Published in final edited form as:

*JAMA Oncol*. 2015 November ; 1(8): 1051–1059. doi:10.1001/jamaoncol.2015.2639.

# Validity and Reliability of the U.S. National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)

**Amylou C. Dueck, PhD**,
Mayo Clinic, Scottsdale, AZ, USA

**Tito R. Mendoza, PhD**,
University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

**Sandra A. Mitchell, PhD, CRNP, AOCN**,
National Cancer Institute, Rockville, MD, USA

**Bryce B. Reeve, PhD**,
University of North Carolina, Chapel Hill, NC, USA

**Kathleen M. Castro, RN, MS, AOCN**,
National Cancer Institute, Rockville, MD, USA

**Lauren J. Rogak, MA**,
Memorial Sloan Kettering Cancer Center, New York, NY, USA

**Thomas M. Atkinson, PhD**,
Memorial Sloan Kettering Cancer Center, New York, NY, USA

**Antonia V. Bennett, PhD**,
University of North Carolina, Chapel Hill, NC, USA

**Andrea M. Denicoff, MS, RN, ANP**,
National Cancer Institute, Rockville, MD, USA

---

Corresponding author: Amylou C. Dueck, Mayo Clinic, 13400 E. Shea Blvd, Scottsdale, AZ 85259, (480) 301-6159, dueck@mayo.edu.

*Additional National Cancer Institute PRO-CTCAE Study Group members are: Narre Heon, Mary Shaw, Sean Ryan, Liora P. Stark, Laura Sit, Amy Thomassie, Wendy Pettus, Lucy Gansauer, Barbara Marconi, Angela Steele-Tilton, Donna Malveaux, Jennifer Wind
Narre Heon, BA (Memorial Sloan Kettering Cancer Center, New York, NY, USA) heonn@mskcc.org
Mary Shaw, BA (Memorial Sloan Kettering Cancer Center, New York, NY, USA) ShawM@mskcc.org
Sean Ryan, MS (Memorial Sloan Kettering Cancer Center, New York, NY, USA) ryans@mskcc.org
Liora P. Stark, MPH (Memorial Sloan Kettering Cancer Center, New York, NY, USA) Liorastark@gmail.com
Laura Sit, MS (Memorial Sloan Kettering Cancer Center, New York, NY, USA) laura.sit@gmail.com
Amy Thomassie RN, BSN, CCRC (The Cancer Program of Our Lady of the Lake and Mary Bird Perkins, Baton Rouge, LA, USA) Amy.Thomassie@pbrc.edu
Wendy Pettus (Gibbs Cancer Center, Spartanburg, SC, USA) WPettus@srhs.com
Lucy Gansauer, RN, MSN, OCN (Gibbs Cancer Center, Spartanburg, SC, USA) lgansauer@gibbscc.org
Barbara Marconi, RN (Helen F. Graham Cancer Center & Research Institute, Christiana Care Health System, Newark, DE, USA) BMarconi@Christianacare.org
Angela Steele-Tilton, MSN (Helen F. Graham Cancer Center & Research Institute, Christiana Care Health System, Newark, DE, USA) Angela.b.steele-tilton@wilmu.edu
Donna Malveaux (University of Texas M. D. Anderson Cancer Center, Houston, TX, USA) dmalveau@mdanderson.org
Jennifer Wind, MA (Dana-Farber Cancer Institute, Boston, MA, USA) Jennifer_Wind@DFCI.HARVARD.EDU

**Trial Registration:** ClinicalTrials.gov NCT02158637; http://www.clinicaltrials.gov/ct2/show/NCT02158637

**Ann M. O'Mara, PhD, RN, FAAN,**
National Cancer Institute, Rockville, MD, USA

**Yuelin Li, PhD,**
Memorial Sloan Kettering Cancer Center, New York, NY, USA

**Steven B. Clauser, PhD, MPA,**
National Cancer Institute, Rockville, MD, USA

**Donna M. Bryant, MSN, ANP-BC, OCN, CCRC,**
The Cancer Program of Our Lady of the Lake and Mary Bird Perkins, Baton Rouge, LA, USA

**James D. Bearden, MD, FACP,**
Gibbs Cancer Center, Spartanburg, SC, USA

**Theresa A. Gillis, MD,**
Helen F. Graham Cancer Center & Research Institute, Christiana Care Health System, Newark, DE, USA

