

Needs assessments can identify scores on HRQOL questionnaires that represent problems for patients: an illustration with the Supportive Care Needs Survey and the QLQ-C30

Claire F. Snyder · Amanda L. Blackford · Julie R. Brahmer ·
Michael A. Carducci · Roberto Pili · Vered Stearns ·
Antonio C. Wolff · Sydney M. Dy · Albert W. Wu

Accepted: 8 March 2010 / Published online: 26 March 2010
© Springer Science+Business Media B.V. 2010

Abstract

Purpose A barrier to using HRQOL questionnaires for individual patient management is knowing what score represents a problem deserving attention. We explored using needs assessments to identify HRQOL scores associated with patient-reported unmet needs.

Methods This cross-sectional study included 117 cancer patients (mean age 61 years; 51% men; 77% white) who completed the Supportive Care Needs Survey (SCNS) and EORTC QLQ-C30. SCNS scores were dichotomized as “no unmet need” versus “some unmet need” and served as an external criterion for identifying problem scores. We evaluated the discriminative ability of QLQ-C30 scores using receiver operating characteristic (ROC) analysis. Domains with an area under the ROC curve (AUC) $\geq .70$ were examined further to determine how well QLQ-C30 scores predicted presence/absence of unmet need.

Results We found AUCs $\geq .70$ for 6 of 14 EORTC domains: physical, emotional, role, global QOL, pain, fatigue. All 6 domains had sensitivity $\geq .85$ and specificity $\geq .50$. EORTC domains that closely matched the content of SCNS item(s) were more likely to have AUCs $\geq .70$. The appropriate cut-off depends on the relative importance of false positives and false negatives.

Conclusions Needs assessments can identify HRQOL scores requiring clinicians’ attention. Future research should confirm these findings using other HRQOL questionnaires and needs assessments.

Keywords Health-related quality of life · Needs assessment · Clinical practice · Cancer · Interpretation

Background

In recent years, there has been increased focus on the use of health-related quality of life (HRQOL) and other patient-reported outcomes (PROs) in clinical practice for individual patient management [1]. Unlike the use of PROs as outcomes in clinical trials and observational studies where differences between groups are the primary outcomes, using PROs in clinical practice raises a number of methodological issues. For example, to be useful in managing individual patients, outcomes need to be measured far more reliably and precisely than is the case when estimating group means [2]. One of the challenges that needs to be addressed to facilitate using PROs in clinical practice is identifying what score represents a problem that requires attention (“problem score”).

The use of PROs in clinical practice is analogous to the use of lab tests in that both PROs and lab tests provide the clinician with information on the patient’s health. When clinicians seek to identify a problem or determine the differential diagnoses to be considered or ruled out, they order a variety of relevant tests, patients have the lab work completed, the results are reported to the clinician, and values that are outside of a given range motivate action. With PROs, the approach is similar, except that rather than

C. F. Snyder (✉)

Johns Hopkins School of Medicine, 624 N. Broadway,
Room 657, Baltimore, MD 21205, USA
e-mail: csnyder@jhsp.edu

A. L. Blackford · J. R. Brahmer · M. A. Carducci · R. Pili ·
V. Stearns · A. C. Wolff · S. M. Dy · A. W. Wu
Johns Hopkins School of Medicine, Baltimore, MD, USA

C. F. Snyder · S. M. Dy · A. W. Wu
Johns Hopkins Bloomberg School of Public Health, Baltimore,
MD, USA

going to a laboratory, they complete a questionnaire either in the office or via the Internet. As with lab tests that require identification of values that are “abnormal,” the use of PROs in clinical practice requires identification of “problem” scores that merit further attention.

Take, for example, Patient X who has completed the physical function questions from the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) [3]. Patient X has quite a bit of difficulty doing strenuous physical activity, a little bit of difficulty with moderate physical activity, and no problems with activities of daily living. The resulting scale score for Patient X is 60 on a scale from 0 to 100 with 100 representing better function. Does this score of 60 represent limitations in physical function that require attention from the patient’s health care provider? Few guides are currently available to help clinicians interpret the meaning of scores on PRO questionnaires. One approach, used by Velikova et al. [4], is to provide mean scores for the general population as a basis for comparison. While helpful, these normative data still do not provide a clear indication of what score represents a problem for an individual patient that requires attention.

A novel approach that might be useful in identifying problem scores on HRQOL questionnaires is to use needs assessments. Needs assessments generally ask whether a need exists and, frequently, how well existing needs are being met [5]. For example, the Supportive Care Needs Survey-Short Form (SCNS) [6, 7] asks patients whether the need is applicable to them, whether they have that need but it is being satisfied, or whether they have a low, moderate, or high level of unmet need. Thus, needs assessments provide an indication of the extent to which the patient perceives an unmet need in an area. We conducted a preliminary analysis to investigate whether needs assessments could be used to identify problem scores on HRQOL questionnaires using data from a study in which cancer patients completed both the EORTC QLQ-C30 and SCNS.

