Review Article

Cut Points on 0–10 Numeric Rating Scales for Symptoms Included in the Edmonton Symptom Assessment Scale in Cancer Patients: A Systematic Review

Wendy H. Oldenmenger, RN, PhD, Pleun J. de Raaf, MD, Cora de Klerk, PhD, and Carin C.D. van der Rijt, MD, PhD

Department of Medical Oncology (W.H.O., P.J.d.R., C.C.D.v.d.R.), Erasmus MC Daniel den Hoed Cancer Center and Department of Medical Psychology and Psychotherapy (C.d.K.), Erasmus MC, Rotterdam, The Netherlands

Abstract

Context. To improve the management of cancer-related symptoms, systematic screening is necessary, often performed by using 0–10 numeric rating scales. Cut points are used to determine if scores represent clinically relevant burden.

Objectives. The aim of this systematic review was to explore the evidence on cut points for the symptoms of the Edmonton Symptom Assessment Scale.

Methods. Relevant literature was searched in PubMed, CINAHL®, Embase, and PsycINFO®. We defined a cut point as the lower bound of the scores representing moderate or severe burden.

Results. Eighteen articles were eligible for this review. Cut points were determined using the interference with daily life, another symptom-related method, or a verbal scale. For pain, cut point 5 and, to a lesser extent, cut point 7 were found as the optimal cut points for moderate pain and severe pain, respectively. For moderate tiredness, the best cut point seemed to be cut point 4. For severe tiredness, both cut points 7 and 8 were suggested frequently. A lack of evidence exists for nausea, depression, anxiety, drowsiness, appetite, well-being, and shortness of breath. Few studies suggested a cut point below 4.

Conclusion. For many symptoms, there is no clear evidence as to what the optimal cut points are. In daily clinical practice, a symptom score ≥4 is recommended as a trigger for a more comprehensive symptom assessment. Until there is more evidence on the optimal cut points, we should hold back using a certain cut point in quality indicators and be cautious about strongly recommending a certain cut point in guidelines. J Pain Symptom Manage 2013;45:1083–1093. © *2013 U.S. Cancer Pain Relief Committee. Published by Elsevier Inc. All rights reserved.*

Drs. Oldenmenger and de Raaf contributed equally to the final article.

Address correspondence to: Wendy H. Oldenmenger, RN, PhD, Department of Medical Oncology, Erasmus MC Daniel den Hoed Cancer Center, P.O. Box 5201, 3008 AE Rotterdam, The Netherlands. E-mail: w.h.oldenmenger@erasmusmc.nl

Accepted for publication: June 3, 2012.

0885-3924/$ - see front matter
http://dx.doi.org/10.1016/j.jpainsymman.2012.06.007

Key Words
Neoplasms, diagnosis, questionnaires, self-report, signs and symptoms, pain measurement

Introduction

Cancer patients suffer from many physical and psychological symptoms, negatively affecting quality of life and daily activities.[1] To improve the management of these cancer-related symptoms, it is necessary to screen for these symptoms systematically. Many screening instruments measure the intensity of symptoms on a 0−10 numeric rating scale (NRS), in which 0 means "no suffering" and 10 means "unbearable suffering."[2−4]

Before being able to interpret the results of these measurements, it is important to determine the clinical meaning of the scores given on the 0−10 NRS for the various symptoms. NRS scores have been categorized as none, mild, moderate, and severe[5] or as representing clinically relevant burden or not.[6] When studies categorize NRS scores as none, mild, moderate, and severe, they report two cut points: one cut point for the boundary between mild and moderate burden and another cut point for the boundary between moderate and severe burden. In case articles report one cut point for clinically relevant burden, they describe it as a clinically significant cut point, an optimal single cut point, or a cut point for significant burden. The cut point for clinically relevant burden is considered to be equivalent to the cut point between mild and moderate burden.[6]

Cut points are frequently used in research and in daily clinical practice. For example, cut points are frequently used in inclusion criteria and in the definition of end points of clinical trials. Furthermore, they are used in quality indicators to measure the quality of care. In addition, cut points are recommended in guidelines as a starting point for the initiation of treatment and for the evaluation of the treatment[7] (e.g., National Comprehensive Cancer Network [NCCN] guidelines[8,9]).

For various symptoms, cut points on the NRS have been proposed, especially for pain and fatigue.[6,10] However, there is heterogeneity in the cut points being recommended. For example, the NCCN proposed a cut point ≥4 for fatigue,[8] whereas an expert group of the European Association for Palliative Care suggested the cut point ≥5.[11] Despite this lack of uniformity, as mentioned before, cut points are advised in guidelines and quality indicators, which has consequences for the treatment of the symptoms and the assessment of the quality of care.

Therefore, it is important to define evidence-based cut points that are proven to distinguish between NRS scores with clinically relevant burden or not. The aim of this review was to explore the evidence on cut points for the respective symptoms of the Edmonton Symptom Assessment Scale (ESAS) in cancer patients and whether it is possible to recommend an optimal cut point per symptom or to recommend one cut point for all symptoms of the ESAS.