**Jay K. Harness, MD,**
The Center for Cancer Prevention and Treatment, St. Joseph Hospital of Orange, Orange, CA, USA

**Robert D. Siegel, MD, FACP,**
Hartford Hospital-Helen and Harry Gray Cancer Center, Hartford, CT, USA

**Diane B. Paul [Patient Advocate and Cancer Survivor],**
Brooklyn, NY, USA

**Charles S. Cleeland, PhD,**
University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

**Deborah Schrag, MD, MPH,**
Dana-Farber Cancer Institute, Boston, MA, USA

**Jeff A. Sloan, PhD,**
Mayo Clinic, Rochester, MN, USA

**Amy P. Abernethy, MD, PhD,**
Duke University Medical Center, Durham, NC, USA

**Deborah W. Bruner, RN, PhD, FAAN,**
Emory University School of Nursing, Atlanta, GA, USA

**Lori M. Minasian, MD, FACP,**
National Cancer Institute, Rockville, MD, USA

**Ethan Basch, MD, MSc**, and
University of North Carolina, Chapel Hill, NC, USA

**on behalf of the National Cancer Institute PRO-CTCAE Study Group***

Tito R. Mendoza: tmendoza@mdanderson.org; Sandra A. Mitchell: mitchlls@mail.nih.gov; Bryce B. Reeve: bbreeve@email.unc.edu; Kathleen M. Castro: kathleen.castro@nih.gov; Lauren J. Rogak: rogakl@mskcc.org; Thomas M. Atkinson: atkinsot@mskcc.org; Antonia V. Bennett: avbenn@unc.edu; Andrea M. Denicoff: denicofa@mail.nih.gov; Ann M. O'Mara: omaraa@mail.nih.gov; Yuelin Li: LiY12@mskcc.org; Steven B. Clauser: sclauser@pcori.org; Donna M. Bryant: dbryantnp@cox.net; James D. Bearden: jbearden@srhs.com; Theresa A. Gillis: TGillis@christianacare.org; Jay K.

Harness: jkharness@gmail.com; Robert D. Siegel: ROBERT_SIEGEL@bshsi.org; Diane B. Paul: funnylady93@nyc.rr.com; Charles S. Cleeland: ccleeland@mdanderson.org; Deborah Schrag: deb_schrag@dfci.harvard.edu; Jeff A. Sloan: jsloan@mayo.edu; Amy P. Abernethy: amy.abernethy@duke.edu; Deborah W. Bruner: deborah.w.bruner@emory.edu; Lori M. Minasian: minasilo@mail.nih.gov; Ethan Basch: ebasch@med.unc.edu



## Abstract

**Importance**—Symptomatic adverse events (AEs) in cancer trials are currently reported by clinicians using the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE). To integrate the patient perspective, the NCI developed a patient-reported outcomes version of the CTCAE (PRO-CTCAE) to capture symptomatic AEs directly from patients.

**Objective**—To assess the construct validity, test-retest reliability, and responsiveness of PRO-CTCAE items.

**Design**—Participants completed PRO-CTCAE items on tablet computers in clinic waiting rooms at two visits 1-6 weeks apart. A subset completed PRO-CTCAE items during an additional visit one business day after the first visit.

**Setting**—Nine U.S. cancer centers and community oncology practices.

**Participants**—975 adult cancer patients undergoing outpatient chemotherapy and/or radiation enrolled between January 2011 and February 2012. Eligibility required participants to read English and be without clinically significant cognitive impairment.

**Main Outcome(s) and Measure(s)**—Primary comparators were clinician-reported Eastern Cooperative Oncology Group Performance Status (ECOG PS) and the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (QLQ-C30).

**Results**—940/975 (96%) and 852/940 (91%) participants completed PRO-CTCAE items at each visit. 938/940 (99.8%) participants (53% female, median age 59, 32% high school education or less, 17% ECOG PS 2-4) reported having at least one symptom. All PRO-CTCAE items had at least one correlation in the expected direction with a QLQ-C30 scale (111/124 $P<.05$). Stronger correlations were seen between PRO-CTCAE items and conceptually-related QLQ-C30 domains. Scores for 94/124 PRO-CTCAE items were higher in the ECOG PS 2-4 versus 0-1 group (58/124 $P<.05$). Overall, 119/124 items met at least one construct validity criterion. Test-retest reliability was acceptable for 36/49 pre-specified items (median intra-class correlation coefficient .76; range .53-.96). Correlations between PRO-CTCAE item changes and corresponding QLQ-C30 scale changes reached statistical significance for 27 pre-specified items (median $r=.43$, range .10-.56; all $P\leq.006$).

**Conclusions and Relevance**—Evidence demonstrates favorable validity, reliability, and responsiveness of PRO-CTCAE in a large, heterogeneous U.S. sample of patients undergoing cancer treatment. Studies evaluating other measurement properties of PRO-CTCAE are underway to inform further development of PRO-CTCAE and its inclusion in cancer trials.

## Introduction

In cancer clinical trials, adverse events (AEs) are collected and reported using the U.S. National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events

(CTCAE).[1] The CTCAE is a library of items representing 790 discrete AEs, each graded using an ordinal severity scale.[2] Approximately 10% of AEs in the CTCAE are symptoms (e.g., nausea, sensory neuropathy), which in trials have historically been reported by clinical investigators.[3] However, there is empiric evidence that collection of this information directly from patients improves the precision and reliability of symptomatic AE detection in trials,[4-9] and is feasible.[10,11] Moreover, there is substantial evidence that clinical investigators may miss up to half of patients' symptomatic AEs.[5,6,12,13]

To improve precision and patient-centeredness in the capture of symptomatic AEs, the NCI developed a library of patient-reported outcome (PRO) items to supplement the CTCAE, called the PRO-CTCAE,[14] as has been previously described.[15] Of the 790 AEs in the CTCAE, 78 were identified as amenable to patient self-report. For each of these AEs, PRO items were created reflecting the attributes of frequency, severity, interference with usual or daily activities, amount, or presence/absence. One to three attributes were selected for any given AE depending on the content of the CTCAE criteria for that AE and the nature of that particular AE. In total, 124 individual items represent the 78 symptomatic AEs currently in the PRO-CTCAE item library.