Methods

Patient population and data collection

The study population and data collection methods have been described previously [8, 9]. Briefly, the patients of seven medical oncologists involved in the treatment of breast, prostate, and lung cancers were recruited for participation through flyers handed out by clinic staff. Eligibility criteria included (1) diagnosis of breast, prostate, or lung cancer at any stage, (2) aged 18 or older, (3) currently undergoing

treatment with chemotherapy, radiation therapy, hormonal therapy, biologic therapy, or therapy as part of a clinical trial, (4) physically and cognitively able to complete the questionnaire, (5) able to read and write in English, and (6) able and willing to provide oral informed consent. Based on our plans to use the data for initial exploratory analyses, we aimed to enroll 35–50 patients per tumour type for a total target sample size of 105–150 patients.

The questionnaires for this study included the SCNS [6, 7] followed by the EORTC QLQ-C30 [3] and complied with the requirements of the instruments’ copyright holders. The QLQ-C30 includes five function domains (physical, role, emotional, cognitive, social), eight symptom domains (fatigue, pain, nausea and vomiting, dyspnoea, insomnia, appetite loss, constipation, diarrhoea), plus financial impact and global health/quality of life ratings. Most questions use a four-point scale from “not at all” to “very much” and have a 1-week recall period. Domains are transformed to a 0–100 scale with higher scores on function domains representing better function and higher scores on the other domains representing greater burden. The SCNS addresses five domains of need (psychological, health system and information, physical and daily living, patient care and support, sexual). Patients respond using a five-point scale from “not applicable” to “high need” using a one-month recall period. To calculate domain scores, we averaged the items in the domain, and scores > 2.0 represent presence of an unmet need. This cut-off of >2.0 was also applied in analyses involving individual SCNS items.

Patients also reported their age, sex, race, and education level. Clinicians completed a form that detailed patients’ Eastern Cooperative Oncology Group (ECOG) performance status, cancer type, extent of disease, and current and previous treatments. The study was reviewed and approved by the Institutional Review Board of the Johns Hopkins School of Medicine (#NA_00001797) and complies with the 1964 Declaration of Helsinki.

Research questions and hypotheses

Before we could address the main research question of whether needs assessments help identify scores on HRQOL questionnaires that represent a problem, we first had to establish whether HRQOL scores differ by level of need reported. If, for example, patients’ average HRQOL pain scores were the same regardless of whether they reported on the needs assessment that pain was “not applicable” or they had a “high need,” then a needs assessment would clearly not be helpful in identifying problem scores on the HRQOL questionnaire. We hypothesized that needs assessments would be most useful where there is a close relationship between the content of the needs assessment and the HRQOL scale. Specifically, we examined the content of both the QLQ-C30 and

Table 1 Hypothesized relationship between QLQ-C30 and SCNS domains/items and resulting areas under the curve (AUC)

EORTC Domain	SCNS Domain/Item(s)	AUC
Hypothesized Strong Relationship		
Physical Function	Physical & Daily Living Needs Domain ^a (overall domain score and individual items)	.69-.81
Role Function	Work around the home Not being able to do the things you used to	.71-.73
Emotional Function	Psychological Needs Domain ^a (overall domain score and individual items)	.56-.74
Pain	Pain	.78
Fatigue	Lack of energy/tiredness	.74
Hypothesized Moderate Relationship		
Social Function	Not being able to do the things you used to	.64
Sleep	Lack of energy/tiredness Feeling unwell a lot of the time Being given information...about aspects of managing your illness and side-effects at home	.41-.51
Hypothesized Weak Relationship		
Cognitive Function	Not being able to do the things you used to	.54
Global Health /QOL	Feeling unwell a lot of the time	.73
Nausea/Vomiting	Feeling unwell a lot of the time Being given information...about aspects of managing your illness and side-effects at home	.19-.36
Dyspnea		.37-.48
Appetite Loss		.47-.49
Constipation		.31-.37
Diarrhea		.34

^a When SCNS domains were used, we examined both the overall domain score and the scores for the individual items that comprise the domain.

SCNS, and for each QLQ-C30 domain, identified the SCNS domain/item(s) with the most similar content. While some QLQ-C30 domains (e.g., pain) had clearly similar item(s) in the SCNS, other domains (e.g., dyspnoea) had no similar item(s) in the SCNS, requiring us to use a more generic item (e.g., “feeling unwell a lot of the time”). Based on how close the content match was between the QLQ-C30 domain and the SCNS, we hypothesized whether there would be a strong, moderate, or weak association between the QLQ-C30 domain and the SCNS domain/item (Table 1).