Methods

We conducted a systematic review on cut points for the symptoms of the ESAS: pain, tiredness, nausea, depression, anxiety, drowsiness, appetite, well-being, and shortness of breath. We searched for studies that measured these symptoms on an NRS or an equivalent instrument, that is, a visual analogue scale (VAS),[12] the Brief Pain Inventory (BPI),[13] the Brief Fatigue Inventory (BFI),[14] the ESAS,[2] the M. D. Anderson Symptom Inventory,[3] the Fatigue Symptom Inventory,[15] or the Symptom Monitor.[4]

Relevant literature was searched in PubMed, using the search strategy: "Neoplasms"[Mesh] AND (cut OR cut-off OR "cut off" OR cutpoint* OR "*symptom* severity") AND (*symptom* OR VAS OR "Visual Analogue Scale" OR "Visual Analog Scale" OR "Visual Scale" OR NRS OR "Numeric Rating Scale" OR BPI OR "Brief Pain Inventory" OR BFI OR "Brief Fatigue Inventory" OR ESAS OR "Edmonton Symptom Assessment Scale" OR FSI OR "Fatigue Symptom Inventory" OR "M. D. Anderson Symptom Inventory" OR "symptom monitor"). We used this strategy for the nine symptoms of the ESAS and also included the synonyms used by the revised ESAS:[16] pain, tiredness (including fatigue and lack of energy), nausea, depression (including feeling sad), anxiety (including nervousness and feeling nervous), drowsiness (including sleepiness and feeling sleepy), appetite

(including loss of appetite, lack of appetite, poor appetite, and anorexia), well-being, and shortness of breath (including dyspnea and breathlessness).

The search was limited to English articles published until July 2011 and to original articles. Studies were included in this review if they were performed with cancer patients, measured one or more symptoms of the ESAS on a 0–10 NRS, and performed statistical tests to determine the optimal cut point. To identify supplementary studies, we studied the reference lists of the selected articles and searched for cross-references. We conducted an additional search in CINAHL®, Embase, and PsycINFO® using the same search strategy.

Articles were reviewed for eligibility independently by two authors (W. H. O. and P. J. d. R.). The results were summarized, and conclusions were independently drafted by these two authors. If the reviewers disagreed about a conclusion, the assumptions leading to the conclusion were discussed until consensus was reached.

Per study, we reported patient characteristics (e.g., disease stage, antitumor treatment), inclusion criteria with respect to symptom scores, methodological characteristics, and quality criteria (e.g., prospective or retrospective design, if a primary or secondary analysis was performed, sample size), specification of the type of symptom intensity asked for (e.g., usual, worst), the method used to determine the cut point, and the number of optional cut points explored. We studied which NRS scores were tested as possible cut points for the various symptoms and which scores were finally selected as the most optimal cut points.

In some articles, the reported cut point reflected the lower bound of a category,[14] whereas in other studies the reported cut point represented the upper bound of a category.[5] We chose to report the lower bound of a category as the cut point. For example, when we report cut points 5 and 7 (CP57), we mean that mild burden is defined with scores 1–4, moderate burden is defined with scores 5–6, and severe burden is defined with scores 7–10.

Results

We found 1524 articles through the original search, of which 14 were relevant. The additional search produced four supplementary articles. In total, we found 18 relevant articles that determined cut points for symptoms covered by the ESAS (Fig. 1). The main characteristics and the quality aspects of these articles are summarized in Tables 1 and 2. The majority of the studies included patients with various stages of cancer; five studies only included patients with advanced cancer.[5,10,17–19] In seven articles, patients were only eligible when they met a certain inclusion criterion on symptom burden.[5,6,19–23] Four articles determined cut points for multiple symptoms.[6,10,21,24] Six studies were primarily designed to calculate cut points.[6,10,23,25–27] All studies measured the symptom on a 0–10 NRS, and no studies used a VAS. For pain, all studies defined the type of symptom intensity (e.g., worst or usual) they asked for. Four of 10 studies on fatigue[10,18,21,26] and all studies on the other symptoms did not define the type of symptom intensity (Table 2). The recall time of the question about symptom intensity varied between "right now"[10] to "last week."[5] Seven studies did not describe the recall time.[17,19–21,23,24,26]

Methods Used to Determine Cut Points

Fifteen studies determined the cut point for a certain symptom using the interference of that symptom with daily life as reference.[5,6,14,17–23,25,27–30] Three studies used another symptom-related questionnaire as reference,[6,24,26] and one study assessed the cut point using the severity of that particular symptom on a verbal scale (none, mild, moderate, severe) as reference[10] (Table 2).

Twelve studies described two cut points: a cut point for mild/moderate burden and a cut point for moderate/severe burden.[5,10,14,17,19–21,25,27–30] The other studies assessed one cut point for clinically relevant burden[6,18,22–24,26] (Table 2).

Studies used regression models or a receiver operating characteristic (ROC) curve to determine the optimal cut point(s). The studies using regression models (multivariate analysis of variance [MANOVA] or general linear model [GLM]) studied multiple cut points or combinations of cut points to categorize NRS scores by symptom severity (clinically relevant/not clinically relevant or mild/moderate/severe). The cut point or combination of cut points that best differentiated the symptom severity categories with respect to the level the