The generic structure for PRO-CTCAE items and response options are shown in Table 1. Each item includes a plain language term for the AE, the attribute of interest, and the standard recall period of "the past 7 days". Cognitive interviews previously determined a high level of patient understanding and meaningfulness of the items.[16] Software was developed for administering PRO-CTCAE items to patients either via web or an automated telephone interactive voice response (IVR) interface, and was refined through usability testing.[15,17]

For any new measurement tool in clinical research (e.g., biomarkers, imaging, diagnostic test), it is essential to establish that the new instrument accurately and reliably captures the underlying phenomenon it is intended to measure. To accomplish this for the PRO-CTCAE, this study was designed to evaluate the measurement properties of the 124 items in the PRO-CTCAE item library including validity (degree to which an instrument accurately measures the underlying phenomenon), reliability (ability of an instrument to produce similar scores on repeated measurements under similar conditions), and responsiveness (capacity of an instrument to show a change when there has been a change in the underlying phenomenon). These properties were examined individually for each item since PRO-CTCAE items are individually reported in trials and not aggregated into a single score. Inclusion of patients with diversity with respect to cancer type, treatment modality, and sociodemographic characteristics was considered essential given the intended use of PRO-CTCAE across varying research contexts. To simultaneously evaluate the measurement properties of 124 items within a single study required us to employ a varied set of comparators or "anchors", and warranted a larger and more diverse sample of respondents and settings than is typically employed in most validation studies of fixed length PRO measures.

## Methods

### Patients

Adult patients initiating or undergoing outpatient chemotherapy, radiation, or both at one of nine U.S. cancer centers or community oncology practices were approached in clinical waiting areas and invited to participate in this study. Participating sites with number of patients enrolled included Dana-Farber Cancer Institute, Boston, MA (N=40); Hartford Hospital-Helen and Harry Gray Cancer Center, Hartford, CT (N=104); Helen F. Graham Cancer Center & Research Institute at Christiana Care Health System, Newark, DE (N=105); Mayo Clinic, Rochester, MN (N=9); Memorial Sloan Kettering Cancer Center, New York, NY (N=280); Our Lady of the Lake and Mary Bird Perkins Cancer Center, Baton Rouge, LA (N=133); Gibbs Cancer Center, Spartanburg, SC (N=113); St. Joseph Hospital of Orange, Orange, CA (N=104); and University of Texas M. D. Anderson Cancer Center, Houston, TX (N=52).

Eligibility criteria required that all participants be able to read and comprehend English, be without clinically significant cognitive impairment based on site investigator judgment, have a cancer diagnosis, and be actively undergoing cancer treatment or be initiating treatment in the next 7 days. Patients with any cancer type were eligible, but an accrual strategy was used to enrich for specific cancer types in order to facilitate planned comparisons between groups based on cancer type in the validity analysis, including breast; aerodigestive tract (head/neck and esophageal cancer); genitourinary (prostate and bladder); lung; colorectal; and lymphoma/myeloma. An enrichment strategy was also employed to ensure that a minimum of 15% of participants had impaired performance status (PS) defined as Eastern Cooperative Oncology Group (ECOG) PS ≥2.

Study sites were selected to encompass geographic, racial/ethnic, economic, and educational diversity reflective of the U.S. population with the understanding that the requirement to be English speaking would limit the enrollment of Hispanic patients (a separate study evaluating the Spanish language version of the PRO-CTCAE has been conducted[18]). Race/ethnicity was self-reported by patients.

Institutional review board approval was obtained at all sites and at the NCI, and all patients completed written informed consent. The trial was registered on ClincialTrials.gov (NCT02158637). Each participant received a $20 gift card or parking voucher.

### Questionnaire

The previously developed PRO-CTCAE item library consists of 78 symptomatic AEs represented by 124 distinct items.[14,15] To limit burden, a maximum of 58 symptomatic AEs (82 items) was presented to each participant. Seven electronic surveys targeted towards different cancer types (eTable 1) were created in the central PRO-CTCAE web survey administration platform. As part of the registration process, the site coordinator selected a single survey based on the patient's diagnosis, and that survey was then automatically scheduled for completion at each visit. All surveys included a set of 20 "core" symptomatic AEs[15], predetermined based on high prevalence across cancer types in prior NCI-sponsored clinical trials.[19] Remaining symptomatic AEs were classified a priori as likely to be

prevalent or non-prevalent in specific cancer types based on expert consultation, patient representative input, and literature review. These items were included on surveys for selected cancer types to facilitate planned comparisons between groups based on cancer type. When 80% of accrual was reached, to increase sample size for the 58 symptomatic AEs which were not systematically administered to all patients, a new survey containing exactly these 58 symptomatic AEs was administered to all subsequently enrolled patients.

## Procedure

PRO-CTCAE items were completed by participants prior to clinic appointments on tablet computers via the PRO-CTCAE measurement system hosted on a secure server at the NCI.[17] To optimize usability by individuals with disabilities, PRO-CTCAE software is compliant with Section 508 of the U.S. Rehabilitation Act. The PRO-CTCAE measurement system employs conditional branching for AEs that contain more than a single attribute, such that subsequent items about severity or interference are skipped if respondents indicate that they are not experiencing a specific symptomatic AE. Participants were required to answer questions without assistance, but could request technical assistance with using the tablet computer from study staff.

