

The Piper Fatigue Scale-12 (PFS-12): psychometric findings and item reduction in a cohort of breast cancer survivors

Bryce B. Reeve · Angela M. Stover · Catherine M. Alfano ·
Ashley Wilder Smith · Rachel Ballard-Barbash · Leslie Bernstein ·
Anne McTiernan · Kathy B. Baumgartner · Barbara F. Piper

Received: 24 May 2012 / Accepted: 12 August 2012
© Springer Science+Business Media, LLC. 2012

Abstract Brief, valid measures of fatigue, a prevalent and distressing cancer symptom, are needed for use in research. This study's primary aim was to create a shortened version of the revised Piper Fatigue Scale (PFS-R) based on data from a diverse cohort of breast cancer survivors. A secondary aim was to determine whether the PFS captured multiple distinct aspects of fatigue (a multidimensional model) or a single overall fatigue factor (a unidimensional model). Breast cancer survivors ($n = 799$; stages in situ through IIIa; ages 29–86 years) were recruited through three SEER registries (New Mexico, Western Washington, and Los Angeles, CA) as part of the Health, Eating, Activity, and Lifestyle (HEAL) study. Fatigue was measured approximately 3 years post-diagnosis using the 22-item PFS-R that has four subscales (Behavior, Affect, Sensory, and Cognition). Confirmatory

factor analysis was used to compare unidimensional and multidimensional models. Six criteria were used to make item selections to shorten the PFS-R: scale's content validity, items' relationship with fatigue, content redundancy, differential item functioning by race and/or education, scale reliability, and literacy demand. Factor analyses supported the original 4-factor structure. There was also evidence from the bi-factor model for a dominant underlying fatigue factor. Six items tested positive for differential item functioning between African-American and Caucasian survivors. Four additional items either showed poor association, local dependence, or content validity concerns. After removing these 10 items, the reliability of the PFS-12 subscales ranged from 0.87 to 0.89, compared to 0.90–0.94 prior to item removal. The newly developed PFS-12 can be used to assess fatigue in African-American

B. B. Reeve (✉)
Department of Health Policy & Management, Gillings School of
Global Public Health, University of North Carolina at Chapel
Hill, 1101-D McGavran-Greenberg Hall, 135 Dauer Drive,
CB 7411, Chapel Hill, NC 27599-7411, USA
e-mail: bbreeve@email.unc.edu

B. B. Reeve · A. M. Stover
Lineberger Comprehensive Cancer Center, University of North
Carolina at Chapel Hill, CB# 7295, Chapel Hill,
NC 27599-7295, USA

A. M. Stover
Department of Health Behavior, Gillings School of Global
Public Health, University of North Carolina at Chapel Hill,
Rosenau Hall, CB# 7440, Chapel Hill, NC 27599, USA

C. M. Alfano
Office of Cancer Survivorship, Division of Cancer Control and
Population Sciences, National Cancer Institute, 6116 Executive
Blvd, Suite 404, Bethesda, MD 20892, USA

A. W. Smith · R. Ballard-Barbash
Applied Research Program, Division of Cancer Control and
Population Sciences, National Cancer Institute, EPN 4005,
6130 Executive Blvd, Bethesda, MD 20892-7344, USA

L. Bernstein
Division of Cancer Etiology, Department of Population
Sciences, Beckman Research Institute of the City of Hope,
1500 East Duarte Road, Duarte, CA 91010, USA

A. McTiernan
Prevention Center, Fred Hutchinson Cancer Research Center,
P.O. Box 19024, Seattle, WA 98109, USA

K. B. Baumgartner
Epidemiology and Population Health, University of Louisville,
485 E. Gray St., Louisville, KY 40292, USA

B. F. Piper
University of Arizona School of Nursing, 10684 N. 113th Street,
Scottsdale, AZ 85259-4034, USA

and Caucasian breast cancer survivors and reduces response burden without compromising reliability or validity. This is the first study to determine PFS literacy demand and to compare PFS-R responses in African-Americans and Caucasian breast cancer survivors. Further testing in diverse populations is warranted.

Keywords Fatigue · Breast cancer survivors · Patient-reported outcomes · Piper Fatigue Scale · Psychometrics

Introduction

Cancer-related fatigue is defined as the perception of unusual tiredness that varies in pattern or severity and can affect the functional ability of cancer survivors [1–3]. A recent literature review of 18 studies measuring symptoms in adults during active treatment found fatigue to be experienced by 62 % of patients [4]. Fatigue continues to be a prevalent and distressful symptom for cancer survivors years after active treatment ends [5]. Among disease-free breast cancer survivors approximately 3 years post-diagnosis, 41 % reported moderate to severe fatigue levels [6].

Existing self-report fatigue questionnaires vary on a number of factors including questionnaire length, reference period, and the response scale [7]. One key factor that differentiates these questionnaires is whether they provide a single overall score (a unidimensional measure) or multiple scores that reflect different attributes of the fatigue experience (a multidimensional measure). The Piper Fatigue Scale (PFS) is one of the commonly used multidimensional fatigue measures in the cancer research field and includes subdomains of behavioral, affective, sensory, and cognitive/mood attributes of fatigue [8]. The original PFS consisted of 40 questions (items) and the revised PFS (PFS-R) includes 22 questions [8].

While the PFS-R has been translated into multiple languages [9–13], more researchers and clinicians would likely use the PFS if it were shorter in length. Respondent burden is a concern as most research studies include a battery of self-report measures to capture the health-related quality of life (HRQOL) of cancer patients. In addition, prospective studies are encouraged over cross-sectional studies to capture variations in fatigue experience over time; thus, reducing response burden is a necessity. Finally, to increase research of fatigue in diverse samples of survivors, fatigue instruments must be valid and reliable across different racial/ethnic groups.

The overall goal of this psychometric study was to analyze the PFS-R in a diverse cohort of breast cancer survivors to reduce the number of questions in the scale. To accomplish this goal, the original four subdomain structure of the PFS-R

was re-examined. The assumption guiding this study was that if analyses confirmed the multi-dimensionality of the scale, then the shortened scale must maintain its ability to provide reliable measurement for each subdomain.