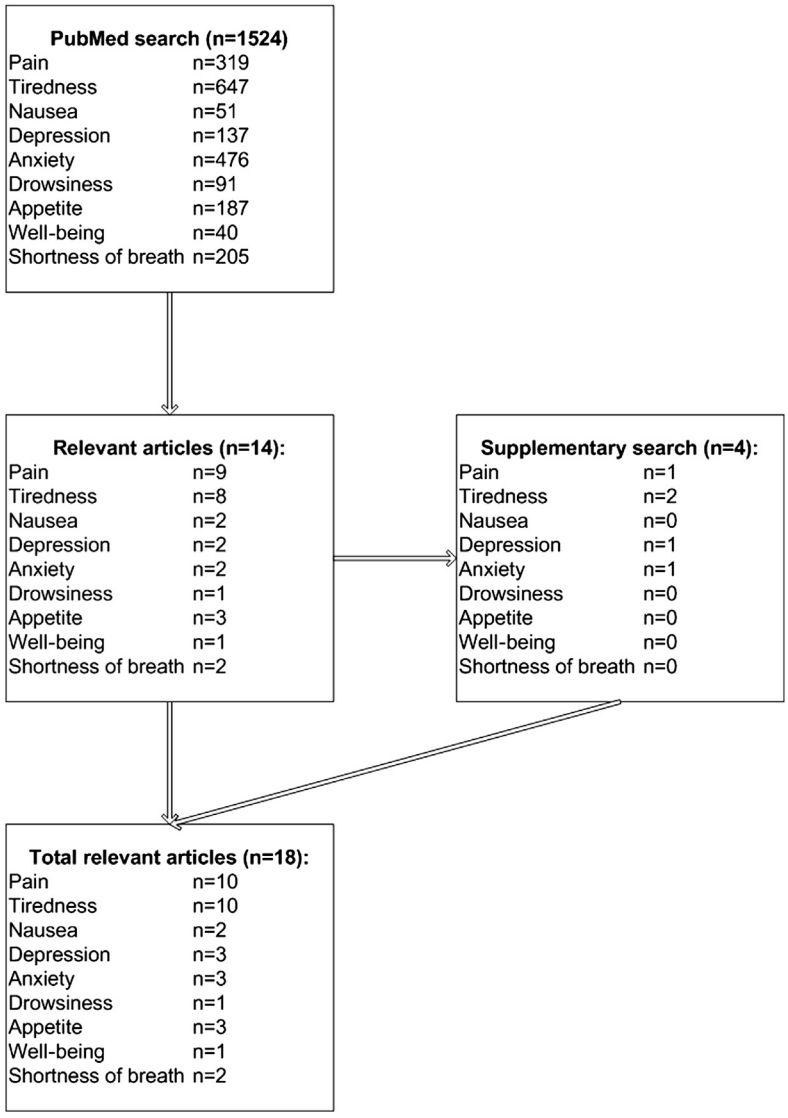

Fig. 1. Flowchart of selection process.

symptom interfered with daily life, as measured with the reference questionnaire, was considered to be the optimal cut point. The number of possible options explored for a single cut point varied from one[25] to seven.[17]

The studies using an ROC curve predefined per patient if a symptom was present, using a reference questionnaire. Thereafter, for each possible cut point on the NRS, the sensitivity and specificity were calculated. The sensitivity was defined as the proportion of the patients suffering from that particular symptom (as predefined using the reference questionnaire) with an NRS score on that possible cut point or higher. The specificity was defined as the proportion of the patients not suffering from that particular symptom (according to the reference questionnaire) with an NRS score below that possible cut point. The optimal cut point is that with the optimal ratio between sensitivity and specificity.

Optimal Cut Points per Symptom

Pain. Ten studies assessed cut points on an NRS for pain.[5,6,10,17,19,21–23,28] Pain was asked as present pain,[10,17] average pain,[17,19,21,23] or worst pain.[5,6,17,19,20,22,28] Seven studies used a MANOVA[5,17,19,20,22,23,28] and one study a GLM,[21] both with the interference items of the BPI as reference. In the seven studies

Table 1
Overview of the Characteristics of the Included Studies

First Author (Year)	Symptoms	Study Design	Type of Analysis	Patient Population			
				N	Disease Stage (%)	Treatment (%)	Symptom Burden in Inclusion Criteria
Serlin[5] (1995)	Pain	Retrospective	Secondary	1897	Metastatic (100)	?	Worst pain >0
Mendoza[14] (1999)	Fatigue	Prospective	Secondary	305	Advanced (85) Early (15)	CT (100)	—
Okuyama[18] (2001)	Fatigue	Prospective	Secondary	157	Advanced lung cancer	—	—
Hwang[25] (2002)	Fatigue	Prospective	Primary	180	NED (7) Localized (5) Locally advanced (21) Advanced (67)	CT (17) RT (10) CRT (4) HT (22) None (47)	—
Okuyama[30] (2003)	Fatigue	Prospective	Secondary	252	Recurrence (36) Metastatic (50)	Surgery (2) CT (22) RT (2)	—
Paul[19] (2005)	Pain	Retrospective	Secondary	160	Metastatic (100)	CT (46) HT (32) RT (18) BT (3) None (12)	Average pain ≥2.5
Temel[26] (2006)	Fatigue	Prospective	Primary	574	?	?	—
Vignaroli[24] (2006)	Depression, anxiety	Retrospective	Secondary	216	?	?	?
Chang[27] (2007)	Fatigue	Prospective	Primary	150	I (10) II (9) III (23) IV (58)	Surgery (15) CT (66) RT (27) None (21)	—
Li[17] (2007)	Pain	Retrospective	Secondary	199	Metastatic (100)	RT (100)	—
Butt[6] (2008)	Pain, fatigue, appetite loss	Prospective	Primary	148	Local (27) Regional (24) Metastatic (42) N/A (7)	?	NRS ≥4 on ≥1/4 symptoms
Given[21] (2008)	Pain, fatigue, nausea, depression, anxiety, poor appetite, dyspnea	Retrospective	Secondary	588	Early (15) Late (85)	CT (100)	NRS ≥2 on ≥1/16 symptoms
Kalyadina[28] (2008)	Pain	Prospective	Secondary	148	I or II (7) III (52) IV (35) Recurrent disease (6)	?	—
Valeberg[23] (2008)	Pain	Prospective	Primary	210	Metastatic (41)	?	Average pain ≥0
Utne[22] (2009)	Pain	Retrospective	Secondary	225	Metastatic (70)	?	Opioid treatment
Mendoza[29] (2010)	Fatigue	Prospective	Secondary	206	?	?	—
Selby[10] (2010)	Pain, tiredness, nausea, depression, anxiety, drowsiness, loss of appetite, well-being, shortness of breath	Prospective	Primary	400	Advanced	?	—
Ferreira[20] (2011)	Pain	Prospective	Secondary	143	Metastatic (66)	None (100)	Chronic cancer-related pain

