Applied Research in Quality of Life (2023) 18:697–707
https://doi.org/10.1007/s11482-022-10103-6

Comparison of Simple-Summated Scoring and Toxicity Index Scoring of Symptom Bother in the NSABP B-30 Clinical Trial

Ron D. Hays[1,2] · Gillian Gresham[3,4] · Patricia A. Ganz[2,5] · Mourad Tighiouart[3]

Received: 19 July 2022 / Accepted: 26 August 2022 / Published online: 9 September 2022

Abstract

Background Level of symptom burden for cancer patients can be summarized using simple-summated scoring of multiple patient-reported symptoms. The Toxicity Index (TI) is an alternative that has been used primarily to summarize clinician-reported toxicities.

Objective To compare the TI with simple-summated scoring of 28 patient-reported symptoms.

Methods This is a secondary analysis of longitudinal data from a clinical trial of women with stage 2 or 3 breast cancer: baseline (n = 2156) and 6 months later (n = 1764). Study participants completed the 28-item Breast Cancer Prevention Trial symptom checklist assessing level of symptom bother in the past 7 days and four criterion items assessing general health and overall quality of life.

Results Associations of simple-summated scoring of the 28 cancer-related symptoms with the general health and overall quality of life items tended to be larger than correlations of the TI summary scoring of the symptoms. For example, the Spearman correlation of change in quality of life was −0.38 with change in the simple-summated score and −0.23 with change in the TI.

Conclusion The findings suggest that simple-summated scoring and differential weighting of the level of symptom bother yield similar results.

Implications for Practice Clinicians can use simple-summated scoring rather than more complicated scoring algorithms to obtain an indication of overall level of symptom burden among cancer patients.

Keywords symptoms · toxicity index · simple summation · cancer

A symptom is a "patient's perception of an abnormal physical, emotional, or cognitive state" (Wilson & Cleary, 1995,p. 61) The Assessing the Symptoms of Cancer

Extended author information available on the last page of the article

Using Patient-reported Outcomes (ASCPRO) working group asserted that symptom burden is the patient-reported counterpart to disease and tumor burden (Cleeland & Sloan, 2010). The ASCPRO group also noted that more stable and informative self-reports of symptoms are obtained by asking about symptom severity rather than simply present or absent of symptoms.

The National Cancer Institute (NCI) Symptom Management and Health-Related Quality of Life Steering Committee recommended a core set of 12 symptoms be "assessed across oncology trials for the purposes of better understanding treatment efficacy and toxicity and to facilitate cross study comparisons" (p. 1): fatigue, insomnia, pain, anorexia, dyspnea, cognitive problems, anxiety, nausea, depression, sensory neuropathy, constipation, and diarrhea (Reeve et al., 2014). Each symptom can be examined separately, but there is often interest in combining different symptoms into a summary score (McColl, 2004). The NCI committee did not recommend how level of symptom burden should be summarized. Often, symptom items are summed or averaged (Niu et al., 2021).

For example, the "symptom distress" scale is the sum of responses to 13 symptoms: nausea, appetite, insomnia, pain, fatigue, bowel pattern, concentration, appearance, breathing, outlook, and cough (Stapleton et al., 2016). While "there are some instances where sum scores are justified; the problem … is employing methods without any justification" (McNeish & Wolf, 2020, p. 2301). More generally, Ferrando & Lorenzo-Seva (2020, p. 207)commented that "all the models and scoring schemas are approximations, and the key issue is to decide whether the approximation is good enough for the purposes of the study."

The Toxicity Index (TI) is an alternative to simple summation (Razaee et al., 2021). The TI was inspired by hash functions and summarizes all n observed toxicity grades as defined in the Common Terminology Criteria for Adverse Events (CTCAE). Each of the n toxicity grades $X\,i$ ($i=1, \ldots, n$) for an individual is represented in descending order: $X_1 \geq X_2 \geq \ldots \geq X_n$. An individual's TI score is a function of the ordered toxicity grades. Any TI ≥ 3 corresponds to a dose-limiting toxicity, and the maximum toxicity grade is the integer part of the final score. For example, a TI of 3.0 indicates a single grade 3 toxic event, whereas a TI of 3.5 means that the patient experienced at least 1 grade 3 toxic event plus additional toxic events. All toxicity grades are represented in the score, although lower grades contribute less to the final score than higher grades. The TI has been used to summarize clinician-reported toxicities from the CTCAE (Henry et al., 2021) and may also be applied to and useful for summarizing patient-reported symptom measures (Hays et al., 2021).

