Title Page

Title;

Defining the cut-off point of clinically significant postoperative fatigue in three common fatigue scales.

The study was conducted at;

Department of Anaesthesia, Telemark Hospital Trust, Skien, Norway.

Authors;

Torkjell Nøstdahl; MD, Senior Consultant
- Department of Anaesthesia, Telemark Hospital, Skien, Norway

Tomm Bernklev; Research Director, PhD, Professor
- Research & Development, Vestfold Hospital Trust, Tønsberg, Norway
and
- Institute of Clinical Medicine, University of Oslo, Norway

Olav Fredheim; MD, PhD, Professor
- Pain and Palliation Research Group, Department of Circulation and Medical Imaging, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway
and
- National Competence Centre for Complex Symptom Disorders, Department of Pain and Complex Disorders, St. Olav University Hospital, Trondheim, Norway
and
- Centre of Palliative Medicine, Surgical Division, Akershus University Hospital, Lørenskog, Norway

Johanna S Paddison; MD, PhD, Department of Medicine, Wellington Hospital, New Zealand

Johan Raeder; MD, PhD, Consultant/Professor.
- Institute of Clinical Medicine, University of Oslo, Norway
and
- Department of Anaesthesiology, Oslo University Hospital, Norway

Corresponding author;

Torkjell Nøstdahl, Department of Anaesthesia, Telemark Hospital, Skien, Norway.
Tel: +47-35003500. E-mail: nost@sthf.no

Keywords;

Cut-off point, Fatigue, Postoperative, Quality-of-Life, Surgery

Word count;

3917 words

Source of support/Conflict of Interest;

Telemark Hospital Trust, Department of Research and Development has been the sole funding source, providing a partial buy-out from consultant practice for Torkjell Nøstdahl.

None of the authors have any conflict of interests to declare.

Defining the cut-off point of clinically significant postoperative fatigue in three common fatigue scales.

Nøstdahl T, Bernklev T, Fredheim OM, Paddison JS, Ræder J.

Introduction

Fatigue is a subjective experience, often defined as a persistent tiredness or weakness, being physical, mental or both. It is common in the general population [1,2] but is also present in a wide range of diseases. It may influence physical, behavioural, cognitive and social functioning, interfere with daily activities and restrict recreational activities [3,4].
Postoperative fatigue (POF) is an often underestimated problem after surgery [5,6]. It may impede patient recovery, with an impact similar to that of pain [7,8] and is often reported as one of the most distressing symptoms by patients [9,10]. It is most prominent during the first postoperative days, but may last several weeks [5,9,11]. Although POF is often proportional to the invasiveness of a surgical procedure [6,11-13], in some cases, surgical procedures with extensive tissue damage (e.g. hip or knee arthroplasty) are associated with less POF than would be expected [14]. The aetiology of POF is poorly understood but appears to be multifactorial [15,16].

Being a subjective experience fatigue is a difficult construct to define and measure [10,17]. Although some question the existence of more than one dimension [18], fatigue symptomatology is often considered to fall into two dimensions: physical and mental [17]. In addition, how it impacts daily activities may be of interest, especially in recovery after day-surgery. The dimensions may be influenced differently depending on the origin, intervention or population. It is therefore argued that multidimensional assessment measures ideally should be applied in descriptive research or for identification of underlying mechanisms [10]. A multitude of fatigue scales are available [4,10,19]. Furthermore, some health related quality of life (HRQoL) instruments have multi-item subscales that assess fatigue as part of broad assessments of HRQoL; such measures include the EORTC QLQ-C30 [20] (developed for cancer research) where 3 of 30 items assess fatigue and the generic MOS 36-item Short Form (SF-36) [21] where 4 of 36 items assess energy/fatigue.

One problem in POF research has been the use of non-validated measures [6], and also the use of single-item or one-dimensional measures [5]. Additionally, use of different measures across studies makes systematic reviews and pooling of evidence difficult. Further, the numeric scores generated have rarely been related to the clinical significance these experiences have for patients.

This paper aims to define operational cut-off points for clinically significant fatigue in a postoperative setting in three different fatigue instruments. These instruments are the Postoperative Fatigue Scale (PO-FS) [22], the Christensen Fatigue Scale (ChrFS) [23], and finally the Chalder Fatigue Questionnaire (CFQ) [17] (see below). Cut-off points will allow for analyses of the clinical impact of POF as a complement to statistical comparisons of mean scores.

Methods

Study design and study population.

The data in the present study have been collected as part of two previously published fatigue-related studies from our research group; a validation study in a mixed day surgery population [22], and a randomised controlled trial (RCT) on fatigue after laparoscopic cholecystectomy [24]. For the present analyses, data from the two studies were combined. See Table 1 for basic patient characteristics and Table 2 for details on the categories of surgery performed.

Table 1. Basic characteristics for enrolled patients

Age; years	Mean (SD)	44 (14.7)
Gender	Male; n (%)	182 (43.1)
	Female; n (%)	240 (56.9)
ASA[a]	1/2/3; n	167/244/11
BMI[b]; kg/m^2	Mean (SD)	26.5 (4.2)
Duration Surgery; minutes	Mean (SD)	40 (24.3)
Duration Anaesthesia; minutes	Mean (SD)	63 (28.3)

[a]**ASA**; American Society of Anesthesiologists physical status classification system (see Table A1)

[b]**BMI**; Body Mass Index

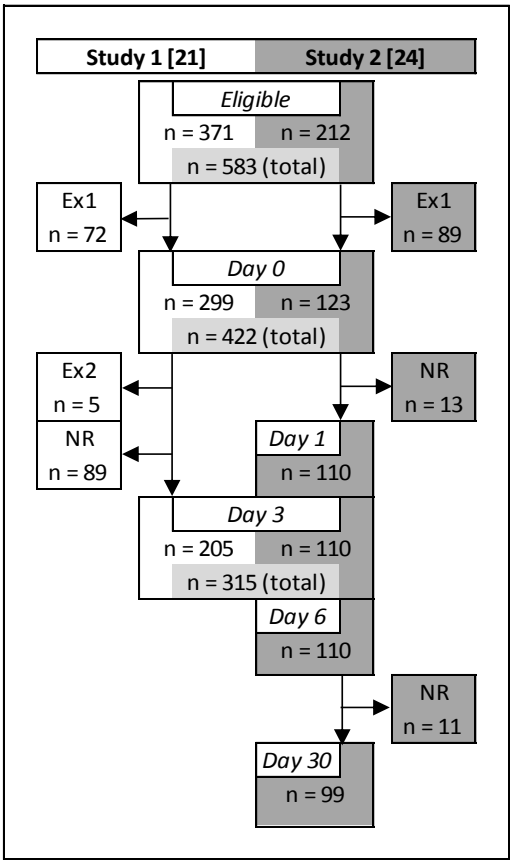
Table 2. Categories of surgery performed

		N	(%)
ENT[a]	Tonsillectomy	41	(9.7)
	Nose surgery	5	(1.2)
Orthopaedic	Joint endoscopy	73	(17.3)
	Other surgery; minor	31	(7.3)
Plastic surg.	Plastic surgery; medium	22	(5.2)
	Plastic surgery; minor	25	(5.9)
General surg.	Laparoscopy	128	(30.3)
	Breast surgery	3	(0.7)
	Transurethral surgery; minor	14	(3.3)
	Other surgery; minor	55	(13.0)
Gynaecology	Laparoscopy	2	(0.5)
	Other surgery; minor	23	(5.5)
	Total	**422**	

[a]**ENT;** Ear, Nose and Throat

In both studies fatigue was measured using the PO-FS, ChrFS, and the CFQ. Data were recorded pre-operatively on the day of surgery (day 0), and on postoperative days 1, 3, 6 and 30. Recall interval was the last two days, except for day 1 which referred to the last 24h. Figure 1 summarises enrolment in both studies, lost participants and the number of forms available at each time point.