Methods

Participants

As described elsewhere [6, 14, 15], participants were female breast cancer survivors enrolled in the Health, Eating, Activity, and Lifestyle (HEAL) study. A total of 1,183 women were recruited through the population-based Surveillance, Epidemiology, and End Results (SEER) cancer registries in New Mexico, Los Angeles County, and Western Washington. These women participated in a baseline in-person interview within 1 year after diagnosis. Of those, 944 (80 %) participated in a follow-up assessment approximately 2 years after the first interview, and 858 (73 %) completed an additional HRQOL questionnaire approximately 3 years (40.5 months) after diagnosis.

For this study, 57 women were excluded for recurrent or new primary breast cancer before completing their HRQOL survey, as were 2 women with incomplete fatigue data. The final sample of 799 women included 436 from New Mexico, 195 from Los Angeles County, and 168 from Western Washington. All African-Americans were recruited from Los Angeles and most Hispanic women were recruited from New Mexico. In addition, the African-Americans were restricted to 35–64 years of age to focus on younger breast cancer survivors. Participants were diagnosed with in situ, Stage I, II, or IIIA breast cancer.

Informed consent was obtained from each participant at each assessment. The study was approved by the Institutional Review Board at each site, in accord with assurances filed with and approved by the U.S. Department of Health and Human Services.

Measures

Demographic and background variables

Age, education, and race/ethnicity data were collected at baseline. Data on marital status, household income, height, employment, menopausal status, smoking, and current weight were collected at the first follow-up survey. Self-reported comorbidities were captured on the HRQOL survey.

Clinical variables

Stage of disease was obtained from each SEER registry database. Medical records were abstracted to obtain

treatment information and these data were supplemented with SEER registry data. Women were asked if they were taking tamoxifen (yes/no) at the 24-month follow-up. In the HRQOL assessment, women self-reported about reconstructive surgery and lymphedema.

Fatigue

The PFS-R was included in the HRQOL survey. It is a 22-item scale that measures four subscales: behavior (6 items), affect (5 items), sensory (5 items), and cognition/mood (6 items) [8]. Each item has 11 response categories on a 0–10 metric with verbal descriptors anchoring the endpoints. Each subscale is scored individually and then aggregated together for an overall score, with higher scores reflecting more fatigue. The HEAL study used an adapted version of the PFS-R that asked survivors to rate their fatigue over the past month rather than the past week. This extended reference period was used to minimize the effect of acute situational events and to enhance the assessment of the survivor's general state of fatigue. Previous studies have found evidence that the PFS-R has acceptable internal consistency and evidence for validity with cancer patients [16–18] including the adapted version used in this study [6].

The HEAL HRQOL questionnaire included an initial screening question that had respondents skip the PFS-R behavioral and affective subscales if they indicated that they had “no fatigue” over the past 4 weeks. These subscales were skipped because the questions asked the respondent to further clarify the fatigue they were currently experiencing. Approximately 36 % of the 799 breast cancer survivors did not complete those two subscales ($N = 291$), but did complete the sensory and cognitive/mood subscales.

Analyses

Before selecting items for removal from the PFS-R, the dimensionality of the PFS was re-confirmed to determine if psychometric analyses should be performed at the subscale level (i.e., referred to as the “multidimensional model” with four subdomains) or at the overall fatigue level (i.e., referred to as the “unidimensional model” with all 22 items loading on a single fatigue factor).

Confirmatory factor analysis was used to evaluate the fit of the multidimensional and unidimensional models to the HEAL data. In addition, a bifactor model was fit to the data to examine the extent to which the items loaded on a general fatigue factor compared with the items' loadings on their respective fatigue subdomains. In the bifactor model, larger loadings on the general factor than the subdomain-specific factors would suggest that a dominant single

fatigue factor accounted for a majority of the variation observed among the response data. Larger loadings on the subdomain-specific factors than the general factor would suggest that the multidimensional model was a better solution [19].

Figure 1 provides illustration of the three models tested. Model fit was assessed with multiple indices, with specific criteria noted for good model fit including: root mean squared error of approximation ($RMSEA < .06$), comparative fit index ($CFI > .95$), and Tucker–Lewis Index ($TLI > .95$) [20–23]. MPLUS software (version 6.1) was used for each factor analysis that uses a maximum-likelihood estimator for continuous data.

The next step included reviewing the psychometric properties of the PFS-R with specific attention given to shortening scale length. Six criteria, described below, were considered for making item selections: (1) content validity, (2) strength of relationship with fatigue, (3) content redundancy, (4) differential item functioning by race and/or education, (5) reliability, and (6) literacy demand.