Simple-summated scoring and the TI represent conceptually distinct scoring approaches. The former treats each symptom as equivalent in its contribution to the total score (i.e., each additional symptom has equal input on the simple sum). The TI heavily weights having one bothersome (i.e., *very much* bothered) symptom and each additional symptom has decreasing input. If the 28 items are scored from 0 (*not at all*) to 4 (*very much*), as shown in the Appendix, the sum for a patient who reports that they are *very much* bothered by one symptom but *not at all bothered* by the 27 other symptoms is 4 on the 0-112 possible range (average score = 0.14 on 0–4 possible range). The TI for a person reporting being *not at all* bothered for 27 symptoms

and *very much* bothered by one symptom is much higher: 4.00 on the 0-4.99 possible range.

Simple-summated scoring of symptoms has not yet been compared to the TI. This paper provides an initial comparison using secondary analyses of patient-reported data collected in the NSABP B-30 cancer clinical trial.

Methods

Sample

Patient-reported data in the NSABP B-30 clinical trial (Swain et al., 2010) at baseline (n=2156) and 6 months later (n=1764) were analyzed. Study participants had stage 2 or 3 breast adenocarcinoma and had undergone primary surgery with lumpectomy or mastectomy and axillary lymph node dissection. Combinations of doxorubicin (A), cyclophosphamide (C), and docetaxel (T) were evaluated by randomizing participants to either: 1) sequential A 60mg/m^2 plus C 600mg/m^2 every 3 weeks for four cycles followed by T 100mg/m^2 every 3 weeks for four cycles; 3) concurrent A 60mg/m^2, C 600mg/m^2, plus T 60mg/m^2 every 3 weeks for four cycles; or A 60mg/m^2 plus T 60mg/m^2 every 4 weeks for four cycles.

Measures

Study participants completed questionnaires to assess symptoms before chemotherapy, at cycle 4day 1 of chemotherapy, at 6 months, and then every 6 months through 24 months following chemotherapy initiation. A 28-item Breast Cancer Prevention Trial symptom checklist with five response options (*Not at All, A little bit, Somewhat, Quite a Bit, Very Much*) assessing symptom bother in the past 7 days was administered (Ganz et al., 2011).

Criterion Measures

Four self-report items hypothesized to be associated with level of symptom burden were also included on the same survey as the symptom items:

1. In general, would you say that your health is *Excellent, Very good, Good, Fair*, or *Poor*.
2. In the past 7 days, I feel ill (*Not at All, A little bit, Somewhat, Quite a Bit, Very Much*).
3. In the past 7 days, I am bothered by side effects of treatment (*Not at All, A little bit, Somewhat, Quite a Bit, Very Much)*.
4. Please score your overall quality of life in the past 7 days on an 11-point scale between *dead* (0) and *perfect health* (10).

The first is the most widely used self-reported health item (Hays et al., 2015), the second is a FACT-G item (Cella et al., 1993), the third is the "GP5" item (Pearman et al., 2018), and the fourth item is a global quality of life item (Diehr et al., 2007).

We scored all four criterion items so that a higher score represented better health. We also created a 3-item general health scale score by summing responses to the first three items (Cronbach, 1951, coefficient alpha = 0.75 at the 6-month follow-up; product-moment correlations among the items ranged from 0.46 to 0.57). We hypothesized that the symptom bother simple sum and the TI would be positively and monotonically related to the criterion items and the general health scale.

Analyses

Analyses were conducted for those with complete data on the 28 symptom items (88% of the baseline sample). We estimated item-scale correlations and internal consistency reliability for the symptom items. We also conducted categorical exploratory factor analysis with diagonally weighted least squares to evaluate the dimensionality of the symptom items, following by categorical confirmatory factor analysis. We evaluated model fit using the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). CFI > 0.95 and RMSEA < 0.06 are indicators of excellent model fit (Hu & Bentler, 1999).