Figure 1. Combined flowchart for the validation study (Study 1) and the RCT (Study 2)



Ex1: Excluded due to inclusion/exclusion criteria,
Ex2: Excluded due to protocol violation after enrolment,
NR: Number of forms not returned at each time point.

Assessment tools

Postoperative Fatigue Scale (PO-FS).

In 2006 Paddison et al presented a measure specifically developed for POF research; the 31-item Identity-Consequences Fatigue Scale (ICFS) [25]. Even though this instrument has been validated and displays good psychometric qualities, it may be perceived as too inclusive and cumbersome for clinical research, particularly if POF is not the primary outcome. Recently, a secondary validation of the ICFS revealed significant item redundancy, and hence an abridged version was derived and validated in a large, mixed surgical population [22]. The abridged version, the PO-FS, consists of 10 items capturing 3 dimensions: performance of daily activities, fatigue and vitality. The scale has no mental/cognitive dimension, as these items in the original four-dimension/31-item ICFS showed lack of sufficient change over time, and were thus eliminated during generation of the 10-item version. Scores on this scale range from 0 to 100.

Christensen Fatigue Scale (ChrFS).

The ChrFS [23] is a single, numeric rating scale (1-10) with four verbal anchors. See Figure A1 (Appendix). Due to its simplicity, it is user-friendly and it may thus fit well into any questionnaire. It has been used in several clinical studies on POF. However, as the scale is one-dimensional and primarily related to physical fatigue it will not be able to distinguish potential different aspects of POF. Scores on this scale range from 1 to 10.

Chalder Fatigue Questionnaire (CFQ).

The CFQ [17] is a widely used measure in chronic fatigue research; it has good psychometric qualities and has been validated in several languages. It was primarily developed for research on chronic fatigue syndrome (CFS) and general fatigue in community settings. It has gained widespread use also outside of CFS research [10]. It consists of 11 items, representing two dimensions; mental fatigue (four items) and physical fatigue (seven items). However, unpublished data from our validation study [22] revealed that similar to the ICFS, in postoperative patients the mental dimension items displayed minimal change over time compared to the physical dimension items. Hence in a surgical recovery setting the CFQ may principally be regarded as a measure of physical fatigue. Scoring may be bimodal or continuous; in our analyses we have applied continuous scoring. Scores on this scale range from 0 to 33 with a calculated [26] cut-off score at ≥18.

For all scales higher scores indicate greater fatigue severity.

To our knowledge neither ChrFS nor CFQ have been validated in a mixed surgical population.

Ethics

Informed written consent was obtained from all patients before inclusion in the studies. Approval for both studies was obtained concurrently from the Regional Committee for Medical Research Ethics; Ref 2009/2171. The RCT was registered in ClinicalTrials.gov; ID: NTC 01125982.

Analysis

The aim of this study was to define at which level the reported fatigue was considered clinically significant by the participants. To accomplish this, one key question was added in each form; “Given your current description of fatigue; would you say it has been of considerable significance to you?”; “Yes/No”. By this dichotomisation, the responses served as an anchor; defining whether clinically significant fatigue was present or not. Against this anchor we analysed each scale’s ability to identify clinically significant fatigue, by performing receiver-operating characteristics (ROC) analyses. ROC analysis is used for diagnostic tests with a dichotomised outcome, where a ROC curve plots Sensitivity and (1-Specificity) against a series of cut-off points. We estimated the area under the ROC curve (AUC) as a measure of the accuracy of the instrument. We calculated the optimal cut-off point between Sensitivity and Specificity by the *Youden Index J* (J = max [Sensitivity + Specificity]-1) and the *Point closest to the (0, 1)* (C^* = √[(1 – sensitivity)^2 + (1 – specificity)^2]) [27,28].

To generate a pragmatic overall cut-off point, for each scale, we calculated a weighted average of the different cut-off points for each day on which fatigue assessments were made. We calculated Likelihood Ratios and Predictive Values related to this cut-off point. Predictive Values were calculated assuming that the positive responses to the key question represented the prevalence of fatigue at the corresponding day.

Scale reliability was measured by Cronbach's alpha (α) coefficient. Correlations between scales were analysed with Pearson’s r.

Scale responsiveness was assessed by calculating the standardised response mean (SRM). Values of .20, .50 and .80 are deemed to represent small, moderate and large responsiveness, respectively [29,30].

The McNemar test was used to analyse the association between proportions.

Missing data were handled by replacing the missing value with the mean score of the subscale.

We performed statistical analyses using MedCalc for Windows®, version 17.9, IBM SPSS® Version 23 and MS Excel 2010®.

Results

A total of 583 patients were considered eligible for participation in the two studies, and 422 were enrolled. The validation study had pre-operative baseline data from day 0 (n=299) and from postoperative day 3 (n=205). The RCT on fatigue after laparoscopic cholecystectomy had equivalent data from day 0 (n=123) and days 1, 3 and 6 (n=110) and day 30 (n=99). See Figure 1 for details.

Returned forms.
The returned forms included all three questionnaires. There were few missing data, with a total of 7 items missing in the PO-FS and 17 items missing in the CFQ for the entire period. Four out of 422 participants did not fill in the ChrFS on day 0, while 5 out of 315 were missing on day 3. ChrFS was completed in all returned forms on days 1, 6 and 30.
Being young and/or male was associated with a lower return rate. Mean age of patients not returning the forms was 34.5 years, while those returning the forms had a mean age of 46.7 years ($p<0.001$). While 36.8 % of men did not return the forms, 17.1 % of women did not return the forms ($p<0.001$). Apart from this no clinical or demographic differences were found, when compared to the responding group.

Demographic analyses.
At baseline there were no differences in fatigue between gender, age, BMI or ASA status. However, at day 3, women reported significantly more fatigue, and there was a negative correlation between age and fatigue; i.e. that younger persons tended to report more fatigue. There was no difference related to BMI or ASA status at any time point.

