Supplementary Material

Section S1: Estimating σ and α_L from the cumulative link model with probit link

When the ordinal outcome (w) is a good proxy for the underlying continuous variable (y), we consider w as categorised from y by using the following thresholds: $1.5, \dots, j + 0.5, \dots, J - 0.5$, for $j = 1, \dots, J - 1$. The relationship between y and the latent continuous variable y^* from the CLM for w (see equation [2] in the main text) is described by a linear transformation function, $y = \alpha_L + \sigma y^*$, which is also applicable to the thresholds:

$$j + 0.5 = \alpha_L + \sigma\theta_j, \qquad \text{[S1]}$$

where θ_js are thresholds for y^*, for $j = 1, \dots, J - 1$.

To determine σ in the transformation function, we note that the thresholds for y are one-unit apart. Specifically, the difference between the j-th and $(j - 1)$-th thresholds is $(j + 0.5) - (j - 0.5) = 1$ for all $j = 2, \dots, J - 1$. Rewriting this difference using equation [S1] gives $(\alpha_L + \sigma\theta_j) - (\alpha_L + \sigma\theta_{j-1}) = 1$, or equivalently $\theta_j - \theta_{j-1} = 1/\sigma$. Hence, when w is good proxy for y, the thresholds θ_js for y^* are equidistant, with $\Delta\theta = \theta_j - \theta_{j-1} = 1/\sigma$ for $j = 2, \dots, J - 1$. Hence, $\sigma = 1/\Delta\theta$, and the shifting factor of the transformation (α_L) can be derived from θ_js by considering $j = 1$ in equation [S1]: $\alpha_L = 1.5 - \sigma\theta_1$ to ensure the smallest threshold for y is 1.5.

Equation [S1] can be rearranged to express θ_js as a linear function of the j index:

$$\theta_j = (0.5 - \alpha_L)/\sigma + j/\sigma. \qquad \text{[S2]}$$

Equation [S2] suggests that with $\hat{\theta}_j$s obtained from the CLM for w, $1/\sigma$ can be estimated via the slope of a simple linear regression model where $\hat{\theta}_j$s are the outcomes and the j index is the predictor. The estimate $\hat{\sigma}$ is subsequently obtained via the reciprocal of the estimated slope.

Section S2: Supplementary tables and figures

Table S1 Characteristics of breast cancer survivors

N = 316	
Scores of MFI subscales[a]: mean (SD)	
General Fatigue	10.1 (3.4)
Physical Fatigue	10.3 (3.6)
Mental Fatigue	8.6 (3.3)
Reduced Activity	9.3 (3.3)
Reduced Motivation	9.0 (3.0)
Years since diagnosis: mean (SD)	6.5 (4.6)
Age: mean (SD)	50.5 (8.9)
Ethnicity: n (%)	
Chinese	247 (78.2)
Non-Chinese	69 (21.8)
Working status: n (%)	
Full-time/Part-time	155 (49.1)
Not working	161 (50.9)
Stage: n (%)	
DCIS	45 (14.2)
I or II	236 (74.7)
III or IV	35 (11.1)
Type of surgery: n (%)	
BSC	126 (39.9)
Mastectomy	190 (60.1)
Used chemotherapy: n (%)	
Yes	204 (64.6)
No	112 (35.4)

a: The 15th to 17th categories of the original score were grouped into a new category and assigned a score of 18.

Abbreviation: BCS: breast-conserving surgery; DCIS: ductal carcinoma in situ; MFI: Multidimensional Fatigue Inventory; SD: standard deviation.

Table S2 Estimated difference in the mean score per year since breast cancer diagnosis when analyzing the original score of General Fatigue using the cumulative link model (CLM) with probit link and the linear regression model (LRM)

		Exposure effect (β) from CLM		Difference in mean scores (β_L)		P-value of likelihood ratio test for equidistant thresholds
Outcome	**Method**	**Estimate**	**95% Confidence interval**	**Estimate**	**95% Confidence interval**	
General Fatigue	CLM	-0.044	-0.071, -0.017	-0.153	-0.246, -0.060	0.930
	LRM			-0.140	-0.227, -0.053	

Abbreviations: CLM, cumulative link model; LRM, linear regression model.

Figure S1 An illustration of thresholds used in the simulation study.