CT = chemotherapy; NED = no evidence of disease; RT = radiation therapy; CRT = chemoradiation therapy; HT = hormonal therapy; BT = biotherapy; NRS = numeric rating scale; N/A = not applicable (i.e., patients with hematologic malignancies).

Table 2
Cut Points per Symptom

First Author (Year)	*N*	Source of NRS	Question	Recall Time	Reference: Type	Reference: Questionnaire	Reference: Cut Point	Statistical Method	NRS Score: 1	2	3	4	5	6	7	8	9	10
Pain																		
Li[17] (2007)	199	BPI	Pain (present)[a]	Now	Interference	BPI-I	—	MANOVA			Mo[b]	Mo[b,c]	Mo[b,c]	Mo[b,c]	S[b,c]	S[b,c]	S[b,c]	S[c]
Selby[10] (2010)	400	ESAS	Pain (present)	Now	Verbal Scale	VRS	Mo or S	ROC curve					Mo	Mo	S	S	S	S
Paul[19] (2005)	160	BPI	Pain (average)	?	Interference	BPI-I	—	MANOVA				[b]	Mo[b]	Mo[b,c]	Mo[c]	S	S	S
Li[17] (2007)	199	BPI	Pain (average)[a]	?	Interference	BPI-I	—	MANOVA			[b]	[b,c]	Mo[b,c]	Mo[b,c]	S[b,c]	S[b,c]	S[b,c]	S[c]
Given[21] (2008)	588	Seven-item[d]	Pain (average)	?	Interference	BPI-I[e]	—	GLM	Mo	Mo	Mo	Mo	S	S	S	S	S	S
Valeberg[23] (2008)	210	BPI	Pain (average)	?	Interference	BPI-I	—	MANOVA				[b]	CR[b]	CR[b]	CR[b]	CR[b]	CR[b]	CR
Serlin[5] (1995)	1897	BPI	Pain (worst)	Seven days	Interference	BPI-I	—	MANOVA				[b]	Mo[b]	Mo	S[c]	S[c]	S	S
Paul[19] (2005)	160	BPI	Pain (worst)	?	Interference	BPI-I	—	MANOVA				[b]	Mo[b]	Mo[b,c]	Mo[c]	S[c]	S	S
Li[17] (2007)	199	BPI	Pain (worst)[a]	?	Interference	BPI-I	—	MANOVA			[b]	[b,c]	Mo[b,c]	Mo[b,c]	S[b,c]	S[b,c]	S[b,c]	S[c]
Butt[6] (2008)	148	Four-item[f]	Pain (worst)	Three days	Interference	BPI-I	BPI-I ≥45	ROC curve				CR	CR	CR	CR	CR	CR	CR
Kalyadina[28] (2008)	148	BPI	Pain (worst)	24 hours	Interference	BPI-I	—	MANOVA				[b]	Mo[b]	Mo	S[c]	S[c]	S	S
Utne[22] (2009)	225	BPI	Pain (worst)	24 hours	Interference	BPI-I	—	MANOVA				[b]	CR[b]	CR[b]	CR[b]	CR	CR	CR
Ferreira[20] (2011)	143	BPI	Pain (worst)	?	Interference	BPI-I	—	MANOVA				[b]	Mo[b]	Mo[b,c]	Mo[c]	S[c]	S	S
Tiredness																		
Selby[10] (2010)	400	ESAS	Tiredness	Now	Verbal Scale	VRS	Mo or S	ROC curve						Mo	Mo	S	S	S
Okuyama[18] (2001)	157	One-item	Fatigue	24 hours	Interference	BFI-I	≥1 item ≥1	ROC curve			CR	CR	CR	CR	CR	CR	CR	CR
Temel[26] (2006)	574	One-item	Fatigue	?	Symptom Scale	FACT-F	FACT-F <30	ROC curve				CR	CR	CR	CR	CR	CR	CR
Given[21] (2008)	588	Seven-item[d]	Fatigue	?	Interference	BPI-I[e]	—	GLM	Mo	Mo	Mo	Mo	S	S	S	S	S	S
Hwang[25] (2002)	180	BFI	Fatigue (usual)[g]	24 hours	Interference	BFI-I	—	MANOVA			Mo[b]	Mo	Mo	Mo[c]	S[c]	S	S	S
Chang[27] (2007)	150	BFI	Fatigue (usual)[g]	24 hours	Interference	BFI-I	—	MANOVA			[b]	Mo[b]	Mo	Mo	Mo[c]	S[c]	S	S
Mendoza[14] (1999)	305	BFI	Fatigue (worst)	24 hours	Interference	BFI-I	—	MANOVA				Mo[b]	Mo[b]	Mo	S[c]	S[c]	S	S
Hwang[25] (2002)	180	BFI	Fatigue (worst)[g]	24 hours	Interference	BFI-I	—	MANOVA				[b]	Mo[b]	Mo	S[c]	S	S	S
Okuyama[30] (2003)	252	BFI	Fatigue (worst)	24 hours	Interference	BFI-I	—	MANOVA				Mo[b]	Mo[b]	Mo	Mo[c]	S[c]	S	S
Chang[27] (2007)	150	BFI	Fatigue (worst)[g]	24 hours	Interference	BFI-I	—	MANOVA			[b]	Mo[b]	Mo	Mo	Mo[c]	S[c]	S	S
Butt[6] (2008)	148	Four-item[f]	Fatigue (worst)	Three days	Symptom Scale	FACT-F	FACT-F ≤42	ROC curve					CR	CR	CR	CR	CR	CR
Mendoza[29] (2010)	206	BFI	Fatigue (worst)	24 hours	Interference	BFI-I	—	MANOVA				Mo[b]	Mo[b]	Mo	S[c]	S[c]	S	S
Nausea																		
Given[21] (2008)	588	Seven-item[d]	Nausea/vomiting	?	Interference	BPI-I[e]	—	GLM				Mo	Mo	Mo	S	S	S	S
Selby[10] (2010)	400	ESAS	Nausea	Now	Verbal Scale	VRS	Mo or S	ROC curve				Mo	S	S	S	S	S	S
Depression																		
Vignaroli[24] (2006)	216	ESAS	Depression	?	Symptom Scale	HADS	HADS-D ≥8	ROC curve	CR	CR	CR	CR	CR	CR	CR	CR	CR	CR
Given[21] (2008)	588	Seven-item[d]	Depression	?	Interference	BPI-I[e]	—	GLM	Mo	Mo	S	S	S	S	S	S	S	S
Selby[10] (2010)	400	ESAS	Depression	Now	Verbal Scale	VRS	Mo or S	ROC curve				Mo	Mo	Mo	S	S	S	S
Anxiety																		
Vignaroli[24] (2006)	216	ESAS	Anxiety	?	Symptom Scale	HADS	HADS-A ≥8	ROC curve	CR	CR	CR	CR	CR	CR	CR	CR	CR	CR
Given[21] (2008)	588	Seven-item[d]	Anxiety	?	Interference	BPI-I[e]	—	GLM				Mo	Mo	S	S	S	S	S
Selby[10] (2010)	400	ESAS	Anxiety	Now	Verbal Scale	VRS	Mo or S	ROC curve					Mo	Mo	S	S	S	S