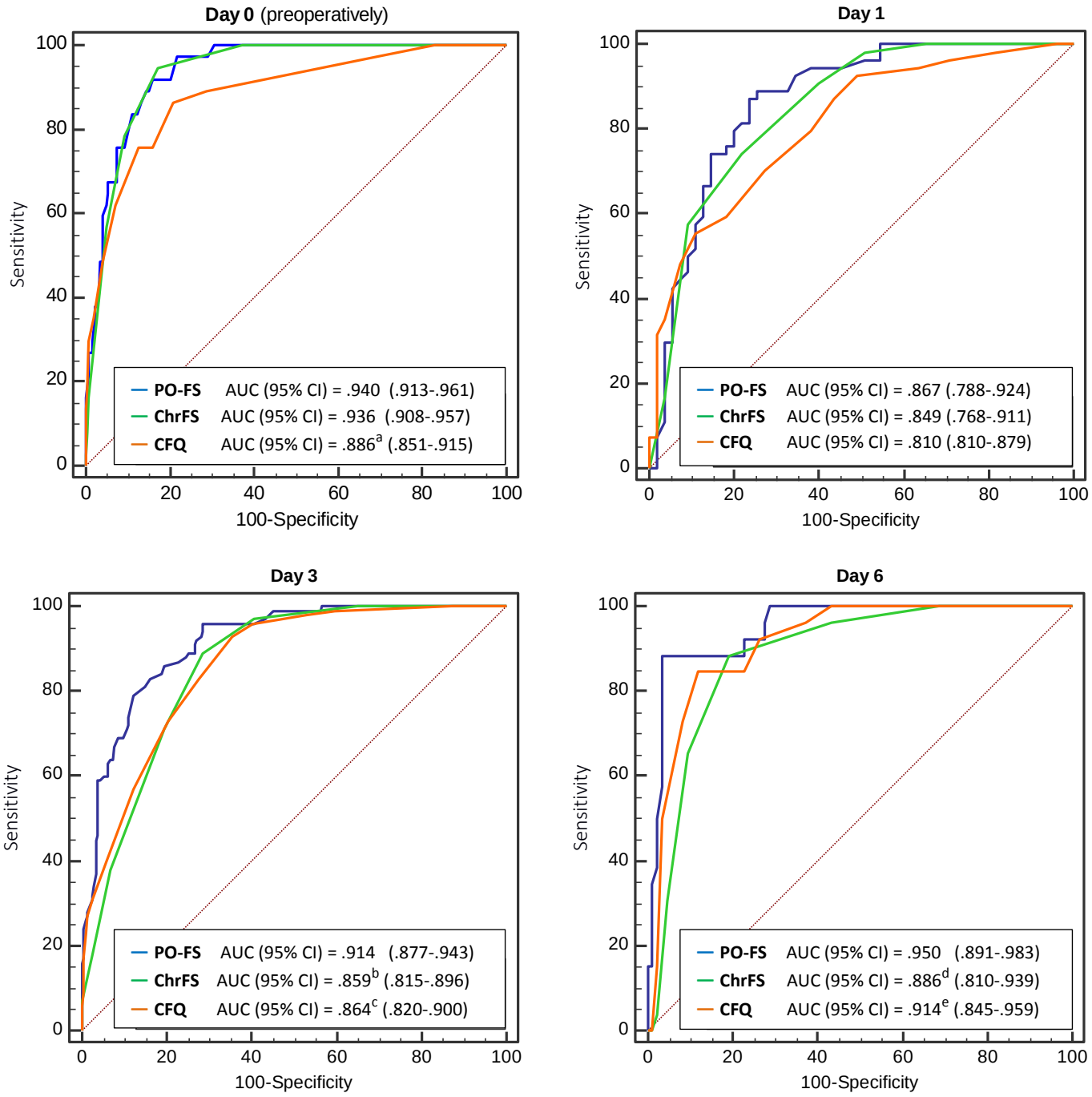
Scale metrics.
Responsiveness of the scales, expressed as Standardized Response Mean (SRM) comparing mean fatigue on Day 0 and Day 1 was: PO-FS: 1.15, ChrFS: 1.35, and CFQ: 0.88. Likewise for the subscales of PO-FS and CFQ: $PO\text{-}FS^{fatigue}$: 0.92, $PO\text{-}FS^{vitality}$: 0.82, $PO\text{-}FS^{daily\ activities}$: 1.17, $CFQ^{physical}$: 1.05 and CFQ^{mental}: 0.17.
Cronbach's alpha for PO-FS on day 0, 1, 3, 6 and 30 was .861, .911, .912, .893 and .868, and for CFQ it was .851, .895, .902, .861 and .834, respectively. Correlation between the (sub)scales at day 3 is shown in Table 3.

Table 3. Correlations between ChrFS and the subscales in PO-FS and CFQ at day 3

	ChrFS	$CFQ^{physical}$	CFQ^{mental}	$PO\text{-}FS^{fatigue}$	$PO\text{-}FS^{vitality}$	$PO\text{-}FS^{daily}$
ChrFS	1	.609 (p<0.001)	.126 (p=0.191)	.643 (p<0.001)	.714 (p<0.001)	.554 (p<0.001)
$CFQ^{physical}$	.609 (p<0.001)	1	.336 (p<0.001)	.670 (p<0.001)	.615 (p<0.001)	.674 (p<0.001)
CFQ^{mental}	.126 (p=0.191)	.336 (p<0.001)	1	.154 (p=0.108)	.178 (p=0.063)	.300 (p=0.002)
$PO\text{-}FS^{fatigue}$	.643 (p<0.001)	.670 (p<0.001)	.154 (p=0.108)	1	.706 (p<0.001)	.540 (p<0.001)
$PO\text{-}FS^{vitality}$	.714 (p<0.001)	.615 (p<0.001)	.178 (p=0.063)	.706 (p<0.001)	1	.514 (p<0.001)
$PO\text{-}FS^{daily}$	.554 (p<0.001)	.674 (p<0.001)	.300 (p=0.002)	.540 (p<0.001)	.514 (p<0.001)	1

ROC curve analyses showed that AUC ranged between 0.810 and 0.950 for all scales, with PO-FS consistently having a larger AUC at every data point compared to the other scales. See Figure 2 for details. ROC analysis for day 30 was not possible, because at that time no patients reported fatigue to be of considerable significance.

Figure 2. Area Under Curve (AUC) for "Clinically Significant Fatigue" measured Pre-operatively (Day 0) and Day 1, 3 and 6 with the Postoperative Fatigue Scale (PO-FS), the Christensen Fatigue Scale (ChrFS) and the Chalder Fatigue Questionnaire (CFQ).



Difference between AUC: [a]PO-FS vs CFQ ($p=0.042/p^{BON}=0.127$); [b]PO-FS vs ChrFS ($p=0.002/p^{BON}=0.006$); [c]PO-FS vs CFQ ($p=0.006/p^{BON}=0.017$); [d]PO-FS vs ChrFS ($p=0.039/p^{BON}=0.118$); [e]PO-FS vs CFQ ($p=0.021/p^{BON}=0.062$). No other differences were significant. **p^{BON}**: adjusted p-value using Bonferroni correction.

The calculated optimal cut-off points for clinically relevant fatigue for PO-FS were 43.1, 58.6, 53.6, and 51.7 on days 0, 1, 3 and 6, respectively. Corresponding values for ChrFS and CFQ were 5, 8, 6 and 5, and 13, 20, 16 and 16, respectively (Table 4). Calculations with either Youden Index *J*, or C* yielded the same results.

Table 4. Sensitivity and Specificity with 95% Confidence Interval at the respective optimal Cut-off point on days 0, 1, 3 and 6.

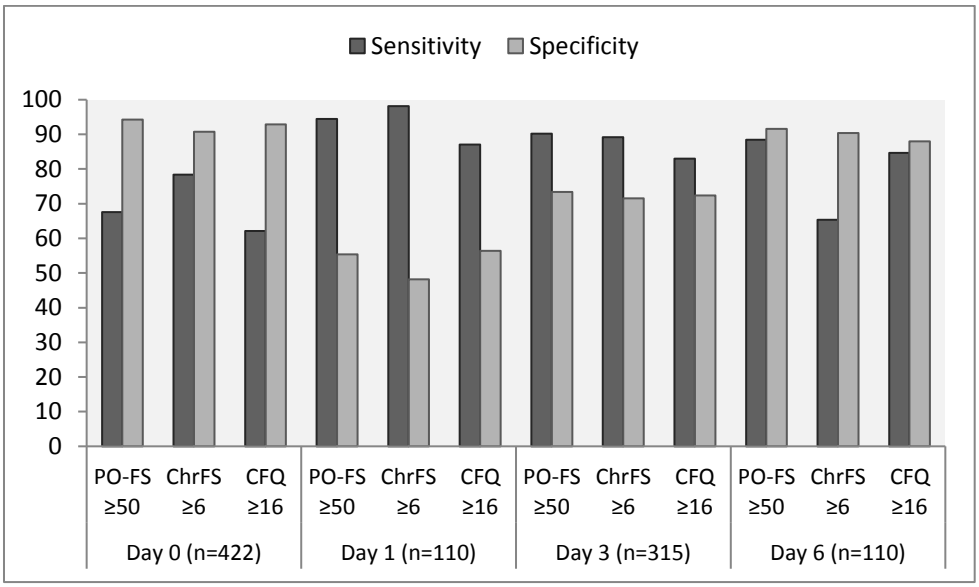
		Optimal Cut-off point	**Sensitivity** (95% CI)	**Specificity** (95% CI)
Day 0 (n=422)	PO-FS	≥ 43.1	91.9 (78.1 - 98.3)	83.7 (79.6 - 87.3)
	ChrFS	≥ 5	94.6 (81.8 - 99.3)	82.9 (78.7 - 86.5)
	CFQ	≥ 13	86.5 (71.2 - 95.5)	79.2 (74.8 - 83.2)
Day 1 (n=110)	PO-FS	≥ 58.6	87.0 (75.1 - 94.6)	76.8 (63.6 - 87.0)
	ChrFS	≥ 8	74.1 (60.3 - 85.0)	76.8 (63.6 - 87.0)
	CFQ	≥ 20	55,6 (41,4 – 69,1)	89,1 (77,8 – 95,9)
Day 3 (n=315)	PO-FS	≥ 53.6	82.4 (73.6 - 89.2)	84.1 (78.3 - 88.8)
	ChrFS	≥ 6	89.2 (81.5 - 94.5)	71.5 (64.8 - 77.5)
	CFQ	≥ 16	83.0 (74.2 - 89.8)	72.3 (65.7 - 78.3)
Day 6 (n=110)	PO-FS	≥ 51.7	88.5 (69.8 - 97.6)	96.4 (89.8 - 99.2)
	ChrFS	≥ 5	88.5 (69.8 - 97.6)	80.7 (70.6 - 88.6)
	CFQ	≥ 16	84.6 (65.1 - 95.6)	88.0 (79.0 - 94.1)