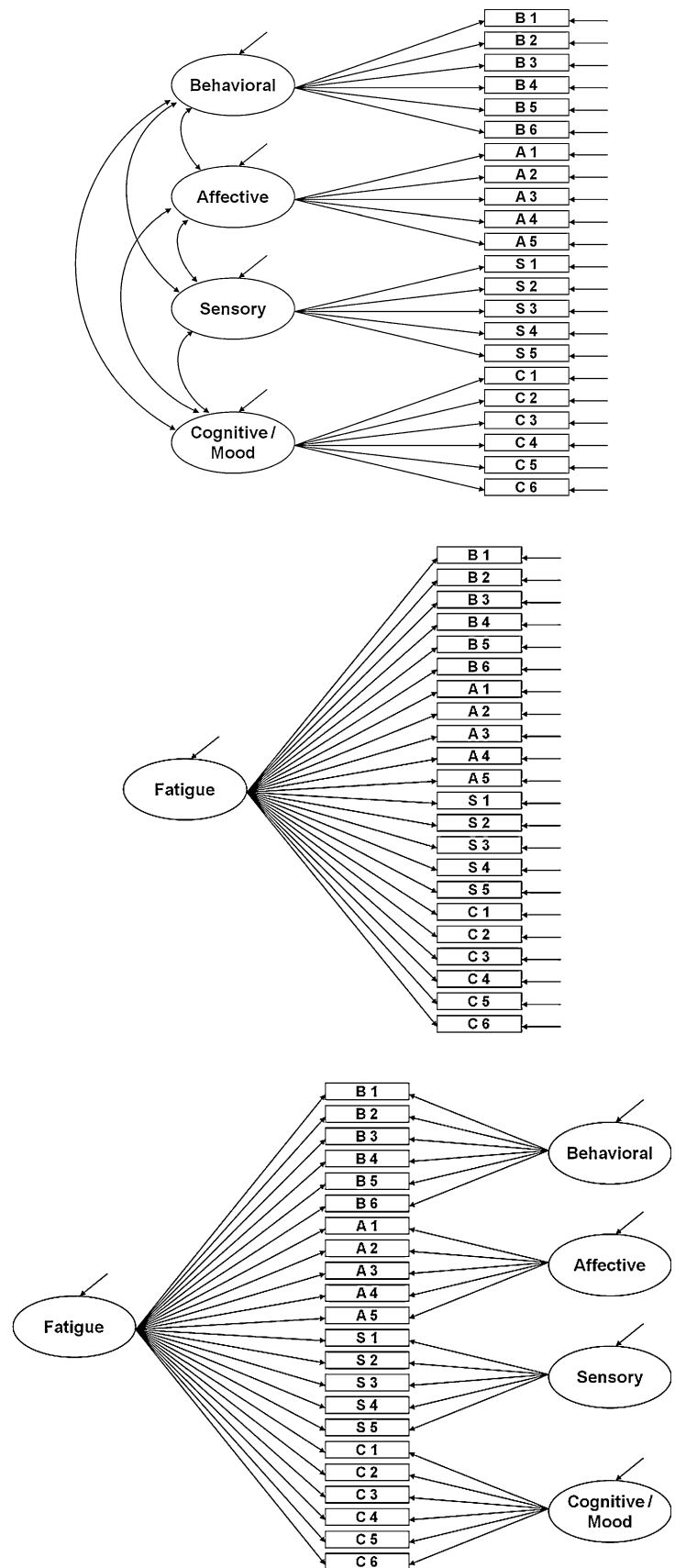
Content validity of each item examined the extent to which the item reflected a critical attribute of the fatigue domain being measured [24]. The content validity review was especially important for those items being considered for removal based on evidence from the psychometric analyses.

The *strength of the relationship* between each item and the overall scale was assessed using three statistical methods. The first method was item-total score correlations, performed in SAS (version 9.2). Secondly, factor loadings were examined from the factor analyses. Finally, item response theory (IRT) models were used to examine how well each item was related to the fatigue domain (i.e., discrimination) and how the item's response categories reflected different levels of fatigue (i.e., threshold). The IRTPRO software (ver. 2.1; Scientific Software, Inc.) was used to estimate IRT parameters for Samejima's Graded Response Model [25–27].

Content redundancy was assessed with inter-item Pearson correlations (from SAS), residual correlations abstracted from the factor analyses (from MPLUS), local dependence matrices, and IRT information functions (from IRTPRO). IRTPRO reports a standardized local dependence χ^2 statistic where values over 10 may be considered problematic [28]. Local dependence suggests there is excessive correlation between the items even after controlling for the underlying fatigue domain being measured. Authors' expertise was also used to judge content redundancy.

Differential item functioning was examined to identify possible response bias by race or by education. Due to the small sample size of Hispanic women and those classified as “other” race, differential item functioning could only be examined between non-Hispanic whites and African-Americans.

Fig. 1 Structural equation models fit to the data on the revised-Piper Fatigue Scale: multidimensional model, unidimensional model, and bifactor model



Education was evaluated by comparing response data between those with a high school education or less to those with more than a high school education. Differential item functioning was tested within an IRT framework (using IRTPRO) using Wald χ^2 to evaluate statistical significance. Because of the large number of response categories (11) per item with the possibility of small sample sizes for each category, a sensitivity analysis was performed comparing the differential item functioning findings using all 11 categories with an alternative 4-category response scale created by collapsing response categories (0 = none; 1–3 = mild; 4–6 = moderate; 7–10 = severe fatigue).

Scale reliability was examined using Cronbach's coefficient alpha. We also examined scale precision with IRT information functions. We ideally wanted to maintain each fatigue subscale's reliability above .80, which is more than adequate precision for group-level comparisons in cancer research settings [29].

Literacy demand was evaluated with the Lexile Framework for Reading [30]. A Lexile value is based on two strong predictors of how difficult a text is to comprehend: word frequency and sentence length. Lexile measures provide corresponding grade levels ranging from first grade to post-high school. Scores were averaged across PFS items to produce the mean literacy demand of the scale and corresponding mean reading grade level for the PFS. Lexile analyses have been previously used to evaluate HRQOL questionnaires [31]. The Institute of Medicine guidelines state that health communication materials should be written at a mean 8th grade reading level or below [32].

Together these six criteria were used to inform the selection of items to remain in the PFS-12. Guiding our decisions, the senior author, Barbara Piper, the originator of the PFS, provided oversight in these judgments bringing her collective experience in the use of her scale in different populations and settings worldwide. At a minimum, it was decided that three questions from each of the four PFS subscales would remain in the new scale to maintain the ability to factor analyze the scale (PFS-12). Only the important findings that led to the decision to remove or keep an item in the PFS-12 are reported in “Results” section. After the final selection of the PFS-12 items, factor analysis and DIF testing were not repeated because the items need to be administered together to a new sample.

Results

Characteristics of the 799 breast cancer survivors are summarized in Table 1. Chi-square tests indicated significant differences in distributions of survivors by race/ethnicity for age, marital status, education, employment

status, income, BMI, menopausal status, comorbid conditions, cancer stage, treatment type, and lymphedema. This is not surprising given the different SEER registries and methods of recruiting at each site [6].

Dimensionality of the 22-item PFS-R

The confirmatory factory analysis model results are shown in Table 2. The unidimensional model did not fit the data (RMSEA = .15, CFI = .73, TLI = .70). The multidimensional solution showed adequate fit (RMSEA = .08, CFI = .92, TLI = .91), consistent with how the PFS-R has conventionally been scored and interpreted. The correlations among the four subscales ranged from 0.67 (affective and cognition/mood subscales) to 0.79 (sensory and cognition/mood subscales). The bifactor model showed good fit (RMSEA = .07, CFI = .95, TLI = .94).