Drowsiness																
Selby[10] (2010)	400	ESAS	Drowsiness	Now	Verbal Scale	VRS	Mo or S	ROC curve			Mo	Mo	S	S	S	S
Appetite																
Butt[6] (2008)	148	Four-item[f]	Appetite loss	Three days	Severity	FAACT-A	≤21	ROC curve				CR	CR	CR	CR	CR
Given[21] (2008)	588	Seven-item[d]	Poor appetite	?	Interference	BPI-I[e]	—	GLM		Mo	Mo	Mo	S	S	S	S
Selby[10] (2010)	400	ESAS	Loss of appetite	Now	Verbal Scale	VRS	Mo or S	ROC curve			Mo	Mo	S	S	S	S
Well-being																
Selby[10] (2010)	400	ESAS	Well-being	Now	Verbal Scale	VRS	Mo or S	ROC curve				Mo	S	S	S	S
Shortness of breath																
Given[21] (2008)	588	Seven-item[d]	Dyspnea	?	Interference	BPI-I[e]	—	GLM		Mo	Mo	Mo	S	S	S	S
Selby[10] (2010)	400	ESAS	Shortness of breath	Now	Verbal Scale	VRS	Mo or S	ROC curve			Mo	S	S	S	S	S

NRS = numeric rating scale; BPI = Brief Pain Inventory; BPI-I = BPI interference items; MANOVA = multivariate analysis of variance; Mo = moderate; S = severe; ESAS = Edmonton Symptom Assessment Scale; ROC = receiver operating characteristic; GLM = general linear model; CR = clinically relevant; BFI-I = Brief Fatigue Inventory interference items; FACT-F = Functional Assessment of Cancer Therapy-Fatigue subscale; BFI = Brief Fatigue Inventory; HADS = Hospital Anxiety and Depression Scale; HADS-D = Hospital Anxiety and Depression Scale-Depression subscale; HADS-A = Hospital Anxiety and Depression Scale-Anxiety subscale; FAACT-A = Functional Assessment of Anorexia/Cachexia Treatment-Anorexia subscale; VRS = verbal rating scale.

[a]Authors recommended to ask for worst pain, which better predicted interference scores than average pain or present pain.

[b]Optimal lower bound cut points tested to define a symptom as of moderate burden (only for studies using MANOVA).

[c]Optimal lower bound cut points tested to define a symptom as of severe burden (only for studies using MANOVA).

[d]The authors questioned for the intensity of pain, fatigue, nausea, depression, anxiety, poor appetite, and dyspnea simultaneously, each on a 0–10 NRS.

[e]The authors used four interference items from the Brief Pain Inventory: enjoyment of life, social relationships, general daily activities, and emotions.

[f]The authors questioned for the intensity of pain, fatigue, distress, and appetite loss simultaneously, each on a 0–10 NRS.