The results were weighted according to the number of participants on the respective days. The mean level of clinically significant POF from days 0, 1, 3 and 6 pooled, was 49.3, 5.7 and 15.1 for PO-FS, ChrFS and CFQ respectively. Accordingly, the operational cut-off points are ≥50 (0-100), ≥6 (1-10) and ≥16 (0-33) respectively.

Due to the observed differences in fatigue related to age and gender on day 3, we also analysed whether age or gender influenced the cut-points. There were no significant difference in the cut-off points between age groups or gender, and the few differences observed were small and inconsistent in direction between the three scales. Further, there were some inconsistencies between the Youden Index *J* and C*.

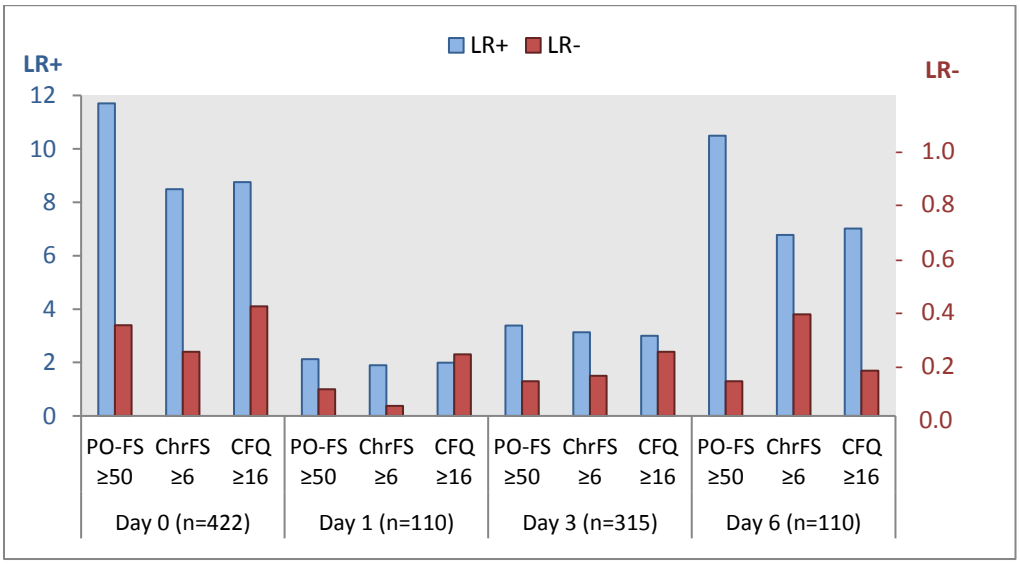
When we applied the averaged cut-off point for each scale, sensitivity ranged from 62.2 to 78.4 % pre-operatively, from 87.0 to 98.2 % on day 1, from 83.0 to 90.2 % on day 3 and from 65.4 to 88.5 % on day 6. Specificity ranged from 90.8 to 94.2 % pre-operatively, from 48.2 to 56.4 % on day 1, from 71.5 to 73.4 % on day 3, and from 88.0 to 91.6 % on day 6. (Figure 3)

Figure 3. Sensitivity and Specificity at the averaged Cut-off points for PO-FS (≥ 50/100), ChrFS (≥6/10) and CFQ (≥16/33).



The Positive Likelihood Ratios (LR+) were low in all scales on day 1 (1.90-2.12), slightly higher on day 3 (3.00-3.39), further increased on day 6 (6.78-10.49), and was highest in all scales pre-operatively/day 0 (8.49-11.70). The Negative Likelihood Ratio (LR-) varied between the scales in a more inconsistent pattern. (Figure 4)

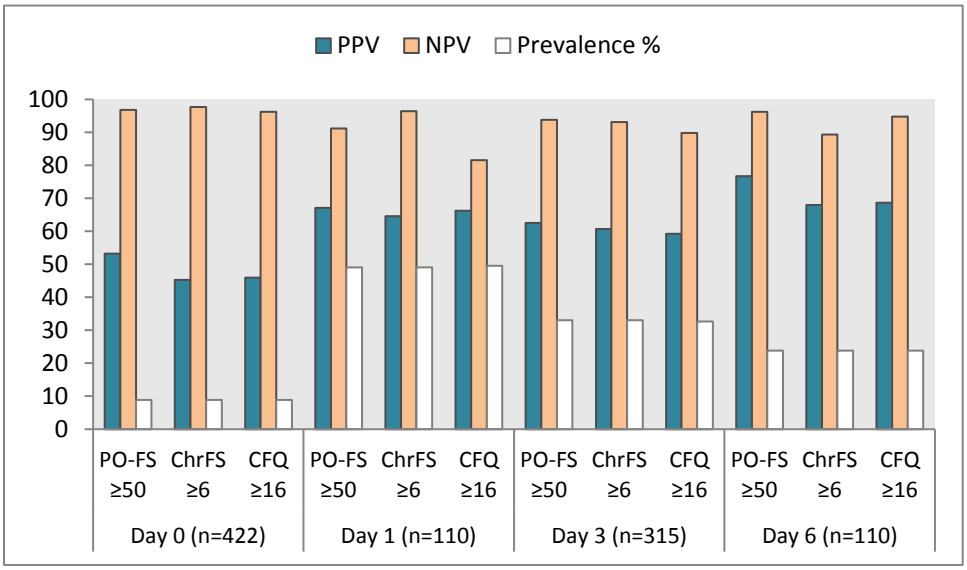
Figure 4. Positive and Negative Likelihood Ratio at the averaged Cut-off points for PO-FS (≥ 50/100), ChrFS (≥6/10) and CFQ (≥16/33).



LR+: Positive Likelihood Ratio **LR-**: Negative Likelihood Ratio

The Negative Predictive Values (NPV) were consistently high (81.6-97.7%) for all scales, across data collection time points. The Positive Predictive Values (PPV) were markedly lower; PO-FS ranged from 53.2% to 76.7%, ChrFS from 45.3% to 68.0%, and CFQ from 46.0 to 68.7%. (Figure 5)

Figure 5. Positive and Negative Predictive values, and proportion of patients self-reporting fatigue being of considerable significance, at the averaged cut-off points for PO-FS (≥ 50/100), ChrFS (≥6/10) and CFQ (≥16/33).