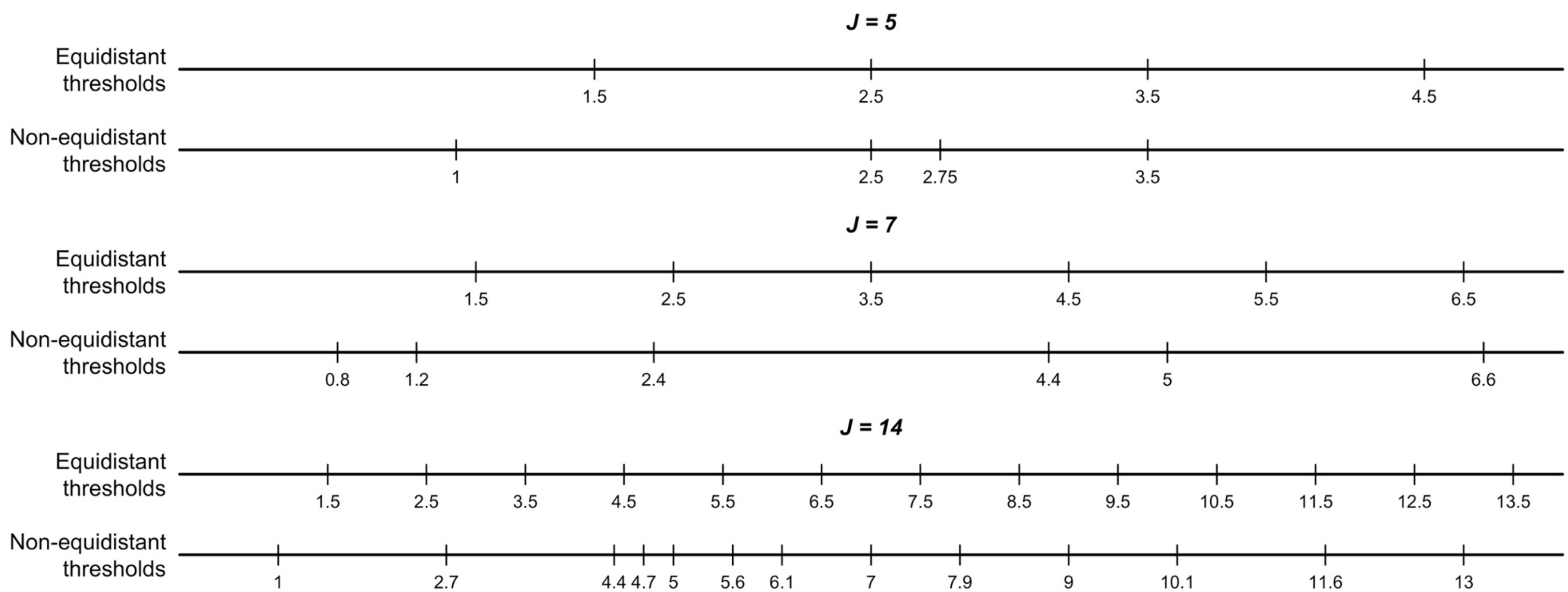
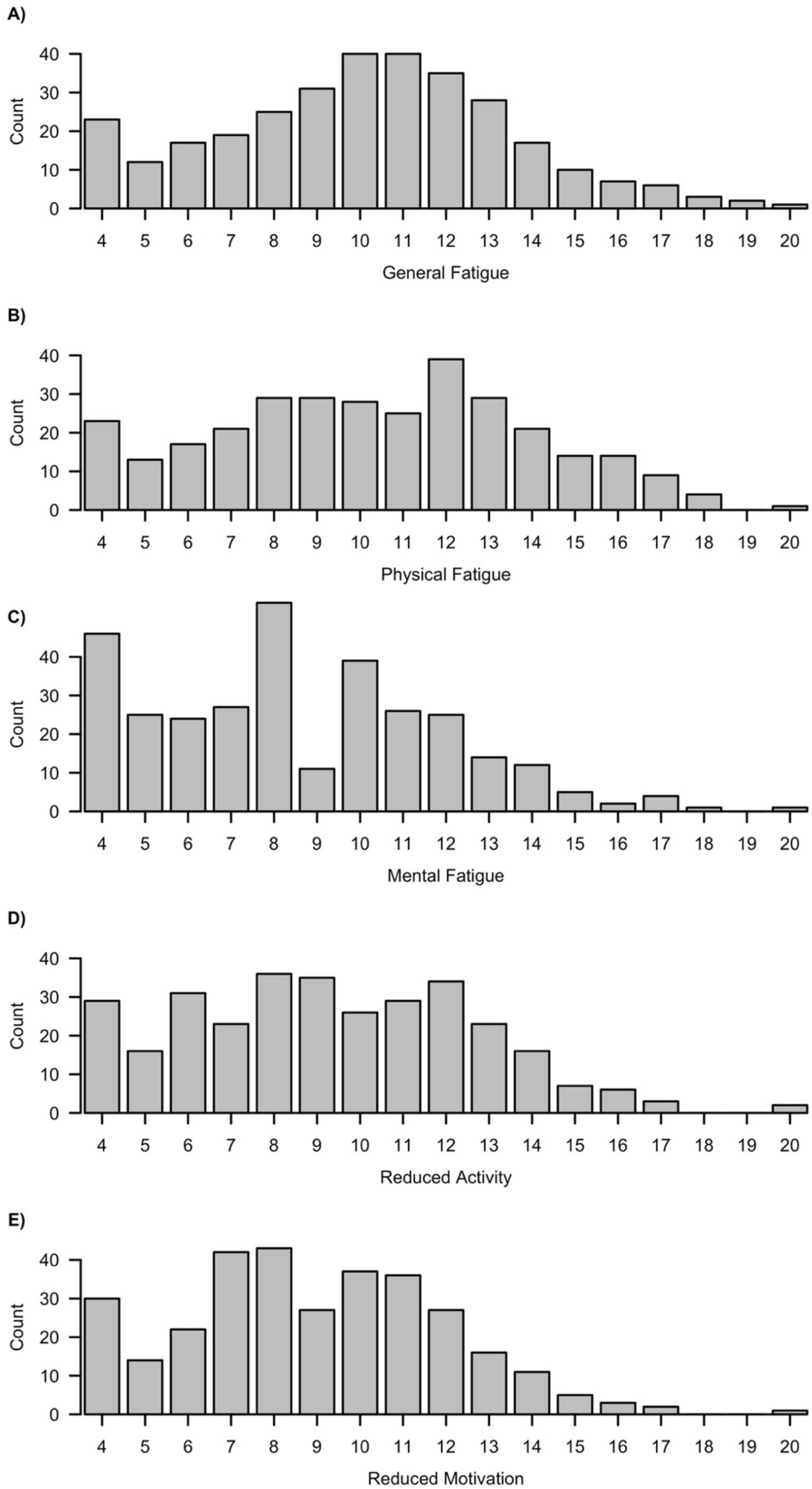


