier, M., Tournier-Rangeard, L., Leonard, I., Montcuquet, P., Rauch, P., Conroy, T., Guyatt, G., Ferrans, C., Halyard, M., Revicki, D., Symonds, T., Varricchio, C., Kotzeva, A., Valderas, J., Alonso, J., Garratt, A., Schmidt, L., Mackintosh, A., Fitzpatrick, R., Aaronson, N., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N., Filiberti, A., Flechtner, H., Fleishman, S., de Haes, J., Osoba, D., Zee, B., Pater, J., Warr, D., Kaizer, L., Latreille, J., Sprangers, M., te Velde, A., Aaronson, N., Cella, D., Tulsky, D., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silberman, M., Yellen, S., Winicour, P., Brannon, J., Ward, W., Hahn, E., Mo, F., Hernandez, L., Tulsky, D., Cella, D., Conroy, T., Mercier, M., Bonneterre, J., Luporsi, E., Lefebvre, J., Lapeyre, M., Puyraveau, M., Schraub, S., Guyatt, G., Feeny, D., Patrick, D., Hawley, D., Wolfe, F., Katz, J., Larson, M., Phillips, C., Fossel, A., Liang, M., Kazis, L., Anderson, J., Meenan, R., Guyatt, G., Walter, S., Norman, G., Guyatt, G., Deyo, R., Charlson, M., Levine, M., Mitchell, A., Conroy, T., Blazeby, J., Guren, M., Dueland, S., Skovlund, E., Fossa, S., Poulsen, J., Tveit, K., Osoba, D., Rodrigues, G., Myles, J., Zee, B., Pater, J., Jaeschke, R., Singer, J., Guyatt, G., Fleiss, J., Shrout, P., Cohen, J., Katz, J., Phillips, C., Fossel, A., Liang, M., Beaton, D., Bombardier, C., Katz, J., Wright, J., Wasserman, S., Bockenholt, U., Deyo, R., Diehr, P., Patrick, D., Terwee, C., Dekker, F., Wiersinga, W., Prummel, M., Bossuyt, P., Bombardier, C., Melfi, C., Paul, J., Green, R., Hawker, G., Wright, J., Coyte, P., Yost, K., Cella, D., Chawla, A., Holmgren, E., Eton, D., Ayanian, J., West, D., Gujral, S., Conroy, T., Fleissner, C., Sezer, O., King, P., Avery, K., Sylvester, P., Koller, M., Sprangers, M., Blazeby, J., Whistance, R., Conroy, T., Chie, W., Costantini, A., Sezer, O., Koller, M., Johnson, C., Pilkington, S., Arraras, J., Josef, E. B., Pullyblank, A., Fayers, P., Blazeby, J., Sprangers, M., Schwartz, C., Visser, M., Oort, F., and Sprangers, M. (2011). Responsiveness of EORTC QLQ-C30, QLQ-CR38 and FACT-C quality of life questionnaires in patients with colorectal cancer. *Health and Quality of Life Outcomes*, 9(1):70.

Uysal-Bozkir, O., Fokkema, T., MacNeil-Vroomen, J. L., van Tilburg, T. G., and de Rooij, S. E. (2017). Translation and Validation of the De Jong Gierveld Loneliness Scale Among Older Migrants Living in the Netherlands. *J Gerontol B Psychol Sci Soc Sci*, 72:109–119.

van Baarsen, B., Snijders, T. A. B., Smit, J. H., and van Duijn, M. A. J. (2001). Lonely but not alone: Emotional isolation and social isolation as two distinct dimensions of loneliness in older people. *Educational and Psychological Measurement*, 61:119–135.

van der Heijden, G., Leffers, P., and Bouter, L. M. (2000). Shoulder disability questionnaire design and responsiveness of a functional status measure. *Journal of clinical*.

Van der Wees, P., Hendriks, E., Van Beers, H., Van Rijn, R., Dekker, J., and De Bie, R. (2012). Validity and responsiveness of the ankle function score after acute ankle injury. *Scand J Med Sci Sports*, 22:170–174.

Van Esch, L., Den Oudsten, B. L., and De Vries, J. (2011). The World Health Organization Quality of Life Instrument-Short Form (WHOQOL-BREF) in women with breast problems. *International Journal of Clinical and Health Psychology*, 11:5–22.

van Verschuer, V. M., Vrijland, W. W., Mares-Engelberts, I., and Klem, T. M. (2015). Reliability and validity of the Dutch-translated Body Image Scale. *Qual Life Res*, 24:1629–1633.

Vaughn, M. J. and Baier, M. E. M. (1999). Reliability and validity of the Relationship Assessment Scale. *American Journal of Family Therapy*, 27:137–147.

Veiga, D., Sabino Neto, M., Ferreira, L., Garcia, E., Veiga Filho, J., Novo, N., and Rocha, J. (2004). Quality of life outcomes after pedicled TRAM flap delayed breast reconstruction. *British Journal of Plastic Surgery*, 57(3):252–257.

Veldhoven, M. V., Meijman, T., Broersen, J., and Fortuin, R. (2002). *Handleiding VBBA*. SKB Vragenlijst Services, Amsterdam, 2nd edition.

Velikova, G., Coens, C., Efficace, F., Greimel, E., Groenvold, M., Johnson, C., Singer, S., van de Poll-Franse, L., Young, T., and Bottomley, A. (2012). Health-Related Quality of Life in EORTC clinical trials - 30 years of progress from methodological developments to making a real impact on oncology practice. *Eur J Cancer Suppl*, 10:141–149.