Analysis

After performing descriptive analyses of the patients’ sociodemographic and clinical characteristics, we conducted bivariate analyses and constructed boxplots to explore how the distribution of QLQ-C30 scores varied by level of need reported.

To evaluate the discriminative ability of the QLQ-C30 domains for their hypothesized SCNS domain/item(s), we calculated the area under the receiver operating characteristic

(ROC) curve (AUC). The AUC is a quantitative measure of how well a continuous variable discriminates two binary classes (see the “[Technical Appendix](#)” for a detailed explanation of ROC analyses and a practical example). The ROC curve represents the trade-off between sensitivity and specificity. If the covariate does well at predicting the outcome, the ROC curve will curve towards the top left corner, where sensitivity and specificity are high and the AUC approaches 1.0. The worst case is where sensitivity = 1 – specificity, where the AUC equals .50.

As shown in Table 1, we generated a list of 38 hypothesized relationships between QLQ-C30 scores and binary groupings of SCNS domains/items, resulting in 38 separate ROC analyses. In our case, if our binary classes are patients with and without an unmet need, we examined how well QLQ-C30 scores predict which need group a patient is in. Of the 38 calculated AUCs, 12 were less than .50; 12 were between .50 and .70; 10 were between .70 and .75; 3 were between .75 and .80; and 1 was greater than .80. Based on this distribution, we further explored the relationship between presence or absence of unmet need and QLQ-C30

scores for pairs whose $AUC \geq .70$. Though there are no firm benchmarks on classifying AUC scores, it has been suggested that values below .70 represent poor discrimination, values between .70 and .80 represent acceptable discrimination, and values above .80 represent excellent discrimination [10]. We then tested potential cut-off scores and calculated the sensitivity and specificity associated with them, as well as the positive and negative predictive values. All analyses were performed using R version 2.7.0.

Results

Sample characteristics

The characteristics of the patient sample have been reported previously [8, 9]. Briefly, a total of 117 patients enrolled in the study, a 91% response rate from the 129 patients referred to the research coordinator by clinic staff. The majority of the patients had either breast (43%) or prostate (41%) cancer, with the remainder having lung cancer (16%). The mean age of the sample was 61.2 years, 77% were white, and 49% were women. As might be expected from a sample attending an outpatient clinic, the patients had good performance status, with 95% having ECOG ratings of 0 or 1. Half of the patients had metastatic disease. The majority of patients were currently taking hormonal therapies and had previously had surgery.

Bivariate analyses of HRQOL scores by need reported

To answer the question of whether HRQOL scores vary by presence or absence of an unmet need, we examined the distribution of HRQOL scores by level of need reported. As hypothesized, HRQOL scores showed more differentiation by need reported when the content of the need domain/item(s) was most similar to the content of the HRQOL questionnaire. Figure 1 provides example boxplots for strong, moderate, and weak relationships. For the hypothesized strong relationship (physical function and work around the home), there was a clear differentiation in scores between patients who did and did not perceive an unmet need. For the hypothesized moderate relationship (sleep and lack of energy/tiredness), there is more overlap in the boxes, but the median scores are different. For the hypothesized weak relationship (nausea/vomiting and feeling unwell a lot of the time), the medians are both at 0. These bivariate analyses confirmed that HRQOL scores do differ by level of need reported, if the content of the need domain/item is closely related to the HRQOL domain.

ROC analysis

We then performed the ROC analysis to assess the strength of the association between the HRQOL domains and need items/domains using the area under the ROC curve. Our hypotheses were largely supported, with all of the EORTC domains hypothesized to have a strong association having