Figure S2 Frequency distribution of the ordinal categories in the five subscales of Multidimensional Fatigue Inventory: General Fatigue (panel A), Physical Fatigue (panel B), Mental Fatigue (panel C), Reduced Activity (panel D) and Reduced Motivation (panel E).



Section S3: Supplementary simulation study with less extreme non-equidistant thresholds

Using the underlying continuous variable y generated for ordinal outcomes with non-zero effect, we grouped y into ordinal categories by considering thresholds with smaller deviations from equidistance than the non-equidistant thresholds used in the main simulation analyses (ie, simulations presented in the main manuscript). Specifically, the thresholds took values, (i) 1.345, 2.549, 3.544 and 4.550 for $J = 5$, (ii) 1.345, 2.549, 3.544, 4.549, 5.644 and 6.556 for $J = 7$, and (iii) 1.571, 2.481, 3.434, 4.588, 5.524, 6.407, 7.383, 8.626, 9.533, 10.512, 11.388, 12.448 and 13.565 for $J = 14$. These thresholds were generated by adding random values to the equidistant thresholds (ie, $1.5, 2.5, \ldots, J - 0.5$), which were generated between -0.2 and 0.2 when $J = 5, 7$ and were between -0.15 and 0.15 when $J = 14$. Ordinal outcomes generated using these thresholds were analyzed using our proposed CLM approach and the LRM, and the results are summarized in Table S3 (see rows indicated with "Non-equidistant, supplemental"). We investigated the impact of the smaller deviations from equidistance by comparing these results with those from the main simulation analyses that considered equidistant thresholds (see rows in Table S3 indicated with "Equidistant, main") and non-equidistant thresholds with considerable deviations from equidistance (see rows indicated with "Non-equidistant, main").

When analyzing simulated outcomes using the non-equidistant thresholds specified in this section, the percent of simulation cycles where the p-value of the likelihood ratio test for equidistant thresholds was less than 0.05 ranged between 18.5% and 24.5% when $n = 300$, and ranged between 50.8% and 68.5% when $n = 1000$. These findings reflect the lower power that would be expected for the likelihood ratio test for equidistant thresholds in scenarios with a smaller deviation from equidistance and/or a smaller sample size, when compared with the non-equidistant thresholds specified in the main manuscript where power was 100%. The magnitude of bias of the CLM estimates in this supplementary study (ie, "Non-equidistant, supplemental") was smaller than the non-equidistant thresholds reported in the main manuscript (ie, "Non-equidistant, main") and larger than the equidistant thresholds reported in the main manuscript (ie, “Equidistant, main”), but the coverage was close to 95%. When the LRM was applied to the supplementary data sets (ie, “Non-equidistant, supplemental”), the magnitude of its bias was smaller than the non-equidistant thresholds reported in the main manuscript (ie, “Non-equidistant, main”) and larger than the equidistant thresholds reported in the main manuscript (ie, “Equidistant, main”) for $J = 7$, and its coverage was better than the non-equidistant thresholds reported in the main manuscript (ie, “Non-equidistant, main”) and poorer than the equidistant thresholds reported in the main manuscript (ie, “Equidistant, main”) for $J = 7$. For $J = 5, 14$, the magnitude of its bias was either similar or larger than those in the main manuscript, and its coverage was either similar or smaller than those in the main manuscript.