Villeneuve, L., Trudel, G., Preville, M., Dargis, L., Boyer, R., and Begin, J. (2015). Dyadic Adjustment Scale: a validation study among older French-Canadians living in relationships. *Can J Aging*, 34:26–35.

Wakefield, C. E., Butow, P. N., Aaronson, N. A., Hack, T. F., Hulbert-Williams, N. J., and Jacobsen, P. B. (2015). Patient-reported depression measures in cancer: A meta-review. *Lancet Psychiatry*, 2:635–647.

Walker, P. S., Walls, B., and Murphy, D. K. (2009). Natrelle Saline-Filled Breast Implants: A Prospective 10-Year Study. *Aesthetic Surgery Journal*, 29(1).

Wan, C., Meng, Q., Yang, Z., Tu, X., Feng, C., Tang, X., and Zhang, C. (2008). Validation of the simplified Chinese version of EORTC QLQ-C30 from the measurements of five types of inpatients with cancer. *Annals of oncology : official journal of the European Society for Medical Oncology*, 19(12):2053–60.

Wang, Y., Guo, X., Lau, Y., Chan, K. S., Yin, L., and Chen, J. (2009). Psychometric evaluation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale. *Int J Nurs Stud*, 46:813–823.

Ware, J. and Sherbourne, C. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical care*.

Whistance, R. N., Conroy, T., Chie, W., Costantini, A., Sezer, O., Koller, M., Johnson, C. D., Pilkington, S. A., Arraras, J., Ben-Josef, E., Pullyblank, A. M., Fayers, P., and Blazeby, J. M. (2009). Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. *Eur J Cancer*, 45:3017–3026.

Whistance, R. N., Gilbert, R., Fayers, P., Longman, R. J., Pullyblank, A., Thomas, M., and Blazeby, J. M. (2010). Assessment of body image in patients undergoing surgery for colorectal cancer. *Int J Colorectal Dis*, 25:369–374.

Wilkins, E., Cederna, P., and Lowery, J. (2000). Prospective analysis of psychosocial outcomes in breast reconstruction: one-year postoperative results from the Michigan Breast Reconstruction Outcome Study. *Plastic and*.

Wiltink, J., Hauck, E. W., Phädayanon, M., Weidner, W., and Beutel, M. E. (2003). Validation of the German version of the International Index of Erectile Function (IIEF) in patients with erectile dysfunction, Peyronie's disease and controls. *Int J Impotence Res*, 15:192–197.

Winters, Z. E., Haviland, J., Balta, V., Benson, J., Reece-Smith, A., and Betambeau, N. (2013). Integration of patient-reported outcome measures with key clinical outcomes after immediate latissimus dorsi breast reconstruction and adjuvant treatment. *British Journal of Surgery*, 100(2):240–251.

Wong, C. K. H., Chen, J., Yu, C. L. Y., Sham, M. S., and Lam, C. L. K. (2015). Systematic review recommends the European Organization for Research and Treatment of Cancer colorectal cancer-specific module for measuring quality of life in colorectal cancer patients. *Journal of Clinical Epidemiology*, 68:266–278.

Wood, G. J., Barnes, S. M., and Waring, E. M. (1988). The criterion validity of the Waring Intimacy Questionnaire in a psychiatric inpatient sample. *J Sex Marital Ther*, 14:63–73.

Woody, J. D. and D'Souza, H. J. (1994). The sexual interaction system scale: A new inventory for assessing sexual dysfunction and sexual distress. *J SEX MARITAL THER*, 20:210–228.

Youngblut, J. M., Brooten, D., and Menzies, V. (2006). Psychometric properties of Spanish versions of the FACES II and Dyadic Adjustment Scale. *J Nurs Meas*, 14:181–189.

Yun, Y. H., Bae, S. H., Kang, I. O., Shin, K. H., Lee, R., Kwon, S. I., Park, Y. S., and Lee, E. S. (2004). Cross-cultural application of the Korean version of the European Organization for Research and Treatment of Cancer (EORTC) Breast-Cancer-Specific Quality of Life Questionnaire (EORTC QLQ-BR23). *Supportive Care in Cancer*, 12(6):441–445.

Zainal, N. Z., Shuib, N., Bustam, A. Z., Sabki, Z. A., and Guan, N. C. (2013). Reliability and validity of the malay version of the breast-impact of treatment scale (MVBITS) in breast cancer women undergoing chemotherapy. *Asian Pac J Cancer Preven*, 14:463–468.

Zhang, J., Xie, S., Liu, J., Sun, W., Guo, H., Hu, Y., and Gao, X. (2014). Validation of EORTC IN-PATSAT32 for Chinese patients with gastrointestinal cancer. *Patient Preference Adherence*, 8:1285–1292.

Zhang, L., Dai, Z., Cheng, S., Xie, S., Woo, S. M. L., Luo, Z., Wu, J., Gao, T., Liu, J., Zhang, K., Zhang, J., Jia, X., Miller, A. R., and Wang, C. (2015). Validation of EORTC IN-PATSAT32 for Chinese cancer patients. *Supportive Care Cancer*, 23(9):2721–2730.

Zhong, T., McCarthy, C., Min, S., Zhang, J., Beber, B., Pusic, A. L., and Hofer, S. O. P. (2012). Patient satisfaction and health-related quality of life after autologous tissue breast reconstruction. *Cancer*, 118(6):1701–1709.

Zigmond, A. and Snaith, R. (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica*.