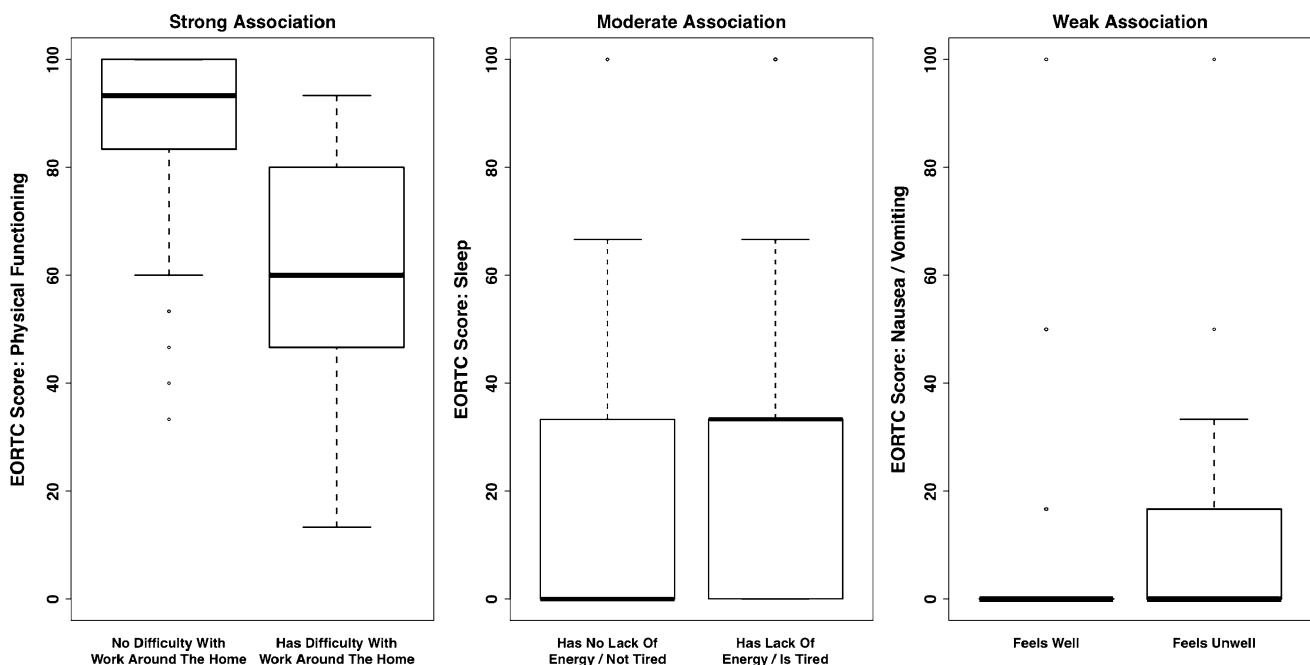


Fig. 1 Example boxplots for hypothesized strong, moderate, and weak associations between QLQ-C30 scores and presence/absence of an unmet need

an $AUC \geq .70$ with at least one need item or domain (Table 1). Where we hypothesized moderate relationships, the best matches had areas under the curve of .51 for sleep and .64 for social function. Finally, with one exception, the hypothesized weak associations had AUCs ranging from .34 to .54. However, global health/quality of life actually had a strong association with “feeling unwell a lot of the time,” with an area under the curve of .73. Based on these results, we further examined the five domains with the hypothesized strong relationship plus the global health/quality of life domain to evaluate the test characteristics associated with various cut-off scores.

Calculation of test characteristics

After calculating the AUCs for all pairings of the QLQ-C30 and SCNS items outlined above, we selected a group of pairings for further analysis. For each of the six QLQ-C30 domains with $AUCs \geq .70$, we examined the SCNS item or domain that was most accurately discriminated, i.e., had the highest AUC. In all cases, it was an individual SCNS item that had the highest AUC, so all additional analyses used these single items only. The resulting QLQ-C30–SCNS item pairings were as follows: physical function-work around the home ($AUC = .81$); role function-work around the home ($AUC = .73$); emotional function-feelings of sadness ($AUC = .74$); pain-pain ($AUC = .78$); fatigue-lack of energy/tiredness ($AUC = .74$); global health/quality of life-feeling unwell a lot of the time ($AUC = .73$) (Fig. 2). For each pairing, we calculated the sensitivity and specificity of the QLQ-C30 score using cut-offs of 0, 33, 50, 60, 70, 80, 90, and 100 for all QLQ-C30 domains except for pain and fatigue, for which we used cut-offs of 0, 5, 10, 20, 30, 33, 40, and 100. The low and high cut-offs often resulted in specificity and sensitivity values of 0 or 100%. Based on these results, we report two candidate cut-off scores for each domain in Table 2, which provide what we considered to be the most optimal trade-off between sensitivity and specificity. For all six domains, we were able to identify a cut-off with sensitivity $\geq .85$ and specificity $\geq .50$. For example, for the physical function domain, a cut-off of 90 resulted in a sensitivity of .85 and a specificity of .58, meaning that with this cut-off, 85% of patients who perceive an unmet need with physical and daily living needs are correctly identified as having a problem and 58% of patients who do not perceive an unmet need with physical and daily living needs are correctly identified as not having a problem.

We also calculated the positive and negative predictive values of the cut-offs applied to our test population. As a reminder, sensitivity and specificity are characteristics of the test, but the positive and negative predictive values vary depending on the prevalence of the condition in the

population being tested. As can be seen in Table 2, these cut-offs when applied to this test population had very high negative predictive values (ranging from .85 to .95), indicating that when the cut-off says the patient does not have a problem, the patient truly does not have a problem. The positive predictive values were somewhat lower (ranging from .39 to .68), indicating that patients identified by the cut-off as having the problem only actually have a problem 39–68% of the time. These positive and negative predictive values can help inform the identification of the appropriate cut-off score to use based on the clinical importance of false positives versus false negatives.