PPV: Positive Predictive Value **NPV**: Negative Predictive Value
Prevalence: Proportion of patients self-reporting fatigue being of considerable significance

Details, with 95% Confidence Intervals on Sensitivity, Specificity, Likelihood Ratios and Predictive Values at the averaged cut-off point for each scale are reported in Table A2 (Appendix).

Figure 6 illustrates the relation between mean fatigue and the optimal cut-off points to the averaged cut-off level in PO-FS, ChrFS and CFQ during the peri-operative period. Details on mean fatigue values in the respective scales are reported in Table A3 (Appendix)

Table 5 shows the percentage of patients with clinically significant fatigue (i.e. above the cut-off point). See also Figure A2 (Appendix)

Figure 6. Relation of mean fatigue and optimal cut-off points to the averaged cut-off point level measured with PO-FS, ChrFS and CFQ.

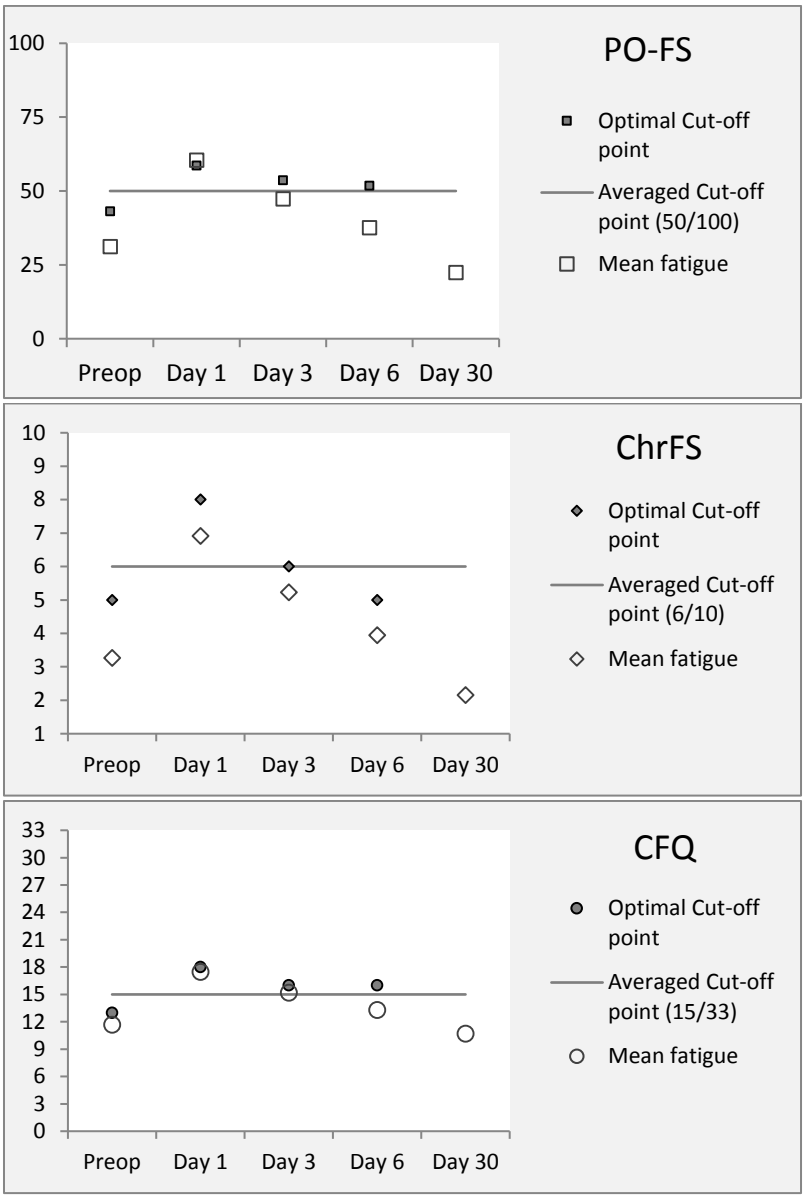


Table 5. Proportion of patients with clinically significant fatigue (i.e. above the cut-off point) when measured with PO-FS, ChrFS and CFQ.

	Day 0 (n=422)	**Day 1** (n=110)	**Day 3** (n=315)	**Day 6** (n=110)	**Day 30** (n=99)
PO-FS ≥50	11.1	69.1 p<0.001	47.3 p<0.001	27.3 p=0.003	4.0 p=0.146
ChrFS ≥6	15.3	74.5 p<0.001	48.7 p<0.001	22.7 p=0.585	3.0 p=0.003
CFQ ≥16	12.0	65.1 p<0.001	46.2 p<0.001	29.1 p=0.024	4.1 p=0.004

p= significance of differences between Day 0 and the respective postoperative days for each scale. There were no differences between the scales within the same day; p>0.05, except PO-FS vs. ChrFS on Day 0: p=0.021. All associations were analysed with McNemar test.

Discussion

It is important that patient reported outcome measures are meaningful and interpretable. Until now, studies on POF have reported fatigue scores and development of fatigue over time without relating these to the clinical relevance for patients. To date authors have tended to use their discretion to define what is 'significant', or 'mild', 'moderate' or 'severe' fatigue; both single studies and reviews on fatigue may present vague and imprecise descriptions of the problem, as exemplified in a review by de Oliveira et al [31]. With a defined cut-off point for clinically significant POF it is possible to analyse the clinical implications of fatigue other than by merely reporting mean fatigue values, and arbitrary definitions of cut-off points and correspondingly ambiguous analyses can be avoided. As an example; in their study on POF Schroeder & Hill [32] set a cut-off point for the ChrFS to ≥4, which, according to our findings probably included a large proportion of patients without clinically significant fatigue in their analyses.

Through the present study we have identified cut-off points for self-reported fatigue which are subjectively experienced to be "of considerable significance" in three scales commonly used in POF research. These cut-off points make it possible to indicate whether statistically significant findings of increased fatigue are also of clinical relevance and thus valuable in outcome evaluation. There is no "gold standard definition" of clinically significant postoperative fatigue. Our definition relies on the patients' own perception of what they regard as "of considerable significance" on the day. Ideally, an optimal cut-off point should be unaltered pre- and postoperatively. The cut-off points vary slightly and in correspondence with the level of fatigue on the respective day. We decided to define an averaged level for the cut-off point using each scale as a weighted average of all days tested. This was a pragmatic trade-off to present a suitable cut-off point for use throughout the peri-operative period. As Figure 3 illustrates, this affects sensitivity and specificity levels differently on different peri-operative days. Generally, with increased fatigue prevalence sensitivity increases, and specificity decreases. Similarly, likelihood ratios and predictive values varied on different days (Figure 4 and 5). We calculated the equivalent values also for other possible joint cut-off points, however this resulted in less optimal results in several parameters on one or more days; leaving the reported averaged cut-off points as the optimal choice for the whole period.

It may be argued that variations in the cut-off points would not allow for the defining of an averaged cut-off point. However, in our opinion it can be reasoned that this pragmatic approach, with a cut-off point that incorporates several days better will reflect the entire peri-operative experience, and allow for fatigue assessment at any day, irrespective of level of fatigue. Further, the concept of postoperative fatigue in a surgical population should not be viewed independently from pre-operative status; we think it is relevant to consider also the pre-operative level of fatigue, as the disease in itself may contribute to the patient's experience of fatigue. Thus, pre-operative level of fatigue was also included in the calculations. This is also the rationale behind choosing a weighted average, as the patient's subjective report of fatigue being "of significance" should count equally at any time-point. However, as we also have presented the data from each assessment point, future studies with clinically significant fatigue as the primary outcome can choose to apply the cut-off value from the most appropriate time point in order to achieve an optimal balance between sensitivity and sensitivity.