The bifactor model results and the high correlations among the subscales suggest there may be evidence of a dominant underlying fatigue factor. A one-factor solution accounted for 58 % of the variance observed in the data, while the second factor accounted for only 7 % of the variance. To be consistent with prior use of the PFS-R and to maintain content validity, item reduction analyses proceeded on a subscale-by-subscale basis.

Selecting items for the PFS-12

Table 3 presents a summary of some of the psychometric analyses performed on the items within each subscale, with unshaded items being retained for the PFS-12. While differential item functioning by education level was performed, no differential item functioning was detected.

In the behavioral subscale, no item tested positive for differential item functioning. The item “engage in sexual activity” had the lowest item-total score correlation and was found by the IRT model (not presented) to have poor discrimination. The item “fatigue causing distress” was judged to have a more affective than behavioral content. While no psychometric issue emerged, the item on “socializing with friends” was thought to be highly related to the item on “engaging in enjoyable activities” but the latter captured a broader range of impact on activities that could be done alone or in a group setting. However, the item “engaging in enjoyable activities” had the highest literacy demand, but was retained for noted content reasons.

In the affective subscale, four items tested positive for differential item functioning between African-Americans and whites. In addition, two item pairs were found to exhibit local dependence. Given this psychometric evidence, we relied on content expertise and literacy demand

Table 1 Survivor characteristics by race and ethnicity

	African-American non-Hispanic breast cancer survivors (<i>n</i> = 196) <i>N</i> (%)	Caucasian non-Hispanic breast cancer survivors (<i>n</i> = 484) <i>N</i> (%)	Hispanic breast cancer survivors (<i>n</i> = 95) <i>N</i> (%)	Other race non-Hispanic breast cancer survivors (<i>n</i> = 24) <i>N</i> (%)	Total sample (<i>n</i> = 799) <i>N</i> (%)
Age at diagnosis**					
Mean age ± SD	51.43 ± 7.84	57.30 ± 10.77	55.82 ± 11.58	51.33 ± 6.13	55.51 ± 10.43
Marital status**					
Married	82 (42)	301 (62)	50 (53)	15 (63)	448 (56)
Widowed	25 (13)	58 (12)	13 (14)	1 (4)	97(12)
Divorced/separated	70 (36)	79 (16)	19 (20)	5 (21)	173 (22)
Never married	19 (10)	25 (5)	4 (4)	3 (12)	51 (6)
Education**					
≤High school graduate	70 (36)	85 (18)	41 (43)	7 (29)	203 (25)
Some college	85 (43)	167 (35)	34 (36)	5 (21)	291 (36)
College graduate	24 (12)	116 (24)	9 (9)	7 (29)	156 (20)
Graduate school	17 (9)	115 (24)	11 (12)	5 (21)	148 (19)
Employment status**					
Working	119 (61)	267 (55)	45 (47)	17 (71)	448 (56)
Not working— leave,retired, unemployed	49 (25)	196 (41)	43 (45)	7 (29)	295 (37)
Not working— disabled	27 (14)	0 (0)	0 (0)	0 (0)	27 (3)
Income**					
≤20,000	55 (28)	60 (12)	19 (20)	3 (12)	137 (17)
20,001–50,000	71 (36)	145 (30)	40 (42)	5 (21)	261 (33)
50,001–70,000	32 (16)	149 (31)	14 (15)	15 (63)	210 (26)
≥70,000	34 (17)	70 (15)	7 (7)	1 (4)	112 (14)
BMI**					
<25	44 (22)	255 (53)	47 (50)	12 (50)	358 (45)
25–29	55 (28)	133 (27)	31 (33)	1 (4)	220 (28)
≥30	97 (50)	96 (20)	17 (18)	11 (46)	221 (28)
Menopausal status*					
Post-menopausal	145 (74)	361 (75)	63 (66)	16 (67)	585 (73)
Pre-menopausal	36 (18)	84 (17)	19 (20)	3 (12)	142 (18)
Unclassifiable	15 (8)	19 (4)	6 (6)	5 (21)	45 (6)
Smoking status—ns					
Not at all (or N/A)	167 (85)	417 (86)	77 (81)	20 (63)	681 (85)
Every day	21 (11)	36 (7)	5 (5)	3 (12)	65 (8)
On some days	8 (4)	11 (2)	6 (6)	1 (4)	26 (3)
Comorbidities—ns					
None	47 (24)	133 (28)	29 (31)	5 (21)	214 (27)
1	61 (31)	131 (27)	20 (21)	9 (38)	221 (28)
≥2	88 (45)	200 (41)	39 (41)	10 (42)	337 (42)
# Comorbid conditions that limit activities*					
0	139 (71)	350 (72)	65 (68)	20 (83)	574 (72)
1	35 (18)	85 (17)	15 (16)	1 (4)	136 (17)
2	8 (4)	20 (4)	6 (6)	1 (4)	35 (4)
3–4	14 (7)	9 (2)	2 (2)	2 (8)	27 (3)