## Anchors

Anchors are measurable criteria pre-specified as comparators in an instrument validation study. Examples of anchors relevant in PRO validation studies include well-validated patient- and clinician-reported outcomes and clinical variables such as disease site or concurrent medication use. For this study, anchors selected a priori included both generic measures (e.g., patient-reported global health-related quality of life [HRQOL] or clinician-reported performance status) and more specific clinical variables (e.g., antiemetic use or receipt of taxane chemotherapy). These anchors were selected based on literature review, expert consensus, and patient representative input.

The PRO anchors were administered to participants using a paper booklet containing the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30),[20] a 30-item instrument which produces a HRQOL summary score,[21,22] a global health status/quality of life (QOL) scale score, 5 functioning (physical, role, emotional, social, cognitive) scale scores, and 9 selected symptom item/scale scores. 28 items are measured on a 1-4 scale (1=not at all; 4=very much) with the remaining two items (overall health and QOL) scored on a 1-7 scale (1=very poor; 7=excellent). Like PRO-CTCAE, the recall period for the QLQ-C30 is "the past week". Patients also completed three Global Impression of Change (GIC)[23,24] items at the primary follow-up visit. These items asked patients to rate their changes in overall QOL; physical condition; and emotional state on a 7-point scale ranging from "very much better", "moderately better", "a little better", "about the same", "a little worse", "moderately worse", to "very much worse".

Clinician-reported ECOG PS was collected at each visit via a case report form. Other clinical anchors were abstracted from medical charts and included whether the participant had received radiation, surgery, and/or chemotherapy in the prior two weeks; type of chemotherapy; and use of specific medication classes, including: hormonal therapy, narcotic

analgesics, laxatives/stool softeners, antiemetics, sleep aids, anti-diarrhea medications, antacids, bronchodilators/inhaled corticosteroids, anxiolytics, and/or antidepressants.

## Study Visits

Participants were assigned to one of three groups with differing questionnaire schedules based on cancer type and clinic visit schedule, to avoid the necessity of extra clinic visits in this symptomatic population (eFigure 1). Group A included patients undergoing daily radiation or chemoradiation to enable analyses of test-retest reliability and varying recall periods (recall period analyses will be reported separately).[25] Group B included patients with at least four planned consecutive weekly clinic visits. Group C included participants whose planned clinic visits precluded participation in Group B but who did have a return clinic visit planned within 1-6 weeks. Irrespective of group assignment, all patients completed PRO-CTCAE items and QLQ-C30 at two visits that were spaced approximately 1-6 weeks apart. At each visit, ECOG PS and other clinical anchors were recorded on case report forms. PRO-CTCAE surveys administered to patients in Group A on the business day following study day 1 were used for the analysis of test-retest reliability, and included 49 pre-specified PRO-CTCAE items.

## Statistical analysis

Construct validity reflects the association between a new measurement tool and an established measure of the underlying concept(s) of interest. Construct validity is often investigated through convergent validity, which determines if the new measure moves in the same direction as an established instrument, and known-groups validity, which determines if the measurement tool can distinguish between groups of patients who are thought to be distinct with respect to the underlying concept being measured. To assess convergent validity, Pearson correlations were computed between each PRO-CTCAE item and QLQ-C30 HRQOL summary and other functioning/symptom scale scores. To aid interpretation, QLQ-C30 HRQOL summary and functioning/global scales were reverse scored such that higher scores represent inferior outcomes, matching the direction of PRO-CTCAE items. Pearson correlation values of .1, .3, and .5 were interpreted as small, medium, and large.[26] To assess known-groups validity, two-sample t-tests for ordinal 0-4 scales and chi-squared tests for binary scales were used to compare each PRO-CTCAE item between patients with high and low performance status (ECOG PS 0-1 versus 2-4). Additional known-groups analyses were pre-specified for PRO-CTCAE items that were expected to be higher in one group of patients versus another on the basis of cancer type, treatment, or other clinically relevant characteristic (e.g., pain in the abdomen in patients with gastrointestinal versus lung cancers). Effect sizes (computed as the difference between group means divided by the pooled standard deviation [Cohen's *d*], or difference between twice the arcsine of the square root of each sample proportion [Cohen's *h*]) of .2, .5, and .8 were interpreted as small, medium, and large.[26]

Test-retest reliability was estimated using the intra-class correlation coefficient (ICC) based on a one-way analysis of variance model[27] with an ICC of .7 or greater interpreted as acceptable.[28] Responsiveness of items was investigated by comparing change from first to second visit in 27 PRO-CTCAE items selected a priori. Comparisons were made using a

one-sided Jonckheere-Terpstra test across respondents who reported their GIC to be worse ("a little worse", "moderately worse", or "very much worse"), unchanged ("about the same"), or improved ("a little better", "moderately better", or "very much better").[29] Standardized response means (SRM) were computed as the mean change score divided by the standard deviation of the change scores within each change category (worse versus no change versus improved) for each PRO-CTCAE item. Pearson correlations were also computed between PRO-CTCAE item changes and QLQ-C30 scale changes. One GIC item and one QLQ-C30 scale were specified a priori for each of the 27 PRO-CTCAE items. See eTable 2 for symptomatic AEs included in each analysis.