Because the TI is a rank-order index, for comparability we estimate Spearman correlations of the TI and simple-summated symptom score at the 6-month follow-up with the 3-item general health scale and the 4 criterion items at 6-months and their change from baseline to 6 months later. We also estimated Spearman correlations for change in the TI and simple-summated score with change in the general health scale and criterion items. Pearson-product moment correlations tended to be similar or not as large for the TI and similar or larger for the simple sum (results not shown). The significance of difference in the dependent rank-order correlations was estimated using z-statistics. (Steiger, 1980).

We regressed (ordinary least squares) change on each of the criterion items on TI at baseline, TI at follow-up, simple sum at baseline, and simple sum at follow-up. Next, we ran three separate Probit regression models for each criterion item, controlling for their baseline values. The first model included the TI at baseline, the second model the simple-summated symptom score, and the third model included both the TI and the simple-summated score. Finally, we regressed (ordinary least squares) the 3-item health scale on dummy variables representing the levels of each symptom item.

Analyses were conducted using SAS 9.4 (TS Level 1MG).

Results

Internal consistency reliability for the 28-item symptom scale was 0.83 and item-scale correlations ranged from 0.13 (vaginal bleeding) to 0.55 (general aches and pains). The exploratory factor analysis suggested two possible underlying factors. The first factor was represented by 23 symptom items and the other factor by 5 items

that represented vasomotor and vaginal symptoms: night sweats, hot flashes, cold sweats, vaginal dryness, pain with intercourse). The 23-item general symptom factor correlated 0.484 with the vasomotor/vaginal symptom factor.

The one-factor confirmatory model did not fit the data (RMSEA=0.077, CFI=0.713). The two-factor confirmatory model fit the data better than the one-factor model (RMSEA=0.054, CFI=0.857) with the general 23-item symptom factor correlated 0.465 with the vasomotor/vaginal symptom factor. The fit of the two-factor model improved after adding correlated residuals (RMSEA=0.037, CFI=0.934). A bifactor model also fit the data reasonably well (RMSEA=0.042, CFI=0.921). The bifactor model revealed larger loadings of the 5 vasomotor/vaginal items on their group factor than on the general factor. Because the pattern of results reported below was similar for all 28 symptoms and the 23 symptoms that loaded on the first factor (results available upon request), we only report TI and simple sum results based on all 28 items.

Spearman correlations between the TI and the simple-summated total symptom score were 0.90 at baseline, 0.86 at follow-up, and 0.66 for change between baseline and follow-up. Table1 presents Spearman correlations of the 6-month TI and simple-summated symptom score with the 6-month general health scale and the 4 criterion items, and with change in the criterion measures from baseline to 6-months later. The simple sum score had similar but consistently stronger correlations than did the TI with the general health scale and criterion items. Spearman correlations of change in the TI and the simple sum with change in the 4 criterion items are provided in Table2. The simple-summated score for the 28 symptom items was consistently more highly correlated than was the TI with change in these criteria (7 of 10 comparisons in Table1 and all 6 of those in Table2).

In the Probit regression models that included either TI or the simple-sum, the standardized coefficients were similar for both, and all p-values were highly significant ($p < .0001$). In the models that included both TI and the simple sum, the TI was significant for the general health scale while the simple sum was significant for the *not* feeling ill and *not bothered* by side effects items (Table3).

The symptom levels with statistically significant associations with the 3-item general health scale at baseline are shown in Table4. As expected, increasing severe levels of symptoms tended to be associated with less positive self-rated health. For example, compared to those *not at all* bothered, the unique decrement in self-rated health was −1.97, -1.92, -3.30 and −5.40 for those reporting they had been bothered by headaches *a little bit*, *somewhat*, *quite a bit*, and *very much* in the past 7 days. The strongest association was seen for those reporting *quite a bit* of bother from vomiting (-27.81 on the 0-100 possible range). Fifteen of the symptoms (mouth sores, diarrhea, difficulty with bladder control, constipation, hot flashes, vaginal discharge, vaginal bleeding or spotting, vaginal dryness, pain with intercourse, cramps, swelling of hands or feet, weight gain, forgetfulness, night sweats, cold sweats) did not have significant unique associations with self-rated health.