Table 1 continued

	African-American non-Hispanic breast cancer survivors (<i>n</i> = 196) <i>N</i> (%)	Caucasian non-Hispanic breast cancer survivors (<i>n</i> = 484) <i>N</i> (%)	Hispanic breast cancer survivors (<i>n</i> = 95) <i>N</i> (%)	Other race non-Hispanic breast cancer survivors (<i>n</i> = 24) <i>N</i> (%)	Total sample (<i>n</i> = 799) <i>N</i> (%)
Stage of breast cancer**					
In situ—stage 0	40 (20)	112 (23)	19 (20)	6 (25)	177 (22)
Localized—stage I	62 (31)	215 (44)	42 (44)	10 (42)	329 (41)
Regional—stage II–IIIA	75 (38)	125 (26)	27 (28)	7 (29)	234 (29)
Unstaged	19 (10)	32 (7)	7 (7)	1 (4)	59 (7)
Surgery*					
No surgical procedure	2 (1)	9 (2)	2 (2)	0 (0)	13 (2)
Partial mastectomy	104 (53)	329 (68)	63 (66)	17 (71)	513 (64)
Total/modified mastectomy	90 (46)	146 (30)	30 (32)	7 (29)	273 (34)
Adjuvant therapy**					
None	73 (37)	150 (31)	31 (33)	5 (21)	259 (32)
Radiation therapy	46 (24)	204 (42)	33 (35)	11 (46)	294 (37)
Chemotherapy	32 (16)	33 (7)	5 (5)	2 (8)	72 (9)
Radiation therapy and chemotherapy	45 (23)	97 (20)	26 (27)	6 (25)	174 (22)
Tamoxifen at follow-up—ns					
No	115 (59)	247 (51)	48 (55)	13 (54)	423 (55)
Yes	81 (41)	217 (45)	40 (45)	11 (46)	349 (45)
Lymphedema**					
Never	139 (71)	401 (83)	74 (78)	22 (92)	636 (80)
Yes, not current	6 (3)	30 (6)	11 (12)	1 (4)	48 (6)
Yes	51 (26)	53 (11)	10 (10)	1 (4)	115 (14)
Bothered by lymphedema in past 3 months*					
Not at all	15 (26)	27 (32)	6 (29)	0 (0)	48 (30)
A little	21 (37)	38 (46)	12 (57)	2 (100)	73 (45)
A fair amount	9 (16)	14 (17)	2 (10)	0 (0)	25 (15)
Much/very much	12 (21)	4 (5)	1 (5)	0 (0)	17 (10)

* $p < .05$, ** $p < .01$

to select the final three items for the affective subscale of the PFS-12.

For the sensory subscale, two items tested positive for differential item functioning. In addition, three sets of items were found to be locally dependent. One item from each pair was removed to resolve the local dependence.

In the cognitive/mood subscale, two items tested positive for differential item functioning and three item pairs had evidence for local dependence.

For each subscale, the differential item functioning analyses were repeated, as a sensitivity analysis, using the collapsed 4-category response scale. There still was no differential item functioning by education level. Consistent with prior findings, the item “exhilarated to depressed” on the cognitive subscale tested positive for differential item functioning by race. However, the item “able to remember” did not demonstrate differential item functioning with the collapsed categories. The item “lively to listless” on

Table 2 Factor loadings for a 4-factor, 1-factor, and bifactor model

Items	4-Factor model: factor loading	1-Factor model: factor loading	Bifactor general model: factor loading	Bifactor specific model: factor loading
<i>Behavioral subscale</i>				
To What Degree is the Fatigue You Are Now Feeling Causing You Distress? (No Distress-Great Deal of Distress)	.77	.71	.64	.41
...Interfering With Your Ability to Complete Your Work or School Activities? (None-Great Deal)	.80	.72	.63	.47
... Interfering With Your Ability to Visit or Socialize With Your Friends? (None-Great Deal)	.88	.77	.69	.56
...Interfering With Your Ability to Engage in Sexual Activity? (None-Great Deal)	.54	.53	.47	.27
...Interfering With Your Ability to Engage in the Kind of Activities You Enjoy Doing? (None-Great Deal)	.91	.80	.70	.58
How Would You Describe the Degree of Intensity or Severity of the Fatigue Which You Are Now Experiencing? (Mild to Severe)	.90	.82	.73	.51
<i>Affective subscale</i>				
To What Degree Would You Describe the Fatigue Which You Are Experiencing Now as Being: Pleasant to Unpleasant	.88	.77	.68	.58
Agreeable to Disagreeable	.88	.73	.65	.64
Protective to Destructive	.80	.71	.67	.40
Positive to Negative	.87	.74	.67	.52
Normal to Abnormal	.79	.76	.69	.35
<i>Sensory subscale</i>				
To What Degree Are You Now Feeling: Strong to Weak	.79	.80	.75	.26
Awake to Sleepy	.80	.78	.71	.35
Lively to Listless	.93	.87	.80	.47
Refreshed to Tired	.92	.86	.78	.50
Energetic to Unenergetic	.92	.84	.77	.51
<i>Cognitive subscale</i>				
To What Degree Are You Now Feeling: Patient to Impatient	.76	.71	.77	.08
Relaxed to Tense	.77	.75	.81	.04
Exhilarated to Depressed	.74	.76	.81	-.03
Able to Concentrate to Unable to Concentrate	.89	.76	.79	.40
Able to Remember to Unable to Remember	.85	.71	.70	.57
Able to Think Clearly to Unable to Think Clearly	.87	.74	.73	.57
<i>Model fit statistics</i>				
RMSEA (Root Mean Square Error of Approximation): Criterion: $\leq .08$.08	.15	.07	
CFI (Comparative Fit Indices): Criterion: $\geq .95$.92	.73	.95	
TLI (Tucker-Lewis Index): Criterion: $\geq .95$.91	.70	.94	

Bifactor general factor loading includes all items loading on a single general fatigue factor. Bifactor specific factor loading includes loadings of items on their respective subdomain of fatigue

the sensory subscale tested positive for differential item functioning by race for both the 11-point and 4-point scale. With the 4-point scale, the sensory item “energetic to unenergetic” no longer showed differential item functioning by race. Different from prior findings, the sexual activity item in the behavioral subscale tested positive for

differential item functioning by race. In the affective subscale, only one item showed differential item functioning by race (“pleasant to unpleasant”), but was retained for content validity reasons.

Overall, the PFS-12 maintained high scale reliability ($r = .92$) with the original 22-item PFS-R having a

Table 3 Review of psychometric findings for items removed and retained in the PFS-12 ordered by subscale ($N = 799$)