Application of the findings

In applying the results of this analysis, it is important to determine the appropriate cut-off scores to use. The appropriate cut-off score to use depends on the clinical importance of false positives versus false negatives. That is, is it more important that every patient who possibly has a problem in one of these areas be identified—which would argue for favouring sensitivity over specificity. Or, is it more important to be sure that if patients are identified as having a problem, they actually have the problem—which would argue for favouring specificity over sensitivity. It is also helpful to think about how the cut-off scores might be applied in practice. There are two possible approaches for how to apply these cut-offs. The first approach would use the HRQOL questionnaire score alone to determine a treatment decision: the patient scored X on the pain scale; therefore, we will prescribe pain medication. The second approach uses cut-off scores to flag potential problems that clinicians should enquire about further: the patient scored X on the pain scale; therefore, we should talk to the patient about his/her pain and potentially recommend action if our follow-up suggests that action is required. Because it is more likely that HRQOL questionnaires will be used to identify potential problems for further evaluation (the second approach), it is probably appropriate to favour sensitivity over specificity.

The positive and negative predictive values are also informative in determining the appropriate cut-off. In this test population, if the cut-off identified the patient as not having a problem, it was correct 85–95% of the time, reassuring clinicians that they need not spend time querying patients about issues the patients do not perceive as being an unmet need (as measured by the SCNS). On the other hand, if the cut-off identified the patient as having an unmet need, it was only correct 39–68% of the time, so many of the patients identified as potentially having a problem may not require intervention. If the test is only used to identify areas for further enquiry (the second approach), the relatively high rate of false positives is not