Although sensitivity, specificity, likelihood ratios and predictive values were reasonably similar in all scales on the different days, the PO-FS consistently performed better; with larger AUC and higher LR+

and PPV compared to the other scales. Nevertheless, based on the findings here, all scales performed adequately and may favourably be used in fatigue research. The AUC values for all scales were "excellent" to "outstanding"; reflecting strength of discrimination [33] as an expression of the scales' accuracy and ability to identify patients with or without fatigue.

All scales showed high responsiveness; SRM: 0.88-1.35. This was also reflected in the analyses of the subscales in PO-FS and CFQ, except in the CFQ^{mental} subscale, with a low value: 0.17. ChrFS had the highest responsiveness, and was highly correlated to the physical subscales in both the PO-FS and CFQ. This may indicate that POF is primarily related to the physical aspect of fatigue. Thus, surgery does not seem to significantly influence mental fatigue. This is consistent with previous findings related to the $ICFS^{concentration}$ items [22,34], which lead to the exclusion of these items in the abbreviated PO-FS. There may however, be reasons to apply a fatigue scale that assesses mental fatigue in POF research, if also other aspects of fatigue is of interest, depending on the intervention or population.

When individuals experience a change in health status they may change their expectation and evaluation of quality of life. This dependency between a patient's expectations and scoring in HRQoL instruments is called response shift, initially described by Calman [35] The theory and appraisal of response shift has since been further developed [36,37]. Within quality of life research fatigue seem to be especially susceptible to this phenomenon [38,39]. The observed temporal variations in the threshold for fatigue of 'considerable significance' may at least partially be viewed as a response shift; i.e. that patients' expectations and interpretations shift during the peri-operative course. The patients may not expect to be as fit on day 1 after surgery as pre-operatively or on day 6, thus reflecting the higher cut-off point for significant fatigue during the first postoperative days.

In a paper comparing patients with chronic fatigue syndrome (CFS) and a nonclinical community group of participants Cella & Chalder [26] found the cut-off point in the CFQ to be ≥18; which is higher than our finding of a cut-off point at ≥16. This difference may be due to how the dichotomizations in the ROC analyses were made, and to apparent differences between the populations. First; the cut-off value in the CFS paper was based on the difference between the community group and patients diagnosed with CFS. This is a different approach than asking the participants to define whether the perceived fatigue is of considerable significance to them. Further, the cut-off point is influenced by fatigue intensity and prevalence. The peri-operative cut-off point is an average of recordings with varying fatigue intensity and prevalence, while the CFS/community analysis included CFS patients with very high fatigue levels. The mean(SD) pre-operative fatigue level found in the present study, 11.7(3.3), is significantly below what was found among the CFS-patients (n=361); 24.4(5.8), $p<0.0001$, but also lower than the mean score reported by the general-population group who participated in Cella & Chalder's British study (n=1615); 14.2(4.6), $p<0.0001$. The mean(SD) pre-operative fatigue level in the present study is comparable to what was found in a general sample of the Norwegian population; 12.2(4.0) [1]. The discrepancy between fatigue scores from general population samples may be due to cultural differences, illustrating the inherent challenge in comparing clinical data between countries.

In our literature search we identified no studies that described the level or development of clinically relevant fatigue during the peri-operative period. Several studies have examined POF in settings similar to those which we examined without reference to the clinical significance of the fatigue intensity. For example, Hill et al [11] compared fatigue assessed with ChrFS in patients undergoing

laparoscopic vs. open cholecystectomy. The laparoscopic group reported fatigue during the first month very similar to our findings, while the open group had more intense and prolonged fatigue (see Table A4 in the Appendix for details). Bisgaard et al [7] also assessed fatigue after laparoscopic cholecystectomy using the ChrFS. Unfortunately, the level of fatigue during the first seven days is reported only in a figure without scores quoted. The graph however, is similar to the findings in the present study, with fatigue scores significantly increased until post-operative day 5.

Being young and/or male was associated with a lower return rate. We do not know whether fatigue may be a factor limiting the ability or drive to return the forms. It may play a role regarding age, as increasing age generally has been associated with more missing responses [40-42]. On the other hand, a lower response rate among men is also common [1], as in our study and may thus explain some of the missing responses.

A proper power calculation was not done for the ROC-analyses, due to data already had been collected, according to power calculations for the corresponding aims of Study 1 and 2. However, we have performed a post-hoc assessment of the given sample sizes by using MedCalc®: Assuming a significance level of 0.05 and power of 0.80, prevalence rates between 24% and 49%, sample sizes between 100-300, and AUC between 0.81-0,89; the null hypothesis value would be AUC ≈ 0.70-0.81. Although admittedly post-hock analyses may be viewed as of limited value, this indicates that the sample sizes can be assumed adequate for the analyses we have presented.

Even though there were significant differences on day 3 in fatigue between younger and older patients, and between genders, these parameters did not influence the cut-off points in a consistent manner. Analysing subgroups necessarily reduces sample size analysed, and the inconsistencies in direction between the scales and between the Youden index *J* and C* indicate that our data were insufficient to examine these questions. This may be an area for future research.

Limitations of our study include that no externally validated measure of the 'clinical significance' of fatigue was included *a priori*. However, given the subjective nature of fatigue and its impacts, an item that enabled participants to state the presence or absence of 'significant' fatigue was deemed to be an appropriate choice. Limitations also include that, given that patients' expectations appear to influence the cut-off levels, the cut-off points identified for a day-surgery population will not necessarily be valid if the instruments are applied in other settings, e.g. operations necessitating hospitalisation. This is an area requiring further research. A further limitation is that no fatigue subscales from comprehensive HRQoL measures, such as the SF-36, were included in our study.

We consider strengths of our research to include the use of multiple data collection time points, and the recruitment of a mixed day-surgery population.

Conclusion

We have analysed three commonly used scales that may be applicable in POF and day-surgical recovery research. The cut-off point for clinically significant fatigue was found to be located near the middle of each scale's response range. With a defined cut-off point for clinically significant POF it is possible to analyse the clinical implications of fatigue better than by merely reporting mean fatigue values. This can be particularly valuable for diagnostic purposes and in treatment evaluation. Further, it may be possible to analyse and review data from earlier studies in light of clinical relevance.

Appendix

Table A1. The American Society of Anesthesiologists classification system

1	A normal healthy patient
2	A patient with mild systemic disease
3	A patient with severe systemic disease
4	A patient with a severe, life-threatening disease
5	A moribund patient that is not expected to survive an operation

Table A2. Sensitivity, Specificity, Likelihood Ratios and Predictive Values at the averaged Cut-off point for PO-FS (≥ 50/100), ChrFS (≥6/10) and CFQ (≥16/33).