Items	Item-subscale correlation	Content redundancy (local dependence)	DIF between African-American and Caucasian survivors	Lexile grade equivalent (literacy demand)
<i>Behavioral subscale</i>				
To What Degree is the Fatigue You Are Now Feeling Causing You Distress? (No Distress-Great Deal of Distress)	.80			5
...Interfering With Your Ability to Complete Your Work or School Activities? (None-Great Deal)	.83			10
... Interfering With Your Ability to Visit or Socialize With Your Friends? (None-Great Deal)	.87			11
...Interfering With Your Ability to Engage in Sexual Activity? (None-Great Deal)	.64			10
...Interfering With Your Ability to Engage in the Kind of Activities You Enjoy Doing? (None-Great Deal)	.88			>12
How Would You Describe the Degree of Intensity or Severity of the Fatigue Which You Are Now Experiencing? (Mild to Severe)	.88			8
<i>Affective subscale</i>				
To What Degree Would You Describe the Fatigue Which You Are Experiencing Now as Being: Pleasant to Unpleasant	.88	LD = 20.4 $r = .83$	$\chi^2(10) = 29.4, p < .01$ (threshold parameter)	8
Agreeable to Disagreeable	.89			9
Protective to Destructive	.84	LD = 18.5 $r = .72$	$\chi^2(1) = 9.9, p < .01$ (discrimination parameter)	8
Positive to Negative	.90			8
Normal to Abnormal	.84			8
<i>Sensory subscale</i>				
To What Degree Are You Now Feeling: Strong to Weak	.84			3
Awake to Sleepy	.85	LD = 14.9 $r = .75$	$\chi^2(1) = 26.6, p < .01$ (discrimination parameter)	3
Lively to Listless	.93			3
Refreshed to Tired	.92	LD = 21.5 $r = .83$	$\chi^2(1) = 18.9, p < .01$ (discrimination parameter)	3
Energetic to Unenergetic	.92			3
<i>Cognitive/Mood subscale</i>				
To What Degree Are You Now Feeling: Patient to Impatient	.86	LD = 17.4 $r = 0.75$		3
Relaxed to Tense	.86			3
Exhilarated to Depressed	.82		$\chi^2(10) = 24.1, p < .01$ (threshold parameter)	3
Able to Concentrate to Unable to Concentrate	.88	LD = 13.8 $r = .79$	$\chi^2(1) = 14.1, p < .01$ (discrimination parameter)	5
Able to Remember to Unable to Remember	.82			4
Able to Think Clearly to Unable to Think Clearly	.81			5

Gray-shaded rows = item removed; unshaded rows = kept in PFS-12 item; LD = local dependence χ^2 value; r = correlation. Differential item functioning (DIF) testing was only performed with African-Americans and non-Hispanic whites (total $n = 680$). All other analyses were performed on the full sample ($n = 799$)

reliability of .96. In addition, reliability for the PFS-12 subscales remained above .80: behavior (.89), affective (.87), sensory (.87), and cognition/mood (.87). The formatted version of the PFS-12 and accompanying scoring manual is provided in [Appendix](#).

Discussion

In a diverse sample of breast cancer survivors approximately 3 years from diagnosis, the 22-item PFS-R was shortened to the 12-item PFS-12 based on multiple criteria including reliability, validity, literacy demand, and response bias (i.e., differential item functioning) by race. This ten-item reduction in the PFS-R has the potential to reduce response burden in future studies while still maintaining a high level of precision for group-level comparisons at the subscale level (i.e., behavioral, affective, sensory, and cognitive/mood aspects of fatigue).

After testing alternative factor analytic models, results indicated that a four-factor model representing the original four PFS-R subdomains fit the data better than a one-factor model. This provides evidence that the four subscales may represent distinct aspects of the fatigue experience as reported by breast cancer survivors.

It is possible, however, that the findings for the multi-dimensional model may be more methodological artifact than distinct fatigue subdomains. For instance, items within each subscale are worded more similarly than items in other subscales; in addition, items presented next to each other on average will be more related than items further away [33]. If the “distinctiveness” of the factors were purely artifact, this would suggest that the fatigue experience itself similarly affects all aspects of a person’s life: behaviors, sensory, affect, and cognition/mood. In support of this perspective, a dominant underlying factor was found accounting for 58 % of the variance and high correlations were found among the four subscales (all correlations >0.67). In addition, the bifactor model supported a dominant general fatigue factor after extracting clustering of items within each subscale. This study cannot disentangle whether evidence for a multidimensional model is due to the distinctness of each fatigue subdomain or due to artifact; likely it is a mixture of both.

Retaining the ability to provide scores on the PFS-12 for both overall fatigue and individual subdomains is valuable for researchers. It allows investigators to characterize the extent to which a health condition or treatment affects different aspects of the fatigue experience. In addition, an intervention may differentially affect specific fatigue subdomains; e.g., meditation may improve cognition/mood and have less effect on behavioral fatigue. For other

research studies, overall fatigue may only be of interest as a mediating variable or an outcome. For each application, the PFS-12 can be used without extensive response burden.

Evidence in the literature from the US and abroad supports the PFS as a multi-dimensional measure. Three psychometric studies of non-English translations of the PFS-R, Greek [9], French [10], and Italian [34], also identified the same four subscales. In contrast, two studies found evidence for a 3-factor solution combining the domains of sensory with cognition/mood. One of these studies used a Brazilian translated version of the PFS-R [13]; the other used the English version in caregivers of stroke survivors [35]. Lending additional validity to the multi-dimensional model of fatigue, recent evidence suggested that the mechanisms driving fatigue may differ by fatigue dimension: increased inflammation (generally thought to produce fatigue in cancer survivors) was related to the behavioral and sensory aspects of fatigue but not to the more psychological aspects (affective and cognitive subscales) [36].

The review of literacy demand of the PFS-R found items in the sensory subscale required a 3rd grade education or higher. The most demanding items were in the behavioral subscale and required an 8th grade education or higher (with one item that required a post-high school education). This more demanding item was kept in the PFS-12 because it captures a broader range of enjoyable activities and could be applied to experiences alone or with a friend or partner. Future versions of the PFS-R or PFS-12 should consider revising this item to read, “To what degree is the fatigue you are now feeling interfering with your ability to do the activities you enjoy?” This simple modification does not change the item content, but drops the literacy demand to a 9th grade education level. Additional cognitive testing of this revised item is recommended.

This study expands on findings from a published study with the same breast cancer cohort, which found fatigue, measured by the PFS-R, was associated with poorer HRQOL [6]. Specifically, fatigue was associated with pain, cognitive problems, antidepressant use, weight gain/personal appearance, and physical inactivity [6]. Because this previous study provided evidence for the validity of the PFS-R in terms of its association with other related clinical and HRQOL factors, these analyses were not duplicated in this study.