Table 1 Spearman Correlations (n = 1085) of 6-month Toxicity Index (TI) and 6-month Simple-summated Symptom Scores with 6-month and Change in the Health Scale and the Criterion Items

Criterion	TI at 6-month follow-up	Simple sum at 6-month follow-up	z-statistic	p-value
Health scale at follow-up	-0.53	-0.61	6.17	< 0.001
Change in health scale	-0.35	-0.38	2.01	0.044
General health item at follow-up	-0.41	-0.48	4.90	< 0.001
Change in general health item	-0.25	-0.26	0.64	0.519
Ill item at follow-up	-0.41	-0.50	6.35	< 0.001
Change in ill item	-0.26	-0.29	1.95	0.051
Side effects item at follow-up	-0.47	-0.52	3.62	< 0.001
Change in side effects item	-0.30	-0.32	1.31	0.189
Quality of Life item at follow-up	-0.41	-0.50	6.35	< 0.001
Change in Quality of Life item	-0.24	-0.29	3.23	0.001

Health scale is the average of the general health, ill, and side effects items. All p-values in columns 2–3 p < .0001. z-statistic tests whether the difference between dependent correlations in each row is zero: Hoerger, M. (2013). Z_H: An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from http://www.psychmike.com/dependent_correlations.php

Discussion

The direction and magnitude of associations of the TI and simple-summated scoring of the 28 cancer-related symptoms with the general health scale and single items assessing health and overall quality of life were similar. Monotonic associations of the levels of the symptom items were found in regressions of self-rated health on the symptom bother items. But several symptoms were not significantly uniquely associated with self-rated health and not all levels of symptoms were significantly different from *not at all* bothered for symptoms with significant associations. The differential magnitude of associations across symptoms at the same level of bother (e.g., *quite a bit* bothered by vomiting had a negative impact of -25.29 while being *quite a bit* bothered by headaches was −4.11) suggests that patients weigh symptoms differentially in terms of their impact on their health (Lee et al., 2012).

While differential weighting of symptoms and levels within symptoms might lead to a better summary of level of symptom burden, it has long been argued that "most attempts at differential item weighting show relatively little improvement over simple scoring" (Wainer, 1976, p. 216; quoting B.F. Green, 1975, personal communication). A more recent study showed that simple-sum scores "perform very well in terms of fidelity and the slight impact of nonlinearity and are quite stable under cross-validation" (Ferrando & Lorenzo-Seva, 2020, p. 223).

Table 2 Spearman Correlations (n = 1085) Between Change in Toxicity Index (TI) and Simple-summated Symptom Score from Baseline to 6-months Follow-up with Change in Criterion Items

Criterion	Change in TI	Change in Simple sum	z-statistic*
Change in health scale	-0.27	-0.46	12.81
Change in general health item	-0.20	-0.34	9.08
Change in ill item	-0.20	-0.36	10.42
Change in side effects item	-0.22	-0.38	10.49
Change in quality-of-Life item	-0.23	-0.38	9.85

*All p-values < 0.001

z-statistic tests whether the difference between dependent correlations in each row is zero: Hoerger, M. (2013). Z_H: An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from http://www.psychmike.com/dependent_correlations.php

Note: General health item: In general, would you say that your health is *excellent, very good, good, fair*, or *poor*. Ill item: In the past 7 days, I feel ill (*Not at All, A little bit, Somewhat, Quite a Bit, Very Much*); Side effects item: In the past 7 days, I am bothered by side effects of treatment (*Not at All, A little bit, Somewhat, Quite a Bit, Very Much*). Quality of life item: Please score your overall quality of life in the past 7 days on an 11-point scale between dead (0) and perfect health (10).