[g]Authors recommended to ask for usual fatigue, which better predicted interference scores than worst fatigue.

that used MANOVA statistics,[5,17,19,20,22,23,28] multiple models were tested to determine cut points, with Li et al.[17] testing most extensively. Two other studies used an ROC curve with the interference items of the BPI[6] or a verbal scale[10] as a reference. Thirteen cut points were calculated for moderate pain or clinically relevant pain (range CP2–CP5), with CP5 most frequently being recommended as the optimal cut point.[5,10,17,19,20,22,23,28] Ten cut points were suggested for severe pain (range CP5–CP8), with CP7 presented as the optimal cut point most frequently[5,10,17,28] (Table 2).

We found no clear differences in cut points between studies asking for different types of pain intensity (present, average, or worst pain) (Table 2).

Tiredness. Ten studies published cut points on an NRS for tiredness.[5,6,14,18,21,25–27,29,30] The question on tiredness was formulated as worst fatigue,[6,14,25,27,29,30] usual fatigue,[25,27] fatigue,[18,21,26] or tiredness.[10] Six studies used a MANOVA[6,14,25,27,29,30] and one study a GLM[21] with the interference items of the BFI as reference. In all studies that determined the cut point using a MANOVA,[6,14,25,27,29,30] conclusions on the optimal cut point were based on the analyses of two possible cut points only (Table 2). The other studies used an ROC curve with the Functional Assessment of Cancer Therapy-Fatigue subscale,[6,26] the interference items of the BFI,[18] or a verbal scale[10] as a reference. Twelve cut points were proposed for moderate tiredness or clinically relevant tiredness (range CP2–CP6),[6,10,14,18,21,25–27,29,30] with CP4 being found as the optimal cut point most frequently.[14,26,27,29,30] Nine cut points were proposed for severe tiredness (range CP5–CP8),[10,14,21,25,27,29,30] with CP7[14,25,29] and CP8[10,27,30] being recommended as the optimal cut points most frequently. We could not investigate if there were differences in cut points between studies asking for different types of tiredness intensity (i.e., usual or worst fatigue) because only two studies asked for usual fatigue.

Nausea. Two studies assessed cut points on an NRS for nausea, using an ROC curve[10] or a GLM.[21] The study that used the severity of nausea expressed on a verbal scale as reference reported CP4 for moderate nausea and CP5

for severe nausea.[10] The other study, which determined the cut point using the interference of nausea in daily life, found CP4 and CP7 to be the optimal cut points for moderate and severe nausea, respectively[21] (Table 2).

Depression. Three studies published cut points on an NRS for depression.[10,21,24] Two studies used an ROC curve with the depression subscale of the Hospital Anxiety and Depression Scale [24] or a verbal scale[10] as a reference. One other study calculated the cut point using the interference of depression with daily life in a GLM.[21] Cut points for moderate/clinically relevant depression were CP2[21,24] or CP4.[10] Severe depression was represented with CP4[21] and CP7[10] (Table 2).

Anxiety. Three studies assessed cut points on an NRS for anxiety.[10,21,24] Two studies calculated the sensitivity per cut point using the anxiety subscale of the Hospital Anxiety and Depression Scale (clinically relevant CP2)[24] or using a verbal scale (moderate CP5),[10] whereas the third study determined the optimal cut point using the interference of anxiety with daily life in a GLM (moderate CP4).[21] Severe anxiety was indicated with CP6[21] or CP7[10] (Table 2).

Drowsiness. One study determined a cut point on an NRS for drowsiness based on the severity of drowsiness as measured with a verbal scale. Moderate drowsiness was reflected best by CP5, and severe drowsiness was indicated by CP7[10] (Table 2).

Appetite. Three studies reported cut points on an NRS for appetite.[6,10,21] The question on appetite was formulated as "appetite loss,"[6] "loss of appetite,"[10] or "poor appetite."[21] Two studies calculated the sensitivity per cut point using the Functional Assessment of Anorexia/Cachexia Therapy (clinically relevant CP6)[6] or using a verbal scale (moderate CP5),[10] whereas the other study calculated the optimal cut point using the interference of poor appetite with daily life in a GLM (moderate CP4).[21] Two studies that determined a cut point for severe appetite loss reported CP7[10,21] (Table 2).

Well-Being. Only one study determined a cut point on an NRS for well-being using an ROC curve with a verbal scale as reference. Moderate impairment of well-being corresponded best with an NRS score of 6 and severe impairment of well-being corresponded best with an NRS score of 7 or higher[10] (Table 2).

Shortness of Breath. Two studies assessed cut points on an NRS for shortness of breath.[10,21] The study that used the severity of shortness of breath expressed on a verbal scale as reference reported CP4 for moderate shortness of breath and CP6 for severe shortness of breath.[10] The other study, which determined the cut point using the interference of dyspnea in daily life in a GLM, found CP3 and CP7 to be the optimal cut points for moderate and severe dyspnea, respectively[21] (Table 2).

Discussion

Based on this review, there is not sufficient evidence for recommending the same cut point for all symptoms of the ESAS questionnaire. The level of evidence of the optimal cut point differs per symptom. The most evidence exists for cut points for pain and tiredness. Concerning pain, there is consensus for CP5 as the cut point for moderate pain and, to a lesser extent, for CP7 as the cut point for severe pain. This implies that mild pain is reflected by NRS scores 1–4, moderate pain by NRS scores 5–6, and severe pain by NRS scores 7–10. For moderate tiredness, CP4 seems to be the most appropriate cut point. For severe tiredness, the evidence is ambiguous; both CP7 and CP8 are suggested frequently. There is conflicting evidence for cut points on the symptoms depression, anxiety, and appetite. For these symptoms, we found three or four studies with inconsistent results per symptom. A lack of evidence exists for cut points for nausea, shortness of breath, and well-being. Possible cut points for these symptoms were only studied once or twice.