To accommodate conditional branching in the PRO-CTCAE software, values for automatically skipped items were assumed to be zero. P-values <.05 were considered statistically significant throughout. To take into consideration potential collinearity and multiplicity, sensitivity analyses employed a stricter p-value cut-off of <.001 and Hochberg's step-up procedure[30] across construct validity analyses within each item. An item was considered valid if statistical significance ($P<.05$) along with a meaningful effect size (Pearson $r \geq .1$ or group difference effect size $d$ or $h \geq .2$) was observed for at least one convergent or known-groups validity analysis.

## Results

Between January 2011 and February 2012, 975 patients initiating or undergoing chemotherapy and/or radiation were enrolled with 940/975 (96%) eligible patients completing PRO-CTCAE items at Visit 1 and 852/940 (91%) completing PRO-CTCAE items at Visit 2 (eFigure 1). Characteristics of the 940 participants included in this analysis are presented in Table 2. Median age was 59 years (range 19-91), 539 (57%) were female, 161 (17%) had impaired PS (ECOG 2-4), and 305 (32%) had no more than a high school education.

Most participants (938/940 [99.8%]) reported presence of at least one symptom (i.e., a score greater than 0) during the two primary visits, with 768/940 (82%) reporting at least one symptom as frequent, severe, and/or interfering "quite a bit" with daily activities. Patients were broadly symptomatic reporting presence of a median of 23 symptoms (range 0-91) with 904/940 (96%) reporting presence of 5 or more symptoms at the first visit. 118/124 (95%) PRO-CTCAE items were reported as present by at least 10% of respondents at both primary visits, with 82/124 (66%) items having at least 25% prevalence. The distribution of item scores for the set of 20 "core" symptomatic AEs appears in eFigure 2.

Detailed results related to construct validity of PRO-CTCAE items using all anchors are provided in eTable 3. With respect to convergent validity, 122/124 (98%) PRO-CTCAE items were associated in the expected direction with the QLQ-C30 HRQOL summary score (102/124 $P<.05$; 87/124 $P<.001$; Figure 1); 107/124 items demonstrated meaningful correlation (Pearson $r \geq .1$). When considering all QLQ-C30 functioning/global scales, 124/124 (100%) PRO-CTCAE items were associated in the expected direction with one or more scales, with 114/124 demonstrating meaningful correlation (Pearson $r \geq .1$), and 111/124 coefficients reaching statistical significance ($P<.05$; 90/124 $P<.001$). PRO-CTCAE

items that were likely to impact physical functioning had the strongest correlations with the QLQ-C30 physical functioning scale (e.g., shortness of breath severity: Pearson $r$=.47, $P$<.001) whereas items likely to impact cognitive functioning had the strongest correlations with the QLQ-C30 cognitive functioning scale (e.g., problems with concentration severity: Pearson $r$=.71, $P$<.001; problems with memory severity: Pearson $r$=.69, $P$<.001). Similar results were seen between PRO-CTCAE items and conceptually-related QLQ-C30 emotional, role, and social functioning scales. For those PRO-CTCAE items with a parallel QLQ-C30 symptom scale/item (e.g., fatigue), large correlations between analogous items (all Pearson $r$>.69, $P$<.001) were consistently observed.

In the known-groups comparison between patients with low and high performance status, 94/124 PRO-CTCAE items had higher mean scores in the ECOG PS 2-4 group versus 0-1 group (58/124, $P$<.05; 37/124, $P$<.001; shown for 37 PRO-CTCAE items in eFigure 3).

In 127 a priori known-groups comparisons involving 87 PRO-CTCAE items based on cancer type, treatment, or other clinically relevant characteristic, 110/127 comparisons demonstrated higher PRO-CTCAE scores in the group expected to have worse symptom experience (85/127, $P$<.05; 53/127, $P$<.001, eTable 3).

Most PRO-CTCAE items (119/124) reached a statistically significant and meaningful effect size on one or more construct validity criteria. The five items that did not exhibit at least one statistically significant and meaningful effect had low prevalence in this sample, thereby limiting our analysis. These items were: nosebleeds (prevalence 14.9% [frequency] and 14.0% [severity]); pain, swelling or redness at site of drug injection or intravenous therapy (prevalence 12.5%); pain during vaginal sex (prevalence 20.7%); and rash (prevalence 17.5%). A majority of PRO-CTCAE items (99/124 and 101/124) remained statistically significant under stricter criteria ($P$<.001 and Hochberg's $P$<.05) in sensitivity analyses (eTable 3).

In the subset of 80 respondents who completed PRO-CTCAE on consecutive business days (median 1 day, range 1-3 days), the test-retest reliability for the 49 pre-specified items ranged from .53 to .96 (median ICC .76) with 36/49 items having an ICC ≥.7 (eTable 4).