Table 3 Probit Regression of Criterion Items at the 6-Month Follow-up on Toxicity Index (TI) and Simple-summated Symptom Score at Baseline, Controlling for the Criterion at Baseline

	General health	Quality of Life	Not Feeling ill	Not Bothered by side effects
Model 1				
TI at baseline	−0.20 (<0.0001)	−0.15 (<0.0001)	−0.21 (<0.0001)	−0.16 (<0.0001)
Model 2				
Simple-summated score at baseline	−0.19 (<0.0001)	−0.15 (<0.0001)	−0.23 (<0.0001)	−0.17 (<0.0001)
Model 3				
TI	−0.16 (0.0017)	NS	NS	NS
Simple-summated score	NS	NS	−0.11 (0.0008)	−0.11 (0.0406)

Note: All models control for baseline score on the Criterion. Standardized estimates are shown (p-values within parentheses). NS = not significant.

Alternative approaches include weighting different subscales to obtain a summary score. For example, the Kidney Disease Quality of Life (KDQOL) instrument has a summary score that combines scores from a 12-item symptoms/problems scale, 8-item effects of kidney disease scale, and a 4-item burden of kidney disease scale by weighting mean scores by the number of items per domain (Peipert et al., 2019).

The 11 respondents who reported they were *very much* bothered on 29% (8 of the 28) of the symptom items had a rounded TI score of 5.00000 (the TI never reaches

Table 4 Ordinary least squares regression of 3-item Health Scale on Symptom Levels at Baseline (Coefficients significant at $p < .05$ shown)

Symptom Level	Estimate	Probability
Headache (1)	-1.97	0.0044
Headache (2)	-1.92	0.0390
Headache (3)	-3.30	0.0276
Headache (4)	-5.40	0.0443
Vomiting (1)	-5.37	0.0038
Vomiting (3)	-27.81	<0.0001
Numbness or tingling in hands or feet (1)	-2.18	0.0138
Numbness or tingling in hands or feet (2)	-3.51	0.0222
Fever or shivering (shaking, chills) (1)	-3.59	0.0012
Fever or shivering (shaking, chills) (2)	-7.26	0.0005
Mood swings (1)	-3.25	0.0003
Mood swings (3)	-3.28	0.0239
Mood swings (4)	-8.09	0.0001
Joint pains (1)	-1.93	0.0125
Joint pains (4)	-10.92	0.0009
Weight Loss (1)	-1.72	0.0383
Difficulty concentrating (1)	-1.56	0.0215
Difficulty concentrating (2)	-3.33	0.0008
Skin problems (including rash, dry skin, irritation, or redness) (2)	-2.77	0.0198
Genital itching or irritation (2)	-4.35	0.0083
Muscle stiffness (2)	-2.38	0.0212
Muscle stiffness (3)	-3.57	0.0436
General aches and pains (2)	-2.23	0.0239
General aches and pains (3)	-6.02	0.0003
Unhappy with appearance of my body (3)	-3.87	0.0067

Note. Health scale is the average of the general health, ill, and side effects items. Adjusted $R^2 = 0.25$. (1) *A little bit*; (2) *Somewhat*; (3) *Quite a bit*; (4) *Very much*

5 exactly). But the average of the 28 symptom items on the 0–4 possible range for these 11 patients ranged from 1.4 to 2.3 (mean = 1.78) and the range of 0–10 overall quality of life ratings was 3–8 (mean = 4.9, SD = 1.6). Moreover, there were 18 cases endorsing *very much* for more than 8 symptoms that also had a rounded TI score of 5.00000 and their average symptom scores ranged from 1.6 to 2.5 (mean = 2.04) and their overall quality of life ratings ranged from 3 to 9 (mean = 5.5, SD = 1.9).

The method used to combine items to produce a summary score needs to be justified for the specific context of use (U.S. Food and Drug Administration, 2009). Future research may help identify circumstances in which the present findings do not hold. The results of this study need replication in other datasets using different patient-reported symptom measures. It may also be useful to compare the summary approaches examined here with latent profile analysis (Whisenant et al., 2022). The TI was shown to have greater power than the maximum grade method to detect differences between side effects of treatments assessed using the CTCAE (Gresham et al., 2020). Because patient-reported symptoms include disease-related and treatment-related effects, it is not possible to generalize from this study to clinician reported

CTCAE data. Future studies comparing simple-summated scoring and the TI with the CTCAE may also be informative.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11482-022-10103-6.