	Averaged Cut-off point	**Sensitivity** (95% CI)	**Specificity** (95% CI)	**Likelihood Ratio** **LR+** (95% CI)	**LR-** (95% CI)	**Predictive Value %** **PPV** (95% CI)	**NPV** (95% CI)	**Accuracy %** (95% CI)
Day 0	PO-FS ≥50	67.6 (50.2-82.0)	94.2 (91.4-96.4)	11.70 (7.4-18.6)	0.34 (0.22-0.55)	53.2 (41.7-64.4)	96.8 (94.9-98.0)	91.9 (88.8-94.3)
	ChrFS ≥6	78.4 (61.8-90.2)	90.8 (87.4-93.5)	8.49 (5.9-12.1)	0.24 (0.13-0.44)	45.3 (36.7-54.2)	97.7 (95.9-98.8)	89.7 (86.3-92.4)
	CFQ ≥16	62.2 (44.8-77.5)	92.9 (89.8-95.3)	8.75 (5.62-13.61)	0.41 (0.27-0.62)	46.0 (35.4-57.0)	96.2 (94.3-97.5)	90.2 (86.9-92.9)
Day 1	PO-FS ≥50	94.4 (84.6-98.8)	55.4 (41.5-68.7)	2.12 (1.57-2.85)	0.10 (0.03-0.31)	67.1 (60.2-73.3)	91.2 (77.0-97.0)	74.6 (65.4-82.4)
	ChrFS ≥6	98.2 (90.1-99.9)	48.2 (34.7-62.0)	1.90 (1.47-2.45)	0.04 (0.01-0.27)	64.6 (58.6-70.2)	96.4 (79.2-99.5)	72.7 (63.4-80.8)
	CFQ ≥16	87.0 (75.1-94.6)	56.4 (42.3-69.7)	1.99 (1.45-2.74)	0.23 (0.11-0.48)	66.2 (58.8-72.9)	81.6 (68.1-90.2)	71.6 (62.1-79.8)
Day 3	PO-FS ≥50	90.2 (82.7-95.2)	73.4 (66.9-79.3)	3.39 (2.68-4.30)	0.13 (0.07-0.24)	62.6 (56.9-67.9)	93.8 (89.4-96.5)	79.0 (74.0-83.4)
	ChrFS ≥6	89.2 (81.5-94.5)	71.5 (64.8-77.5)	3.13 (2.50-3.92)	0.15 (0.09-0.27)	60.7 (55.2-65.9)	93.08 (88.4-96.0)	77.4 (72.3-81.9)
	CFQ ≥16	83.0 (74.2-89.8)	72.3 (65.7-78.3)	3.00 (2.36-3.81)	0.24 (0.15-0.37)	59.3 (53.4-64.9)	89.8 (84.9-93.2)	75.8 (70.6-80.5)
Day 6	PO-FS ≥50	88.5 (69.9-97.6)	91.6 (83.4-96.5)	10.49 (5.09-21.60)	0.13 (0.04-0.37)	76.7 (61.5-87.1)	96.2 (89.7-98.7)	90.8 (83.7-95.5)
	ChrFS ≥6	65.4 (44.3-82.8)	90.4 (81.9-95.8)	6.78 (3.32-13.88)	0.38 (0.22-0.65)	68.0 (51.0-81.3)	89.3 (83.0-93.4)	84.4 (76.2-90.6)
	CFQ ≥16	84.6 (65.1-95.6)	88.0 (79.0-94.1)	7.02 (3.84-12.85)	0.17 (0.07-0.43)	68.8 (54.6-80.1)	94.8 (88.1-97.8)	87.2 (79.4-92.8)

LR+: Positive Likelihood Ratio, **LR-**: Negative Likelihood Ratio, **PPV**: Positive Predictive Value, **NPV**: Negative Predictive Value

Table A3. Mean (SD) fatigue level measured with PO-FS, ChrFS and CFQ

	Day 0 (n=422)	**Day 1** (n=110)	**Day 3** (n=315)	**Day 6** (n=110)	**Day 30** (n=99)
PO-FS	31.2 (14.4)	60.4 (21.1)	47.3 (19.0)	37.5 (16.3)	22.4 (12.6)
ChrFS	3.26 (2.03)	6.91 (2.43)	5.23 (2.39)	3.94 (1.90)	2.15 (1.37)
CFQ	11.7 (3.27)	17.5 (4.76)	15.2 (4.41)	13.3 (3.33)	10.7 (2.43)

Table A4. Comparison of fatigue after laparoscopic (lap.) or open cholecystectomy measured with ChrFS ; from Hill [11] and Nostdahl [24]. Values are mean (SD).

	Day 0	Day 1	Day 3	Day 7/6	Day 28/30
Hill (open) (n=15)	3.1 (1.86)	6.3 (2.56)	5.2 (2.59)	6.1 (2.17)	3.4 (1.24)
Hill (lap.)(n=15)	2.7 (1.94)	6.6 (2.29)	5.8 (2.48)	3.9 (2.56)	2.3 (1.24)
Nostdahl (lap.)	3.48 (1.96) (n=120)	6.91 (2.43) (n=110)	5.43 (1.76) (n=110)	3.94 (1.90) (n=110)	2.15 (1.37) (n=99)

Figure A1. Christensen Fatigue Scale (ChrFS) [23]

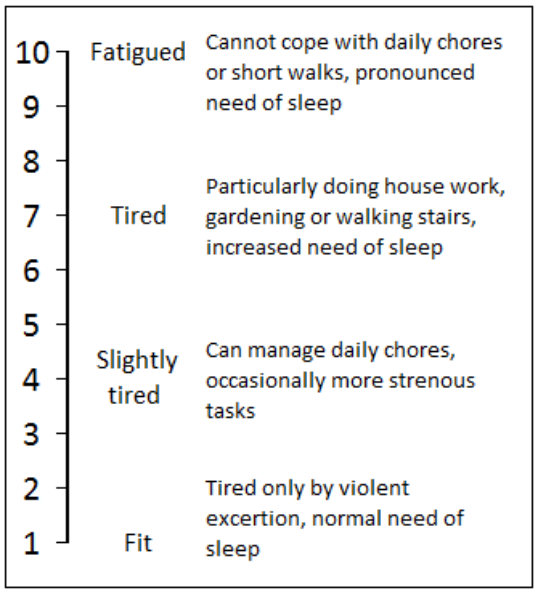
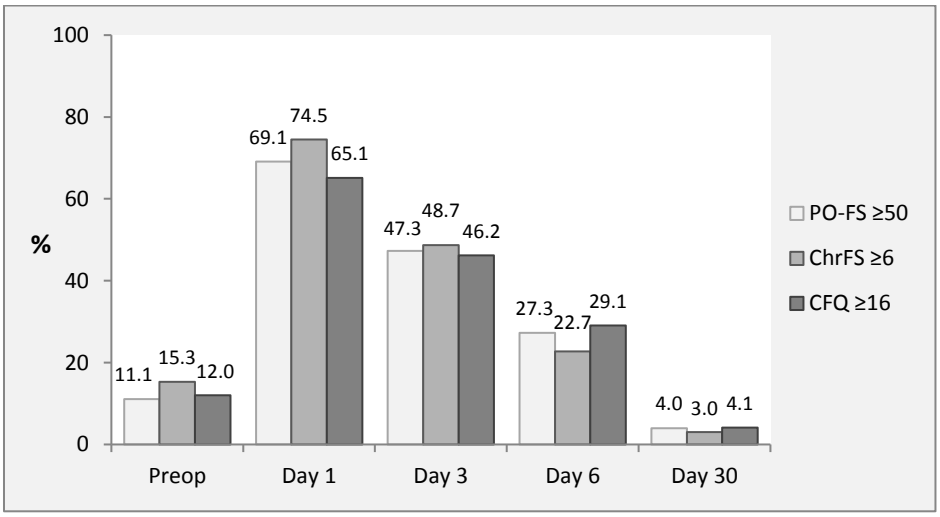


Figure A2. Percentage of day surgery patients with clinically significant fatigue level (i.e. above the cut-off point) measured with PO-FS, ChrFS and CFQ