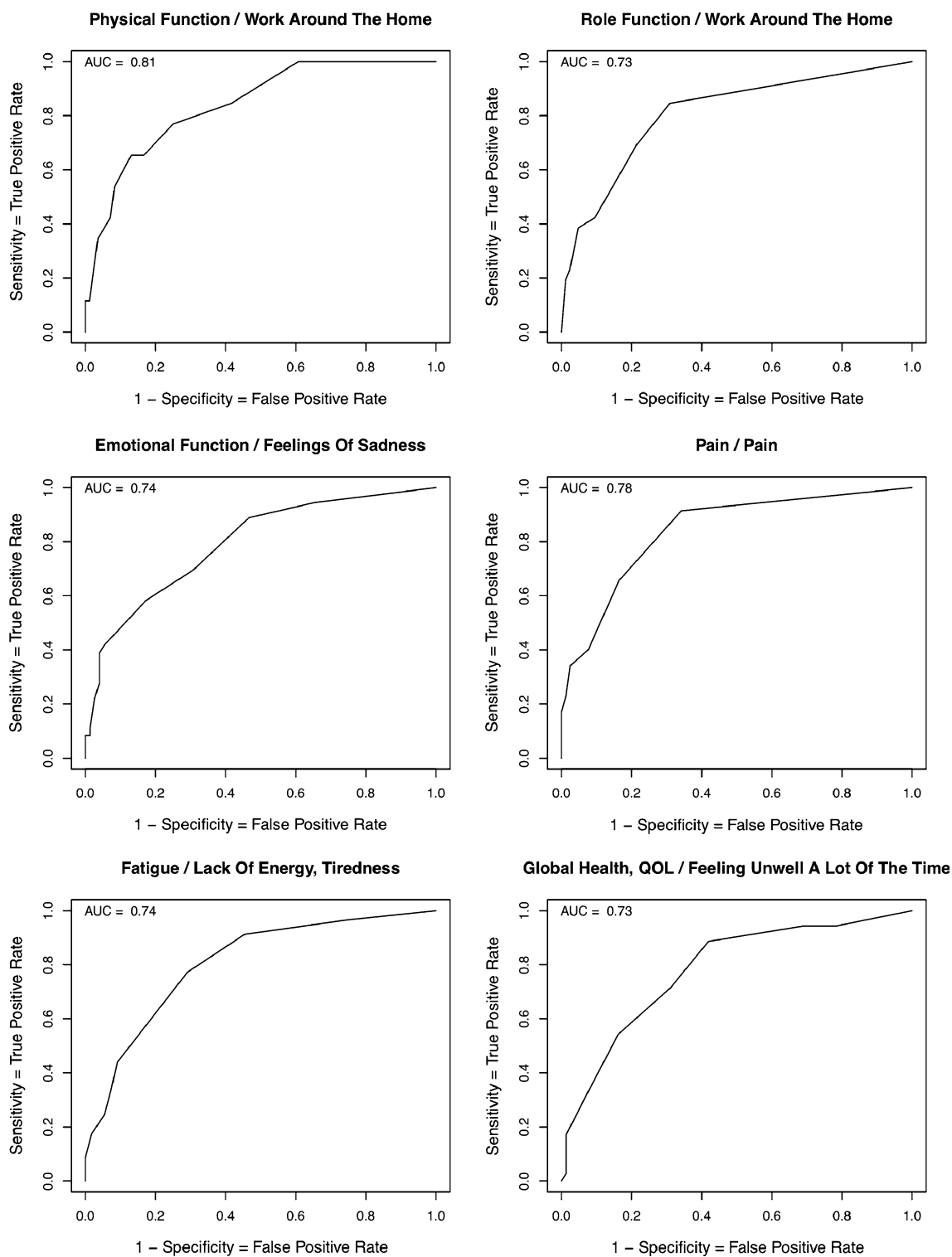


Fig. 2 ROC curves for select QLQ-C30 domains that discriminate SCNS classification with an AUC of .70 or higher

overly concerning. That being said, too many false positives can lead to “alert fatigue” (i.e., clinicians growing tired of asking patients about issues that are not really a problem) and may result in clinicians ignoring the PRO results.

Discussion

This analysis provides preliminary support for the use of needs assessments to help identify problem scores on HRQOL questionnaires. As expected, the needs assessment

Table 2 Sensitivity, specificity, positive predictive value, and negative predictive value with various cut-off scores

EORTC scale	SCNS item	Cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Physical function	Work around the home	80	.65	.83	.55	.89
		90	.85	.58	.39	.92
Role function	Work around the home	80	.69	.79	.50	.89
		90	.85	.69	.46	.94
Emotional function	Feelings of sadness	90	.89	.53	.48	.91
		100	.94	.35	.41	.93
Global health/QOL	Feeling unwell a lot of the time	80	.89	.58	.50	.91
		90	.94	.31	.39	.92
Pain	Pain	20	.66	.84	.64	.85
		10	.91	.66	.54	.95
Fatigue	Lack of energy/tiredness	20	.91	.55	.68	.86
		10	.96	.25	.57	.88

worked best for this purpose when there was good overlap between the content of the HRQOL domains and the SCNS items. For QLQ-C30 domains with close SCNS matches, there were cut-off scores with both high sensitivity and specificity, meaning that the test performs well in identifying both patients who do perceive an unmet need and patients who do not perceive an unmet need. Because there are currently few guides available to indicate what score represents a problem, these results suggest that this needs assessment approach may be useful in identifying problem scores on HRQOL questionnaires.

While the results of this analysis are encouraging, they are preliminary. Before this approach can be applied in practice, further research addressing the limitations of this study should be conducted. Because the sample size for the study was relatively small and included predominantly well-functioning patients with breast and prostate cancer, these results need to be confirmed in larger studies with more diverse samples to improve generalizability. Studies in these other samples are particularly important, given the sample-dependent nature of positive and negative predictive values.

It would also be helpful to know whether this approach also works using other HRQOL measures and/or other needs assessments. Patients' reports of their supportive care needs may differ from patient to patient or change within a given patient over time. A more rigorous evaluation would not rely on a single self-reported measure for the gold standard but use an array of measures, including patient-reported, clinician-reported, and objective disease measures, to triangulate and identify cut-points. Longitudinal studies might help us identify important changes in scores—an area that this analysis does not even begin to address.

One issue that remains is what external criterion to use for the HRQOL domains that have no good needs assessment content to match. There are several options for approaching this issue. In this study, we used a validated needs assessment, which has the benefit of using questions that have been developed and tested using a rigorous process. However, it would be possible to develop needs assessment-type questions to use as anchors specifically for the purpose of identifying problem scores on HRQOL questionnaires. This latter approach would also address another weakness of our study, which is that our two questionnaires had different recall periods. The QLQ-C30 questions generally use a 1-week recall period while the SCNS questions use a 1-month recall period, meaning that the patients' responses were not technically for the same time period. This discrepancy in recall period could lead to differences in patients' answers to the QLQ-C30 questions versus the SCNS. Developing specific questions allows for making the recall period consistent. As noted above, it would be preferable to supplement whatever PRO is used as the gold standard with other clinician-reported and objective measures.

While all of these issues will need to be addressed in future studies, the preliminary evidence from this analysis suggests that there are viable approaches for identifying problem scores on HRQOL questionnaires. Being able to identify scores that require a clinician's attention is an important step in using PROs in clinical practice to assist with individual patient management.