In summary, smaller deviations from equidistance can have an adverse impact on the performance of CLM and LRM. Future work should be conducted to better understand how deviations from equidistance impact the performance of LRM.

Table S3 Mean, standard error (SE) and coverage of the estimated difference in mean ordinal scores from the cumulative link model (CLM) with probit link and linear regression model (LRM) when analyzing ordinal outcome generated with non-equidistant thresholds (ie, "Non-equidistant, supplemental" corresponds to supplementary simulation study where the thresholds had smaller deviations than those presented in the main manuscript: "Non-equidistant, main") or equidistant thresholds in the main manuscript (ie, "Equidistant. main"), and non-zero effects, with varying number of categories ($J = 5, 7, 14$) and sample sizes ($n = 300, 1200$).

Method	Number of ordinal categories	n	Increasing deviation from equidistant thresholds	Bias	Empirical SE	Mean SE	Coverage (%)	Simulation cycles with equidistant thresholds assumption rejected (%)
CLM	$J = 5$, true effect is 0.5	300	Equidistant, main	0.002	0.158	0.158	94.9	4.7
			Non-equidistant, supplemental	-0.028	0.148	0.149	94.3	24.5
			Non-equidistant, main	0.151	0.205	0.207	89.7	100
		1000	Equidistant, main	0.001	0.082	0.081	94.7	5.9
			Non-equidistant, supplemental	-0.027	0.077	0.076	93.7	68.5
			Non-equidistant, main	0.147	0.107	0.106	72.0	100
	$J = 7$, true effect is 0.7	300	Equidistant, main	0.004	0.215	0.216	95.0	5.2
			Non-equidistant, supplemental	-0.023	0.207	0.208	94.6	18.5
			Non-equidistant, main	-0.120	0.180	0.180	89.4	100
		1000	Equidistant, main	0.002	0.112	0.110	94.5	5.5
			Non-equidistant, supplemental	-0.024	0.108	0.106	94.3	50.8
			Non-equidistant, main	-0.121	0.092	0.092	73.1	100
	$J = 14$, true effect is 1.4	300	Equidistant, main	0.006	0.420	0.424	95.3	5.7
			Non-equidistant, supplemental	0.010	0.420	0.425	95.2	21.8
			Non-equidistant, main	0.210	0.483	0.486	92.8	100
		1000	Equidistant, main	0.004	0.218	0.217	94.7	5.0
			Non-equidistant, supplemental	0.008	0.219	0.217	94.7	66.2
			Non-equidistant, main	0.206	0.251	0.248	86.4	100
LRM	$J = 5$, true effect is 0.5	300	Equidistant, main	-0.067	0.134	0.136	91.5	Not available
			Non-equidistant, supplemental	-0.085	0.128	0.130	89.8	Not available
			Non-equidistant, main	0.005	0.158	0.163	96.0	Not available
		1000	Equidistant, main	-0.066	0.070	0.069	83.3	Not available
			Non-equidistant, supplemental	-0.084	0.067	0.066	74.1	Not available
			Non-equidistant, main	0.004	0.082	0.083	95.1	Not available
	$J = 7$, true effect is 0.7	300	Equidistant, main	-0.068	0.190	0.193	93.6	Not available
			Non-equidistant, supplemental	-0.087	0.185	0.188	92.9	Not available
			Non-equidistant, main	-0.167	0.165	0.165	82.4	Not available
		1000	Equidistant, main	-0.069	0.100	0.098	89.1	Not available
			Non-equidistant, supplemental	-0.087	0.097	0.096	83.5	Not available
			Non-equidistant, main	-0.169	0.085	0.084	48.1	Not available
	$J = 14$, true effect is 1.4	300	Equidistant, main	-0.093	0.386	0.393	94.5	Not available
			Non-equidistant, supplemental	-0.091	0.386	0.393	94.3	Not available
			Non-equidistant, main	0.053	0.440	0.446	94.9	Not available
		1000	Equidistant, main	-0.095	0.201	0.200	92.8	Not available
			Non-equidistant, supplemental	-0.093	0.202	0.200	93.1	Not available
			Non-equidistant, main	0.049	0.229	0.227	93.8	Not available

Abbreviations: CLM, cumulative link model; LRM, linear regression model; SE: standard error.