Author Contributions The analysis and drafting of the manuscript was done by Ron D. Hays. All other authors commented on drafts and approved the final manuscript.

Declarations This work was supported by 1U01CA232859 as part of the NCI Cancer Moonshot. The authors have no competing interests to declare that are relevant to the content of this article.

References

Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., & Brannon, J. (1993). The functional assessment of cancer therapy scale: Development and validation of the general measure. *Journal of Clinical Oncology*, *11*(3), 570–579. doi: https://doi.org/10.1200/JCO.1993.11.3.570

Cleeland, C. S., & Sloan, J. A. (2010). Assessing the symptoms of cancer using patient-reported outcomes (ASCPRO): Searching for standards. *Journal of Pain and Symptom Management*, *39*(6), 1077–1085. doi: https://doi.org/10.1016/j.jpainsymman.2009.05.025

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334

Diehr, P., Lafferty, W. E., Patrick, D. L., et al. (2007). Quality of life at the end of life. *Health And Quality Of Life Outcomes*, *5*, 51

Ferrando, P. J., & Lorenzo-Seva, U. (2020). The appropriateness of sum scores as estimates of factor scores in the multiple factor analysis of ordered categorical responses. *Educational and Psychological Measurement*, *81*(2), 205–228

Ganz, P. A., Land, S. R., Geyer, C. E., Cecchini, R. S., Costantino, J. P., Pajon, E. R., Fehrenbacher, L., Atkins, J. N., Polikoff, J. A., Vogel, V. G., Erban, J. K., Livingston, R. B., Perez, E. A., Mamounas, E. P., Wolmark, N., & Swain, S. M. (2011). Menstrual history and quality-of-life outcomes in women with node-positive breast cancer treated with adjuvant therapy on the NSABP B-30 trial. *Journal of Clinical Oncology*, *29*(9), 1110–1116. doi: https://doi.org/10.1200/JCO.2010.29.7689

Gresham, G., Diniz, M. A., Raznee, Z. S., Luu, M., Kim, S., Hays, R. D., Piantadosi, S., Tighiouart, M., Yothers, G., Ganz, P. A., & Rogatko, A. (2020). Evaluating treatment tolerability in cancer clinical trials using the Toxicity Index. *Journal of the National Cancer Institute*, *11*(12), 1266–1274. https://doi.org/10.1093/jnci/djaa028

Hays, R. D., Ganz, P. A., Spritzer, K. L., & Rogatko, A. (2021). Applying the Toxicity Index to patient-reported symptom data: An example using the EORTC Colorectal Cancer-specific Quality of Life Questionnaire. *Clinical Therapeutics*, *43*(7), 1245–1252. doi: https://doi.org/10.1016/j.clinthera.2021.05.011

Hays, R. D., Spritzer, K. L., Thompson, W. W., & Cella, D. (2015). U.S. general population estimate for "excellent" to "poor" self-rated health item. *Journal of General Internal Medicine*, *30*(10), 1511–1516. doi: https://doi.org/10.1007/s11606-015-3290-x

Henry, N. L., Kim, S., Hays, R. D., Diniz, M. A., Luu, M., Cecchini, R. S., Yothers, G., Rogatko, A., & Ganz, P. A. (2021). Toxicity index, atient-reported outcomes, and early discontinuation of endocrne therapy for breast cancer risk reduction in NRG oncology/NSABP B-35. *Journal Of Clinical Oncology*, *39*, 3800–3812. doi: https://doi.org/10.1200/JCO.21.00910

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6,1*, 1–55. doi: https://doi.org/10.1080/10705519909540118

Lee, S. M., Hershman, D. L., Martin, P., Leonard, J. P., & Cheung, Y. K. (2012). Toxicity burden score: A novel approach to summarize multiple toxic effects. *Annals of Oncology*, *23*, 537–541. doi: https://doi.org/10.1093/annonc/mdr146