Acknowledgments The authors would like to thank David Ettinger, MD, and Charles Rudin, MD, for assistance in recruiting patients and Danetta Hendricks, MA, and Kristina Weeks, BA, BS, for assistance in coordinating the study. Drs. Snyder, Carducci, and Wu and Ms. Blackford are supported by a Mentored Research Scholar Grant

from the American Cancer Society (MRS-08-011-01-CPPB). This research was also supported by the Aegon International Fellowship in Oncology.

Technical Appendix

Receiver operating characteristic (ROC) curve is a common technique used in medical decision-making research to determine how well a potential classifying variable can discriminate between two classes [10]. In the context of this paper, the potential classifying variables are the continuously measured scores on the QLQ-C30, and the two classes are the binary grouping of the SCNS domains/items. As a working example, we will use the physical function domain of the QLQ-C30 as our potential classifier and the “work around the home” item from the SCNS as our binary class. For this specific pairing, there were 6 individuals with a missing result on the SCNS and 1 with a missing result on the QLQ-C30, leaving 110 patients for the analysis.

The ROC curve is a graphical representation of the trade-off between the sensitivity and specificity of a potential classifier. Sensitivity is the probability that an individual will be correctly classified as “positive” given that he or she is truly positive. Specificity is the probability that an individual will be correctly classified as “negative” when he or she is truly negative. Sensitivity is also equal to the true positive rate:

$$\text{Sensitivity} = \frac{\text{True Positive Rate}}{\text{Number of positives correctly classified}} = \frac{\text{Number of true positives}}{\text{Number of true positives}}$$

Specificity is also equal to 1 minus the false positive rate:

$$\text{Specificity} = 1 - \frac{\text{False Positive Rate}}{\text{Number of negatives incorrectly classified}} = \frac{\text{Number of true negatives}}{\text{Number of true negatives}}$$

Two other commonly used quantities in ROC analyses are positive predictive value (PPV) and negative predictive value (NPV). The positive predictive value is the probability of being a true positive when you are classified as positive:

$$\text{Positive Predictive Value} = \frac{\text{Number of true positives correctly classified}}{\text{Number of classified positives}}$$

The negative predictive value is the probability of being a true negative when you are classified negative:

$$\text{Negative Predictive Value} = \frac{\text{Number of true negatives correctly classified}}{\text{Number of classified negatives}}$$

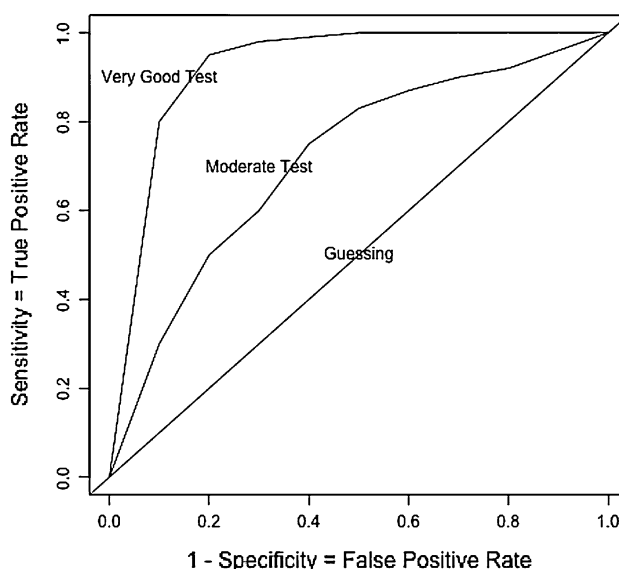


Fig. 3 Examples of three ROC curves representing predictors with chance classification ($x = y$ line), moderately good classification (middle curve), and very good classification (upper curve)

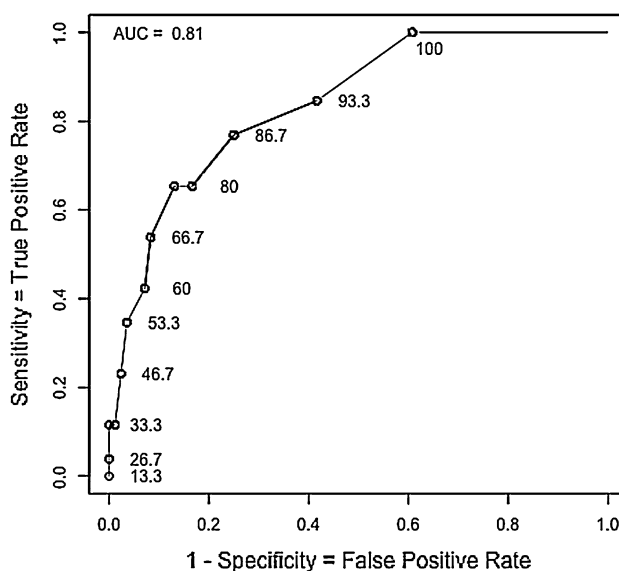


Fig. 4 ROC curve for the discriminative ability of the physical function domain of the QLQ-C30 in predicting patients who report unmet needs performing work around the home. The cut-offs of the QLQ-C30 scores that yield each possible sensitivity and specificity are shown along the curve