This study had limitations. First, though this psychometric analysis used a population-based sample of breast cancer survivors, results cannot be generalized to other survivor groups or health conditions. In addition, race comparisons are not based on women equally recruited across sites. The Los Angeles site collected all African-American data and New Mexico collected a majority of the

Hispanics. Thus, the differential item functioning results by race could have been associated with geographic differences [15]. In addition, eligible African-American women were restricted to 35–64 years of age; thus age differences could be another source for differential item functioning findings. Restricting whites to the same age range would not have yielded enough sample sizes for differential item functioning testing.

It is therefore recommended that additional analyses of datasets that included the PFS-R be done to confirm these findings, including datasets that used a 7-day reference period instead of the 30-day reference period used in this study. The psychometric properties of the PFS-12 also need reconfirmation in new samples, so the items are presented next to each other instead of dispersed across the larger PFS-R. Evaluations in women undergoing active treatment and in other cancer populations using both men and women also are recommended.

Despite these limitations, the brief and reliable PFS-12 should have great value in future patient-centered outcomes research studies to capture the multidimensional aspects of the fatigue experience. Additional research is planned to identify clinically meaningful cut-points on the PFS-12 that classify individuals as mildly, moderately, or severely fatigued and to characterize the association between these cut-points with HRQOL decrements. Together, a psychometrically sound, and decision-relevant fatigue measure will enable researchers to provide empirical evidence about the impact of interventions on survivors' lives that may lead to identifying safer and more effective treatments.

Acknowledgments Dr. Reeve's work was supported under a National Cancer Institute contract: HHSN261201000642P.

Conflict of interest Barbara Piper is the developer of the scale we analyzed. There are no other conflicts of interest to report.

Appendix

Piper Fatigue Scale-12 (PFS-12)

Directions: Please circle the number which best describes the fatigue you are experiencing **IN THE PAST 4 WEEKS.**

- To what degree is the fatigue you are feeling interfering with your ability to complete your work or school activities?

0	1	2	3	4	5	6	7	8	9	10
None										A great deal
- Overall, how much is the fatigue which you are experiencing interfering with your ability to engage in the kind of activities you enjoy?

0	1	2	3	4	5	6	7	8	9	10
None										A great deal

ALTERNATIVE WORDING for #2.
*NOTE: THIS ITEM HAS NOT BEEN TESTED.

To what degree is the fatigue you are feeling interfering with your ability to do the activities you enjoy?

0	1	2	3	4	5	6	7	8	9	10
None										A great deal

- How would you describe the degree of intensity or severity of the fatigue which you are experiencing?

0	1	2	3	4	5	6	7	8	9	10
Mild										Severe
- To what degree would you describe the fatigue which you are experiencing as being:

0	1	2	3	4	5	6	7	8	9	10
Pleasant										Unpleasant
- To what degree would you describe the fatigue which you are experiencing as being:

0	1	2	3	4	5	6	7	8	9	10
Positive										Negative
- To what degree would you describe the fatigue which you are experiencing as being:

0	1	2	3	4	5	6	7	8	9	10
Normal										Abnormal
- To what degree are you feeling:

0	1	2	3	4	5	6	7	8	9	10
Strong										Weak
- To what degree are you feeling:

0	1	2	3	4	5	6	7	8	9	10
Awake										Sleepy
- To what degree are you feeling:

0	1	2	3	4	5	6	7	8	9	10
Refreshed										Tired
- To what degree are you feeling:

0	1	2	3	4	5	6	7	8	9	10
Patient										Impatient
- To what degree are you feeling:

0	1	2	3	4	5	6	7	8	9	10
Able to concentrate										Unable to concentrate
- To what degree are you feeling:

0	1	2	3	4	5	6	7	8	9	10
Able to think clearly										Unable to think clearly

References

- Andrykowski MA, Curran SL, Lightner R (1998) Off-treatment fatigue in breast cancer survivors: a controlled comparison. *J Behav Med* 21(1):1–18
- Barsevick AM, Cleeland CS, Manning DC, O'Mara AM, Reeve BB, Scott JA, Sloan JA, Ascpo (2010) ASCPRO recommendations for the assessment of fatigue as an outcome in clinical trials.