Determination of the optimal cut point depends on the purpose of the test in a specific context, as well as the costs of misses and false alarms. In research and quality assessment of care, one usually aims for optimal accuracy when using a screening measure such as the NRS. In daily practice, clinicians who screen

for cancer-related symptoms in their patients generally aim to minimize the amount of false-negative test results. In this context, symptom screening with a brief, easy-to-administer screening tool is usually followed by a more comprehensive symptom history to identify patients who actually experience clinically relevant burden. The present review showed that few existing studies have recommended a cut point below 4. Therefore, we argue that using CP4 as a cut point for further screening of symptoms will result in identifying most patients with clinically relevant burden.

The interpretation of the results of this review is hampered by the limited comparability of the studies included because the patients in these studies varied greatly in tumor type, disease stage, and treatment received. The comparability of the studies also is limited by the heterogeneity in the symptom assessment questionnaires used, which, for instance, differed in the wording of the probe question and recall time. In addition, the studies varied in the type of symptom intensity asked for (e.g., worst, usual, or current) or they did not specify this (Table 2). Besides this, the studies differed in the reference questionnaire used, the cut point used on the reference questionnaire, the number of possible cut points explored, and the method used to determine cut points (Tables 1 and 2).

The interpretation of the results of this study also is complicated by differences in the quality of the included studies. For example, several studies explored only one optional cut point for a distinction between two categories of pain or fatigue and we cannot rule out that the potential cut point above[5,25,27,28] or below[14,25,29,30] the studied cut point has better characteristics. Unfortunately, there are no quality assessment tools for observational studies that are sufficiently validated.[31] Moreover, there are no quality assessment tools available that contain criteria on prerequisites for a reliable assessment of optimal cut points. Therefore, we decided to describe several aspects of study quality in the Results section and in Tables 1 and 2 instead of performing a quantitative quality assessment with a non-validated tool.

In this review, we identified three methods to determine cut points: based on daily interference, based on another symptom-related questionnaire, and based on verbal rating of symptom severity (Table 3). Every method had its advantages and disadvantages, and it is not clear which method is most suitable to determine the optimal cut point. In the future, thinking aloud studies[32] have to be performed to investigate whether patients rate their symptom intensity on the NRS mainly on the basis of perceived disabilities caused by that particular symptom (daily interference) or by word descriptors of the symptom intensity (mild/moderate/severe). More insight into the cognitive processes underlying the scores on the questionnaires will help to determine whether cut points should be determined based on the interference of a certain symptom with daily life or based on the subjective severity on a verbal scale. Also, we must investigate whether various approaches to

Table 3
Advantages and Disadvantages of the Various Methods to Determine Cut Points

Determination of Cut Points Using	Advantages	Disadvantages
Daily interference	- Gives insight in symptom-related impairments in daily activities - No cut point needed on reference questionnaire	- Difficult for patients to discriminate which symptom causes impairments in daily life in case of suffering from multiple symptoms - Does not take patients' opinion on acceptability of certain symptom scores into account
Other symptom-related questionnaire	- Sensitivity and specificity of cut points can be calculated - Facilitates comparison with other questionnaires	- Assumption needed on cut point on reference lists because of lack of gold standards - Does not take patients' opinion on acceptability of certain symptom scores into account - Does not give insight in symptom-related impairments in daily activities
Symptom intensity on verbal rating scale (none, mild, moderate, severe)	- Professionals' prejudices not needed for determination of cut points - Fits with the subjective nature of symptoms	- Does not give insight in symptom-related impairments in daily activities

determine cut points in the same population result in different cut points.

Most importantly, in future research, cut points should be reported unambiguously. Fourteen of the 18 included articles described the cut points as boundaries of the created categories. The four other articles reported the ranges of the categories created,[18,27,29,30] without mentioning the actual cut point. Six articles, all pain literature, described upper boundaries of a category,[5,17,19,20,22,23] whereas eight articles reported the lower boundaries of a category.[6,10,14,21,24–26,28] Uniformity in reporting cut points is important to avoid confusion. For example, the NCCN guideline "Adult Cancer Pain" categorized mild pain as 1–3,[9] referring to Serlin et al.[5] In the original study, however, Serlin et al.[5] categorized mild pain as 1–4.

Little is known about the validity of cut points in different situations, for example, in the different stages of cancer or for inpatients and outpatients. Besides this, it is possible that the type of symptom intensity asked for (e.g., worst, usual, or current) will affect the cut points. Moreover, cut points could differ depending on the pathophysiology of the symptom (e.g., nociceptive pain and neuropathic pain; physical fatigue and mental fatigue). Furthermore, it is unclear whether cut points are stable over time. It is conceivable that cut points change after a long duration of suffering of a certain symptom. Prospective studies are needed to determine the factors that influence the cut points.

In conclusion, cut points are frequently used in clinical practice and scientific research. In this review, we found some evidence on cut points for pain (moderate pain CP5 and severe pain CP7) and fatigue (moderate fatigue CP4). Until there is more evidence on the optimal cut points, we should hold back in using a certain cut point in quality indicators and be cautious about strongly recommending a cut point in guidelines. In daily clinical practice, symptom scores ≥ 4 should trigger a more comprehensive symptom assessment to properly identify the patients with clinically relevant symptom burden.