References

1. Loge, J. H., Ekeberg, O., & Kaasa, S. (1998). Fatigue in the general Norwegian population: normative data and associations. *J.Psychosom.Res., 45*(1 Spec No), 53-65, doi:S0022-3999(97)00291-2 [pii].
2. Pawlikowska, T., Chalder, T., Hirsch, S. R., Wallace, P., Wright, D. J., & Wessely, S. C. (1994). Population based study of fatigue and psychological distress. *BMJ, 308*(6931), 763-766.
3. Bennett, J. A., Stewart, A. L., Kayser-Jones, J., & Glaser, D. (2002). The mediating effect of pain and fatigue on level of functioning in older adults. *Nurs Res, 51*(4), 254-265.
4. Whitehead, L. (2009). The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. *J Pain Symptom Manage, 37*(1), 107-128, doi:10.1016/j.jpainsymman.2007.08.019.
5. Zargar-Shoshtari, K., & Hill, A. G. (2009). Postoperative fatigue: a review. *World J Surg, 33*(4), 738-745, doi:10.1007/s00268-008-9906-0.
6. Rubin, G. J., Hardy, R., & Hotopf, M. (2004). A systematic review and meta-analysis of the incidence and severity of postoperative fatigue. *J Psychosom Res, 57*(3), 317-326, doi:10.1016/S0022-3999(03)00615-9.
7. Bisgaard, T., Klarskov, B., Rosenberg, J., & Kehlet, H. (2001). Factors determining convalescence after uncomplicated laparoscopic cholecystectomy. *Arch Surg, 136*(8), 917-921.
8. Minig, L., Velez, J. I., Trimble, E. L., Biffi, R., Maggioni, A., & Jeffery, D. D. (2013). Changes in short-term health-related quality of life in women undergoing gynecologic oncologic laparotomy: an associated factor analysis. *Support Care Cancer, 21*(3), 715-726, doi:10.1007/s00520-012-1571-z.
9. DeCherney, A. H., Bachmann, G., Isaacson, K., & Gall, S. (2002). Postoperative fatigue negatively impacts the daily lives of patients recovering from hysterectomy. *Obstet Gynecol, 99*(1), 51-57.
10. Dittner, A. J., Wessely, S. C., & Brown, R. G. (2004). The assessment of fatigue: a practical guide for clinicians and researchers. *J Psychosom Res, 56*(2), 157-170, doi:10.1016/S0022-3999(03)00371-4.
11. Hill, A. G., Finn, P., & Schroeder, D. (1993). Postoperative fatigue after laparoscopic surgery. *Aust N Z J Surg, 63*(12), 946-951.
12. Christensen, T., Hougard, F., & Kehlet, H. (1985). Influence of pre- and intra- operative factors on the occurrence of postoperative fatigue. *Br J Surg, 72*(1), 63-65.
13. Rubin, G. J., & Hotopf, M. (2002). Systematic review and meta-analysis of interventions for postoperative fatigue. *Br J Surg, 89*(8), 971-984, doi:10.1046/j.1365-2168.2002.02138.x.
14. Aarons, H., Forester, A., Hall, G., & Salmon, P. (1996). Fatigue after major joint arthroplasty: relationship to preoperative fatigue and postoperative emotional state. *J Psychosom Res, 41*(3), 225-233.
15. Hall, G. M., & Salmon, P. (2002). Physiological and psychological influences on postoperative fatigue. *Anesth Analg, 95*(5), 1446-1450.
16. Rubin, G. J., Cleare, A., & Hotopf, M. (2004). Psychological factors in postoperative fatigue. *Psychosom Med, 66*(6), 959-964, doi:10.1097/01.psy.0000143636.09159.f1.
17. Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., et al. (1993). Development of a fatigue scale. *J.Psychosom.Res., 37*(2), 147-153.
18. Michielsen, H. J., De Vries, J., & Van Heck, G. L. (2003). Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *J Psychosom Res, 54*(4), 345-352.

19. Hewlett, S., Dures, E., & Almeida, C. (2011). Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) for severity, effect, and coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). *Arthritis Care Res (Hoboken), 63 Suppl 11*, S263-286, doi:10.1002/acr.20579.
20. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst, 85*(5), 365-376.
21. Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care, 30*(6), 473-483.
22. Nostdahl, T., Bernklev, T., Raeder, J., Sandvik, L., & Fredheim, O. (2016). Postoperative fatigue; translation and validation of a revised 10-item short form of the Identity-Consequence Fatigue Scale (ICFS). *J Psychosom Res, 84*, 1-7, doi:10.1016/j.jpsychores.2016.03.002.
23. Christensen, T., Bendix, T., & Kehlet, H. (1982). Fatigue and cardiorespiratory function following abdominal surgery. *Br J Surg, 69*(7), 417-419.
24. Nostdahl, T., Fredheim, O. M., Bernklev, T., Doksrod, T. S., Mohus, R. M., & Raeder, J. (2017). A randomised controlled trial of propofol vs. thiopentone and desflurane for fatigue after laparoscopic cholecystectomy. *Anaesthesia, 72*(7), 864-869, doi:10.1111/anae.13909.
25. Paddison, J. S., Booth, R. J., Hill, A. G., & Cameron, L. D. (2006). Comprehensive assessment of peri-operative fatigue: development of the Identity-Consequence Fatigue Scale. *J Psychosom Res, 60*(6), 615-622, doi:10.1016/j.jpsychores.2005.08.008.
26. Cella, M., & Chalder, T. (2010). Measuring fatigue in clinical and community settings. *J Psychosom Res, 69*(1), 17-22, doi:10.1016/j.jpsychores.2009.10.007.
27. Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. *Biom J, 47*(4), 458-472.
28. Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.
29. Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol, 53*(5), 459-468.
30. Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol, 56*(5), 395-407.
31. Oliveira, M., Oliveira, G., Souza-Talarico, J., & Mota, D. (2016). Surgical Oncology: Evolution of Postoperative Fatigue and Factors Related to Its Severity. *Clin J Oncol Nurs, 20*(1), E3-8, doi:10.1188/16.CJON.E3-E8.
32. Schroeder, D., & Hill, G. L. (1993). Predicting postoperative fatigue: importance of preoperative factors. *World J Surg, 17*(2), 226-231.
33. Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2ed.). New York: John Wiley & Sons Inc.
34. Paddison, J. S., Sammour, T., Kahokehr, A., Zargar-Shoshtari, K., & Hill, A. G. (2011). Development and validation of the Surgical Recovery Scale (SRS). *J Surg Res, 167*(2), e85-91, doi:10.1016/j.jss.2010.12.043.
35. Calman, K. C. (1984). Quality of life in cancer patients--an hypothesis. *J Med Ethics, 10*(3), 124-127.
36. Rapkin, B. D., & Schwartz, C. E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health Qual Life Outcomes, 2*, 14, doi:10.1186/1477-7525-2-14.

37. Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health Qual Life Outcomes, 2*, 16, doi:10.1186/1477-7525-2-16.
38. Jansen, S. J., Stiggelbout, A. M., Nooij, M. A., Noordijk, E. M., & Kievit, J. (2000). Response shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. *Qual Life Res, 9*(6), 603-615.
39. Schwartz, C. E., Bode, R., Repucci, N., Becker, J., Sprangers, M. A., & Fayers, P. M. (2006). The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Qual Life Res, 15*(9), 1533-1550, doi:10.1007/s11136-006-0025-9.
40. Coste, J., Quinquis, L., Audureau, E., & Pouchot, J. (2013). Non response, incomplete and inconsistent responses to self-administered health-related quality of life measures in the general population: patterns, determinants and impact on the validity of estimates - a population-based study in France using the MOS SF-36. *Health Qual Life Outcomes, 11*, 44, doi:10.1186/1477-7525-11-44.
41. Peyre, H., Coste, J., & Leplege, A. (2010). Identifying type and determinants of missing items in quality of life questionnaires: Application to the SF-36 French version of the 2003 Decennial Health Survey. *Health Qual Life Outcomes, 8*, 16, doi:10.1186/1477-7525-8-16.
42. Loge, J. H., Ekeberg, O., & Kaasa, S. (1998). Fatigue in the general Norwegian population: normative data and associations. *Journal of Psychosomatic Research, 45*(1 Spec No), 53-65, doi:S0022-3999(97)00291-2 [pii].