McColl, E. (2004). Best practice in symptom assessment: A review. *Gut 53(Suppl IV)*, iv49–iv54. doi: https://doi.org/10.1136/gut.2003.034355

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. doi: https://doi.org/10.3758/s13428-020-01398-0

Niu, S. F., PhD, R. N., Cheng, Sue-Yueh PhD, R. N., & Chin, Chia-Hui, M. S. N., (2021). RN Quality of life and severity of symptom differences between post open colectomy and laparoscopic colectomy in colorectal cancer patients. Cancer Nursing, *44*(4), doi: https://doi.org/10.1097/NCC.0000000000000793

Peipert, J. D., Nair, D., Klicko, K., Schatell, D., & Hays, R. D. (2019). Kidney Disease Quality of Life 36-item short form survey (KDQOL-36™) normative values for the United States dialysis population and new single summary score. *Journal Of The American Society Of Nephrology*, *30*(4), 654–663. doi: https://doi.org/10.1053/j.ajkd.2017.07.020

Pearman, T. P., Beaumont, J. L., Mroczek, D., O'Connor, M., & Cella, D. (2018). Validity and usefulness of a single-item measure of patient-reported bother from side effects of cancer therapy. *Cancer*, *124*(5), 991–997. doi: https://doi.org/10.1002/cncr.31133

Razaee, Z. S., Amini, A. A., Diniz, M. S., et al. (2021). On the properties of the toxicity index and its statistical efficiency. *Statistics In Medicine*, *40*, 1535–1552. doi: https://doi.org/10.1002/sim.8858

Reeve, B. B., Mitchell, S. A., Dueck, A. C., Basch, E., Cella, D., Reilly, C. M., Minasian, L. M., Denicoff, A. M., O'Mara, A. M., Fisch, M. J., Chauhan, C., Aaronson, N. K., Coens, C., & Bruner, D. W. (2014). Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *Journal of the National Cancer Institute*, *106*(7), dju129. doi:https://doi.org/10.1093/jnci/dju129

Stapleton, S. J., Holden, J., Epstein, J., & Wilkie, D. J. (2016). A systematic review of the Symptom Distress Scale in advanced cancer studies. *Cancer Nursing*, *39*(4), E9–E23. doi:https://doi.org/10.1097/NCC.0000000000000292

Steiger, J. H. (1980). Test for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251

Swain, S. M., Jeong, J. H., Geyer, C. E. Jr., Costantino, J. P., Pajon, E. R., Fehrenbacher, L., Atkins, J. N., Polikoff, J., Vogel, V. G., Erban, J. K., Rastogi, P., Livingston, R. B., Perez, E. A., Mamounas, E. P., Land, S. R., Ganz, P. A., & Wolmark, N. (2010). Longer therapy, iatrogenic amenorrhea, and survival in early breast cancer. *The New England Journal of Medicine*, *362*(22), 2053–2065. doi:https://doi.org/10.1056/NEJMoa0909638

U. S. Food & Administration, D. (2009). Guidance for industry - patient-reported outcome measures: Use in medical product development to support labeling claims. Retrieved July 18, 2022, from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, *83*(2), 213–217

Whisenant, M. S., Williams, L. A., Mendoza, T., Cleeland, C., Chen, T. H., Fisch, M. J., & Shi, Q. (2022). Identification of breast cancer survivors with high symptom burden.Cancer Nurs, *45*(4):253–261. doi: https://doi.org/10.1097/NCC.0000000000001019

Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life: A Conceptual model of patient outcomes. *Journal Of The American Medical Association*, *273*(1), 59–65. doi:https://doi.org/10.1001/jama.1995.03520250075037

Authors and Affiliations

Ron D. Hays[1,2] · Gillian Gresham[3,4] · Patricia A. Ganz[2,5] · Mourad Tighiouart[3]

✉ Ron D. Hays
drhays@ucla.edu

1 Division of General Internal Medicine & Health Services Research, UCLA Department of Medicine, 1100 Glendon Avenue, 90024 Los Angeles, CA, USA

2 Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, CA, USA

3 Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

4 Brigham and Women's Hospital, Harvard University, Boston, MA, USA

5 Center for Cancer Prevention and Control Research, Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, USA