In the analysis of responsiveness (Figure 2), statistically significant ($P$<.05) monotonically decreasing mean PRO-CTCAE change scores were observed for 23 of 27 pre-specified items ($P$<.001 for 13 items). The median SRM in patients reporting worsening was .19 (range .03-.40), whereas median SRM in patients reporting improvement was -.14 (range -.30-.09). Statistically significant correlations were observed between PRO-CTCAE item changes and corresponding QLQ-C30 scale changes for all 27 pre-specified items (median $r$=.43, range .10-.56; all $P$≤.006).

## Discussion

This large-scale multicenter study in adults undergoing active cancer therapy provides evidence supporting the validity, reliability, and responsiveness of the items in the PRO-CTCAE library. The PRO-CTCAE is unique in its intended use to complement the CTCAE

by providing comprehensive data on symptomatic AEs in cancer clinical trials from the patient perspective.

The design of this study posed a unique methodological challenge, due to the goal of assessing, within a single investigation, the measurement properties of 124 individual items representing a broad spectrum of symptomatic toxicities. Typically, PRO validation studies will test the properties of a single composite index score or a small number of domains that encompass related concepts. For the assessment of validity in the current study, the primary strategy to address this challenge was inclusion of both broad generic anchors (e.g., global HRQOL, ECOG PS) and more specific clinical variables (e.g., receipt of specific medication classes such as antiemetics). Interestingly, all of the PRO-CTCAE items were associated in the expected direction with at least one generic functioning measure, suggesting the impact that even a single toxicity may have on the patient experience.

Strengths of this study include the diverse sample, reflecting a wide range of cancer types and treatment modalities, and enrichment for less common cancer types. The sample was also successfully enriched for patients with impaired performance status (ECOG PS ≥2), enabling demonstration of the meaningfulness of PRO-CTCAE among those with substantial symptom burdens, as well as the feasibility of survey administration in debilitated patients. Moreover, participants were accrued at both academic and community sites across the U.S., including rural and urban settings, and reflected a range of educational and racial backgrounds.

Several caveats should be considered. First, our study was conducted in an English-speaking U.S.-residing patient population. Ongoing research is evaluating linguistic adaptations of PRO-CTCAE, and the measurement properties of both the English and other language versions in settings outside the U.S.[31] Linguistic validation of a Spanish language translation of PRO-CTCAE is being reported elsewhere.[18] Second, we assessed reliability in a subset of 49 items; thus, future studies to examine the test-retest reliability of the remaining PRO-CTCAE items are warranted. Third, a small number of highly specific symptomatic AEs were uncommon in the study sample and received low endorsement rates, thus limiting our ability to evaluate their measurement properties. Specifically, five items reflecting four symptomatic AEs (nosebleeds; pain/swelling/redness at site of drug injection or intravenous therapy; pain during vaginal sex; and rash) did not exhibit a statistically significant and meaningful effect on at least one construct validity criterion. These items are being evaluated in other clinical trial contexts. While the large number of items and anchors evaluated in this study raises the possibility of inflated Type I error, in sensitivity analyses using more stringent significance thresholds, the majority of items retained statistical significance. Lastly, notwithstanding inclusion of diverse malignancies in this study, results may not fully generalize to populations with rare tumor types. However, a prior cognitive interviewing study[16] affirms that PRO-CTCAE items were well understood by respondents with varying disease sites and receiving diverse anti-cancer treatments. Continued evaluation of PRO-CTCAE is currently underway in a variety of trial contexts to support the interpretability and value of patient-reporting of symptomatic treatment-related toxicity.

The CTCAE has historically enabled clinicians to describe the toxicity burden of cancer treatments using a consistent standard language allowing comparisons across trials. The value of patients' input in describing their own experiences is well recognized. Having a measurement system which integrates the patient perspective into AE reporting and which fosters consistency, transparency, and comparability across trials is similarly an important objective. The results of this validation study suggest that PRO-CTCAE can achieve its intended aim of integrating the patient experience into routine clinical trial AE reporting thereby augmenting the capacity for informed decision-making. In conclusion, this large-scale multicenter validation study in individuals undergoing active cancer therapy provides robust evidence for the validity, reliability, and responsiveness of items in the PRO-CTCAE library.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, and approval of the manuscript was supported by contracts from the U.S. National Cancer Institute: HHSN261200800043C, HHSN261201000063C, and HHSN261200800001E. The authors are grateful to the patients and clinicians who participated in this study. We thank Gitana Davila and Kathy Alexander for their contributions to the study including study coordination, patient recruitment, and data management at The Center for Cancer Prevention and Treatment at St. Joseph Hospital of Orange (Orange, CA, USA) and Hartford Hospital-Helen and Harry Gray Cancer Center (Hartford, CT, USA), respectively. We also thank Dr. Neil Aaronson for his methodological input on the EORTC QLQ-C30. The authors declare that they have no competing interests in relation to this research. Dr. Dueck had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References