- J Pain Symptom Manage 39(6):1086–1099. doi:[10.1016/j.jpainsymman.2010.02.006](https://doi.org/10.1016/j.jpainsymman.2010.02.006)
3. Piper BF, Cella D (2010) Cancer-related fatigue: definitions and clinical subtypes. *J Natl Compr Cancer Netw* 8(8):958–966
 4. Esther Kim JE, Dodd MJ, Aouizerat BE, Jahan T, Miaskowski C (2009) A review of the prevalence and impact of multiple symptoms in oncology patients. *J Pain Symptom Manage* 37(4):715–736. doi:[10.1016/j.jpainsymman.2008.04.018](https://doi.org/10.1016/j.jpainsymman.2008.04.018)
 5. Bower JE, Ganz PA, Desmond KA, Rowland JH, Meyerowitz BE, Belin TR (2000) Fatigue in breast cancer survivors: occurrence, correlates, and impact on quality of life. *J Clin Oncol* 18(4):743–753
 6. Meeske K, Smith AW, Alfano CM, McGregor BA, McTiernan A, Baumgartner KB, Malone KE, Reeve BB, Ballard-Barbash R, Bernstein L (2007) Fatigue in breast cancer survivors two to five years post diagnosis: a HEAL Study report. *Qual Life Res* 16(6):947–960. doi:[10.1007/s11136-007-9215-3](https://doi.org/10.1007/s11136-007-9215-3)
 7. Dittner AJ, Wessely SC, Brown RG (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](https://doi.org/10.1016/S0022-3999(03)00371-4)
 8. Piper BF, Dibble SL, Dodd MJ, Weiss MC, Slaughter RE, Paul SM (1998) The revised Piper Fatigue Scale: psychometric evaluation in women with breast cancer. *Oncol Nurs Forum* 25(4):677–684
 9. Theodosopoulou H, Tzavelas G, Kotrosiou E, Raftopoulos V, Konstantinou E, Anatsiou N (2001) Psychometric properties of the revised Piper Fatigue Scale in Greek women with breast cancer. *Icus Nurs Web J* 8:1–15. doi:[1108-7366](https://doi.org/1108-7366)
 10. Gledhill JA, Rodary C, Mahe C, Laizet C (2002) French validation of the revised Piper Fatigue Scale. *Recherche en soins infirmiers* 68:50–65
 11. Dagnelie PC, Pijls-Johannesma MC, Pijpe A, Boumans BJ, Skrabanja AT, Lambin P, Kempen GI (2006) Psychometric properties of the revised Piper Fatigue Scale in Dutch cancer patients were satisfactory. *J Clin Epidemiol* 59(6):642–649. doi:[10.1016/j.jclinepi.2005.09.015](https://doi.org/10.1016/j.jclinepi.2005.09.015)
 12. Ostlund U, Gustavsson P, Furst CJ (2007) Translation and cultural adaptation of the Piper Fatigue Scale for use in Sweden. *Eur J Oncol Nurs* 11(2):133–140. doi:[10.1016/j.ejon.2006.04.034](https://doi.org/10.1016/j.ejon.2006.04.034)
 13. Mota DD, Pimenta CA, Piper BF (2009) Fatigue in Brazilian cancer patients, caregivers, and nursing students: a psychometric validation study of the Piper Fatigue Scale-Revised. *Support Care Cancer* 17(6):645–652. doi:[10.1007/s00520-008-0518-x](https://doi.org/10.1007/s00520-008-0518-x)
 14. McTiernan A, Rajan KB, Tworoger SS, Irwin M, Bernstein L, Baumgartner R, Gilliland F, Stanczyk FZ, Yasui Y, Ballard-Barbash R (2003) Adiposity and sex hormones in postmenopausal breast cancer survivors. *Journal of Clin Oncol* 21(10):1961–1966. doi:[10.1200/JCO.2003.07.057](https://doi.org/10.1200/JCO.2003.07.057)
 15. Bowen DJ, Alfano CM, McGregor BA, Kuniyuki A, Bernstein L, Meeske K, Baumgartner KB, Fetherolf J, Reeve BB, Smith AW, Ganz PA, McTiernan A, Barbash RB (2007) Possible socioeconomic and ethnic disparities in quality of life in a cohort of breast cancer survivors. *Breast Cancer Res Treat* 106(1):85–95. doi:[10.1007/s10549-006-9479-2](https://doi.org/10.1007/s10549-006-9479-2)
 16. Berger AM (1998) Patterns of fatigue and activity and rest during adjuvant breast cancer chemotherapy. *Oncol Nurs Forum* 25(1):51–62
 17. Woo B, Dibble SL, Piper BF, Keating SB, Weiss MC (1998) Differences in fatigue by treatment methods in women with breast cancer. *Oncol Nurs Forum* 25(5):915–920
 18. Wu HS, McSweeney M (2001) Measurement of fatigue in people with cancer. *Oncology Nurs Forum* 28(9):1371–1384; quiz 1385–1376
 19. Reise SP, Morizot J, Hays RD (2007) The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res* 16(Suppl 1):19–31. doi:[10.1007/s11136-007-9183-7](https://doi.org/10.1007/s11136-007-9183-7)
 20. Bentler PM (1990) Comparative fit indexes in structural models. *Psychol Bull* 107(2):238–246
 21. West SG, Finch JF, Curran PJ (1995) SEM with nonnormal variables. In: Hoyle RH (ed) *Structural equation modeling: concepts issues and applications*. Sage Publications, Thousand Oaks, CA
 22. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6(1):1–55. doi:[10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
 23. Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA, Long JS (eds) *Testing structural equation models*. Sage Publications, Newbury Park, CA
 24. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo FDAP-ROCMG (2007) What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 10(Suppl 2):S94–S105. doi:[10.1111/j.1524-4733.2007.00272.x](https://doi.org/10.1111/j.1524-4733.2007.00272.x)
 25. Samejima F (1969) Estimation of latent ability using a response pattern of graded scores. *Psychometrika* 34(4p2):1
 26. Samejima F (1997) Graded response model. In: van der Linden WJ, Hambleton RK (eds) *Handbook of modern item response theory*. Springer, New York, pp 85–100
 27. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D, Group PC (2007) Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 45(5 Suppl 1):S22–S31. doi:[10.1097/01.mlr.0000250483.85507.04](https://doi.org/10.1097/01.mlr.0000250483.85507.04)
 28. Chen WH, Thissen D (1997) Local dependence indexes for item pairs: using item response theory. *J Educ Behav Stat* 22(3):265–289
 29. Nunnally JC, Bernstein IH (1994) *Psychometric theory*, 3rd edn. McGraw-Hill, New York
 30. MetaMetrics, Inc. (2008) What does the Lexile measure mean? http://d1jt5u2s0h3gkt.cloudfront.net/m/cms_page_media/123/What-does-the-Lexile-measure-mean_1.pdf. Accessed 1 May 2012
 31. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, Group PC (2011) Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. *Assessment* 18(3):263–283. doi:[10.1177/1073191111411667](https://doi.org/10.1177/1073191111411667)
 32. Institute of Medicine (2004) *Health literacy: a prescription to end confusion*. The National Academies Press, Washington, DC
 33. Yen WM (1993) Scaling performance assessments—strategies for managing local item dependence. *J Educ Meas* 30(3):187–213
 34. Annunziata MA, Muzzatti B, Mella S, Narciso D, Giacalone A, Fratino L, Tirelli U (2010) The revised Piper Fatigue Scale (PFS-R) for Italian cancer patients: a validation study. *Tumori* 96(2):276–281
 35. Clark PC, Ashford S, Burt R, Aycock DM, Kimble LP (2006) Factor analysis of the revised Piper Fatigue Scale in a caregiver sample. *J Nurs Meas* 14(2):71–78
 36. Alfano CM, Imayama I, Neuhaus ML, Kiecolt-Glaser JK, Smith AW, Meeske K, McTiernan A, Bernstein L, Baumgartner KB, Ulrich CM, Ballard-Barbash R (2012) Fatigue, inflammation, and omega-3 and omega-6 fatty acid intake among breast cancer survivors. *J Clin Oncol* 30(12):1280–1287. doi:[10.1200/JCO.2011.36.4109](https://doi.org/10.1200/JCO.2011.36.4109)