Disclosures and Acknowledgments

No funding was received for this study. The authors declare no conflicts of interest.

References

1. van den Beuken-van Everdingen MH, de Rijke JM, Kessels AG, et al. Quality of life and non-pain symptoms in patients with cancer. J Pain Symptom Manage 2009;38:216–233.

2. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care 1991;7:6–9.

3. Cleeland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the M. D. Anderson Symptom Inventory. Cancer 2000; 89:1634–1646.

4. Hoekstra J, Bindels PJ, van Duijn NP, Schade E. The symptom monitor. A diary for monitoring physical symptoms for cancer patients in palliative care: feasibility, reliability and compliance. J Pain Symptom Manage 2004;27:24–35.

5. Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. Pain 1995;61:277–284.

6. Butt Z, Wagner LI, Beaumont JL, et al. Use of a single-item screening tool to detect clinically significant fatigue, pain, distress, and anorexia in ambulatory cancer practice. J Pain Symptom Manage 2008;35:20–30.

7. Anderson KO. Role of cutpoints: why grade pain intensity? Pain 2005;113:5–6.

8. Berger AM, Pickar Abernethy A, Atkinson A, et al. Cancer-related fatigue. NCCN Clinical Practice Guidelines in Oncology. Fort Washington, PA: NCCN, 2011.

9. Swarm R, Abernethy AP, Anghelescu DL, et al. Adult cancer pain. NCCN Clinical Practice Guidelines in Oncology. Fort Washington, PA: NCCN, 2011.

10. Selby D, Cascella A, Gardiner K, et al. A single set of numerical cutpoints to define moderate and severe symptoms for the Edmonton Symptom Assessment System. J Pain Symptom Manage 2010;39: 241–249.

11. Radbruch L, Strasser F, Elsner F, et al. Fatigue in palliative care patients—an EAPC approach. Palliat Med 2008;22:13–32.

12. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986;27:117–126.

13. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994;23:129–138.

14. Mendoza TR, Wang XS, Cleeland CS, et al. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. Cancer 1999;85:1186–1196.

15. Hann DM, Jacobsen PB, Azzarello LM, et al. Measurement of fatigue in cancer patients:

development and validation of the Fatigue Symptom Inventory. Qual Life Res 1998;7:301–310.

16. Watanabe SM, Nekolaichuk C, Beaumont C, et al. A multicenter study comparing two numerical versions of the Edmonton Symptom Assessment System in palliative care patients. J Pain Symptom Manage 2011;41:456–468.

17. Li KK, Harris K, Hadi S, Chow E. What should be the optimal cut points for mild, moderate, and severe pain? J Palliat Med 2007;10:1338–1346.

18. Okuyama T, Tanaka K, Akechi T, et al. Fatigue in ambulatory patients with advanced lung cancer: prevalence, correlated factors, and screening. J Pain Symptom Manage 2001;22:554–564.

19. Paul SM, Zelman DC, Smith M, Miaskowski C. Categorizing the severity of cancer pain: further exploration of the establishment of cutpoints. Pain 2005;113:37–44.

20. Ferreira KA, Teixeira MJ, Mendoza TR, Cleeland CS. Validation of Brief Pain Inventory to Brazilian patients with pain. Support Care Cancer 2011;19:505–511.

21. Given B, Given CW, Sikorskii A, et al. Establishing mild, moderate, and severe scores for cancer-related symptoms: how consistent and clinically meaningful are interference-based severity cutpoints? J Pain Symptom Manage 2008;35:126–135.

22. Utne I, Miaskowski C, Bjordal K, et al. Differences in the use of pain coping strategies between oncology inpatients with mild vs. moderate to severe pain. J Pain Symptom Manage 2009;38:717–726.

23. Valeberg BT, Miaskowski C, Hanestad BR, et al. Demographic, clinical, and pain characteristics are associated with average pain severity groups in a sample of oncology outpatients. J Pain 2008;9: 873–882.

24. Vignaroli E, Pace EA, Willey J, et al. The Edmonton Symptom Assessment System as a screening tool for depression and anxiety. J Palliat Med 2006; 9:296–303.

25. Hwang SS, Chang VT, Cogswell J, Kasimis BS. Clinical relevance of fatigue levels in cancer patients at a Veterans Administration Medical Center. Cancer 2002;94:2481–2489.

26. Temel JS, Pirl WF, Recklitis CJ, Cashavelly B, Lynch TJ. Feasibility and validity of a one-item fatigue screen in a thoracic oncology clinic. J Thorac Oncol 2006;1:454–459.

27. Chang YJ, Lee JS, Lee CG, et al. Assessment of clinical relevant fatigue level in cancer. Support Care Cancer 2007;15:891–896.

28. Kalyadina SA, Ionova TI, Ivanova MO, et al. Russian Brief Pain Inventory: validation and application in cancer pain. J Pain Symptom Manage 2008;35: 95–102.

29. Mendoza TR, Laudico AV, Wang XS, et al. Assessment of fatigue in cancer patients and community dwellers: validation study of the Filipino version of the Brief Fatigue Inventory. Oncology 2010;79:112–117.

30. Okuyama T, Wang XS, Akechi T, et al. Validation study of the Japanese version of the Brief Fatigue Inventory. J Pain Symptom Manage 2003;25:106–117.

31. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 2007; 36:666–676.

32. Watanabe S, Nekolaichuk C, Beaumont C, Mawani A. The Edmonton Symptom Assessment System—what do patients think? Support Care Cancer 2009;17:675–683.