The ROC curve shows the trade-off between the true and false positive rates. The example in Fig. 3 illustrates three potential curves. The curve where sensitivity = 1 – specificity for all values represents mere guessing, or doing no better at classifying individuals than just flipping a fair coin. A moderately good classifier is represented by the curve in the middle, where the false positive rate

Table 3 Example of two-by-two tables for classifying patients as reporting unmet needs performing work around the home using cut-off scores of 80 and 90 on the QLQ-C30

	QLQ-C30				Total
	Cut-off = 80		Cut-off = 90		
	Classified negative	Classified positive	Classified negative	Classified positive	
SCNS					
True negative	70	14	49	35	84
True positive	9	17	4	22	26
Total	79	31	53	57	110

increases as the true positive rate increases. An even better classifier is shown in the top curve, where the false positive rate increases at a slower rate as the true positive rate increases. A perfect classifier has a single point at (0,1), where the true positive rate is 100% and the false positive rate is 0%.

The area under the ROC curve (AUC) is a commonly used measure to describe the performance of the potential classifier. The AUC is the probability that a randomly chosen true positive will be ranked higher by the classifier than a randomly chosen true negative. Thus, an AUC of .50 represents a chance classification, and an AUC of 1.0 represents perfect classification. In our example, there are 26 individuals who are true positives (report unmet needs performing work around the home) and 84 true negatives (do not report unmet needs performing work around the home). The ROC curve in Fig. 4 for the physical function domain of the QLQ-C30 has an AUC of .81, which means that of all possible pairings of the 26 positives and 84 negatives, the positive will be more highly ranked than the negative 81% of the time.

In our analysis, once a good classifier was identified, we examined potential cut-offs for the classifier by creating 2×2 tables showing the number of patients who are correctly and incorrectly classified. The physical function classifier takes on values ranging from 0 to 100 for each individual. For the purpose of illustration, in the table below we present two potential cut-offs of this score for classifying an individual as having trouble with work around the home: 80 and 90 (Table 3).

There are 31 individuals classified positive with a physical function score higher than or equal to 80 and 79 classified negative with a score less than 80. Among the 31 called positive, 17 are true positives, which results in a sensitivity of $17/26 = .65$. Among the 79 individuals classified negative, 70 are true negatives, which results in a specificity of $70/84 = .83$. The PPV and NPV of this cut-off are equal to $17/31 = .55$ and $70/79 = .89$, respectively.

Using a cut-off of 90 results in a more sensitive (sensitivity = $22/26 = .85$), but less specific test (specificity = $49/84 = .58$). The PPV and NPV of this cut-off are equal to $22/57 = .39$ and $49/53 = .92$, respectively.

References

- Aaronson, N. K., & Snyder, C. (2008). Using patient-reported outcomes in clinical practice: Proceedings of an international society for quality of life research conference. *Quality of Life Research*, 17, 1295.
- Donaldson, G. (2009). Patient-reported outcomes and the mandate of measurement. *Quality of Life Research*, 18, 1303–1313.
- Aaronson, N. K., Cull, A. M., Kaasa, S., & Sprangers, M. A. G. (1996). The European Organization for Research and Treatment of Cancer (EORTC) modular approach to quality of life assessment in oncology: An update. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 179–189). Philadelphia: Lippincott-Raven.
- Velikova, G., Booth, L., Smith, A. B., et al. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology*, 22, 714–724.
- Gustafson, D. H. (2005). Needs assessment in cancer. In J. Lipscomb, C. C. Gotay, & C. Snyder (Eds.), *Outcomes assessment in cancer: Measures, methods, and applications* (pp. 305–328). Cambridge: Cambridge University Press.
- Sanson-Fisher, R., Girgis, A., Boyes, A., et al. (2000). The unmet supportive care needs of patients with cancer. *Cancer*, 88, 226–237.
- Bonevski, B., Sanson-Fisher, R. W., Girgis, A., et al. (2000). Evaluation of an instrument to assess the needs of patients with cancer. *Cancer*, 88, 217–225.
- Snyder, C. F., Garrett-Mayer, E., Brahmer, J. R., et al. (2008). Symptoms, supportive care needs, and function in cancer patients: How are they related? *Quality of Life Research*, 17, 665–677.
- Snyder, C. F., Garrett-Mayer, E., Blackford, A. L., et al. (2009). Concordance of cancer patients' function, symptoms, and supportive care needs. *Quality of Life Research*, 18, 991–998.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Chichester, New York: Wiley.