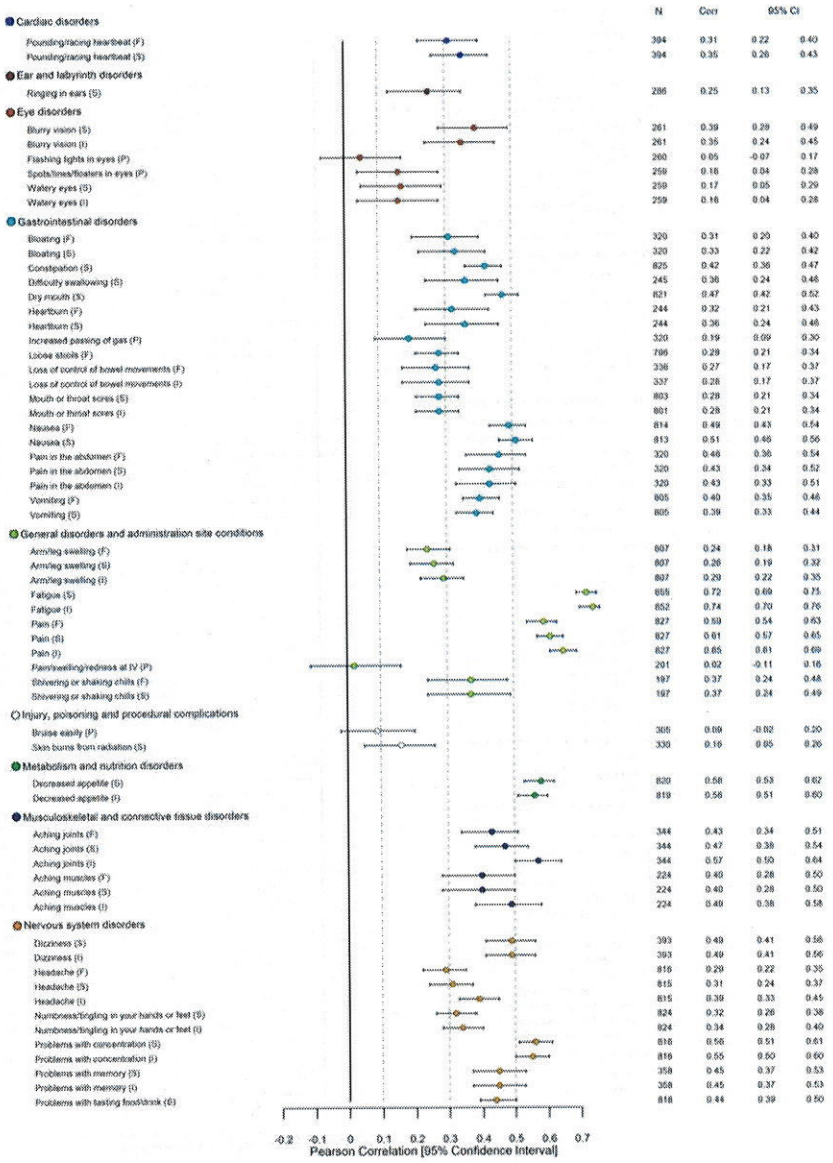
1. National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services. [last accessed March 16, 2015] Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. NIH publication # 09-7473. Published May 29, 2009; Revised Version 4.03 June 14, 2010. Available at http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5×7.pdf
2. National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services. [last accessed March 16, 2015] caBIG Knowledge Center, CTCAE FAQ. Available at http://evs.nci.nih.gov/ftp1/CTCAE/Archive/CTCAE_4.01_2009-07-14_FAQ.doc
3. Trotti A, Colevas AD, Setser A, Basch E. Patient-Reported Outcomes and the Evolution of Adverse Event Reporting in Oncology. J Clin Oncol. 2007; 25(32):5121–7. [PubMed: 17991931]
4. Atkinson TM, Li Y, Coffey CW, Sit L, Shaw M, Lavene D, Bennett AV, Fruscione M, Rogak L, Hay J, Gönen M, Schrag D, Basch E. Reliability of adverse symptom event reporting by clinicians. Qual Life Res. 2012; 21:1159–1164. [PubMed: 21984468]
5. Basch E. The Missing Voice of Patients in Drug-Safety Reporting. N Engl J Med. 2010; 362(10): 865–869. [PubMed: 20220181]
6. Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol. 2004; 22(17):3485–90. [PubMed: 15337796]
7. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008; 14(8):530–9. [PubMed: 18690769]

8. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, Appawu M, Iasonos A, Atkinson T, Goldfarb S, Culkin A, Kris MG, Schrag D. Adverse symptom reporting by patients versus clinicians: relationships with clinical outcomes. J Nat Cancer Inst. 2009; 101(23):1–9.
9. Quinten C, Maringwa J, Gotay CC, Martinelli F, Coens C, Reeve BB, Flechtner H, Greimel E, King M, Osoba D, Cleeland C, Ringash J, Schmucker-Von Koch J, Taphoorn MJ, Weis J, Bottomley A. Patient self-reports of symptoms and clinician ratings as predictors of overall cancer survival. J Natl Cancer Inst. 2011; 103(24):1851–8. [PubMed: 22157640]
10. Meacham, R.; McEngart, D.; O'Gorman, H.; Wenzel, K. Use and compliance with electronic patient reported outcomes within clinical drug trials; Presented at 2008 ISPOR (International Society for Pharmacoeconomics and Outcomes Research) Annual International Meeting; Toronto, Ontario, Canada.
11. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, Hudis CA, Scher HI, Sabattini P, Schrag D, Basch E. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. J Clin Oncol. 2013 Jul 10; 31(20):2580–5. [PubMed: 23733753]
12. Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, Gridelli C, Gebbia V, Ciardiello F, De Placido S, Ceribelli A, Favaretto AG, de Matteis A, Feld R, Butts C, Bryce J, Signoriello S, Morabito A, Rocco G, Perrone F. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. J Clin Oncol. 2015; 33(8):910–5. [PubMed: 25624439]
13. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, Scher HI, Schrag D. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol. 2006; 7(11): 903–9. [PubMed: 17081915]
14. Basch EM, Reeve BB, Mitchell SA, Clauser SB, Minasian L, Sit L, Chilukuri R, Baumgartner P, Rogak L, Blauel E, Abernethy AP, Bruner D. Electronic toxicity monitoring and patient-reported outcomes. Cancer J. 2011 Jul-Aug;17(4):231–4. [PubMed: 21799330]
15. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, Mendoza TR, Hay J, Atkinson TM, Abernethy AP, Bruner DW, Cleeland CS, Sloan JA, Chilukuri R, Baumgartner P, Denicoff A, St Germain D, O'Mara AM, Chen A, Kelaghan J, Bennett AV, Sit L, Rogak L, Barz A, Paul DB, Schrag D. Development of the National Cancer Institute 's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst. 2014 Sep 29.106(9)
16. Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, Minasian LM, Clauser SB, Dénicoff A, O'Mara A, Chen A, Bennett AV, Paul DB, Gagne J, Rogak L, Sit L, Viswanath V, Schrag D, Basch E. NCI PRO-CTCAE Study Group. Cognitive interviewing of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Qual Life Res. 2014 Feb; 23(1):257–69. [PubMed: 23868457]
17. Agarwal, H.; Gulati, M.; Baumgartner, P.; Coffey, D.; Kumar, V.; Chilukuri, R.; Shouery, M.; Reeve, B.; Basch, E. Web interface for the patient-reported version of the CTCAE; NCI caBIG Annual Meeting; July 2009; Washington, DC.
18. Arnold, BJ.; Mitchell, SA.; Lent, L.; Mendoza, TR.; Rogak, L.; Barragan, N.; Willis, GB.; Medina, M.; Lechner, SC.; Harness, J.; Basch, E. Linguistic validation of the Spanish translation of the U S National Cancer Institute's Patient- Reported Outcomes version of the Common Terminology Criteria for Adverse Events; Presentation. International Society for Quality of Life Research 21st Annual Conference; Berlin, Germany. October 15-18, 2014;
19. Reeve BB, Mitchell SA, Dueck AC, Basch E, Cella D, Reilly CM, Minasian LM, Denicoff AM, O'Mara AM, Fisch MJ, Chauhan C, Aaronson NK, Coens C, Bruner DW. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. J Natl Cancer Inst. 2014 Jul 8.106(7)
20. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993 Mar 3; 85(5):365–76. [PubMed: 8433390]

21. Gundy CM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MA, Velikova G, Aaronson NK. Comparing higher order models for the EORTC QLQ-C30. Qual Life Res. 2012 Nov; 21(9):1607–17. [PubMed: 22187352]
22. Kieffer, JM.; Giesinger, JM.; Fayers, P.; Groenvold, M.; Peterson, MA.; Scott, NW.; Sprangers, MAG.; Velikova, G.; Aaronson, NK. Replication and relative validity of the mental and physical higher order model of the EORTC QLQ-C30; Presentation. International Society for Quality of Life Research 21st Annual Conference; Berlin, Germany. October 15-18, 2014;
23. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol. 1998 Jan; 16(1):139–44. [PubMed: 9440735]
24. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002 Apr; 77(4):371–83. [PubMed: 11936935]
25. Mendoza, TR.; Bennett, AV.; Mitchell, SA.; Reeve, BB.; Atkinson, TM.; Li, Y.; Rogak, L.; Dueck, AC.; Basch, E. Impact of recall period on the accuracy of selected items from the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE); Poster. International Society for Quality of Life Research 21st Annual Conference; Berlin, Germany. October 15-18, 2014;
26. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
27. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–8. [PubMed: 18839484]
28. Nunnally, JC.; Berstein, IH. Psychometric Theory. 3rd. New York: McGraw-Hill; 1994.
29. Jonckheere AR. A distribution-free k-sample test against ordered alternatives. Biometrika. 1954; 41:133–45.
30. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988; 75:800–2.
31. Kirsch M, Mitchell SA, Dobbels F, Stussi G, Basch E, Halter JP, De Geest S. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. Eur J Oncol Nurs. 2015 Feb; 19(1):66–74. [PubMed: 25190633]

## At A Glance

- Symptomatic adverse events (AEs) in cancer trials are currently graded by clinicians using the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE)
- This study assessed the measurement properties (validity, reliability, and responsiveness) of the newly developed NCI Patient-Reported Outcomes version of the CTCAE (PRO-CTCAE)
- A large, heterogeneous sample of 940 adult cancer patients undergoing outpatient cancer treatment provided PRO-CTCAE and other patient-reported and clinical data
- A majority of the PRO-CTCAE items (119 out of 124) met at least a validity criterion
- PRO-CTCAE provides a valid and reliable assessment of symptomatic toxicities from the patient's perspective, and is encouraged for use in oncology trials to enhance the accuracy of AE reporting



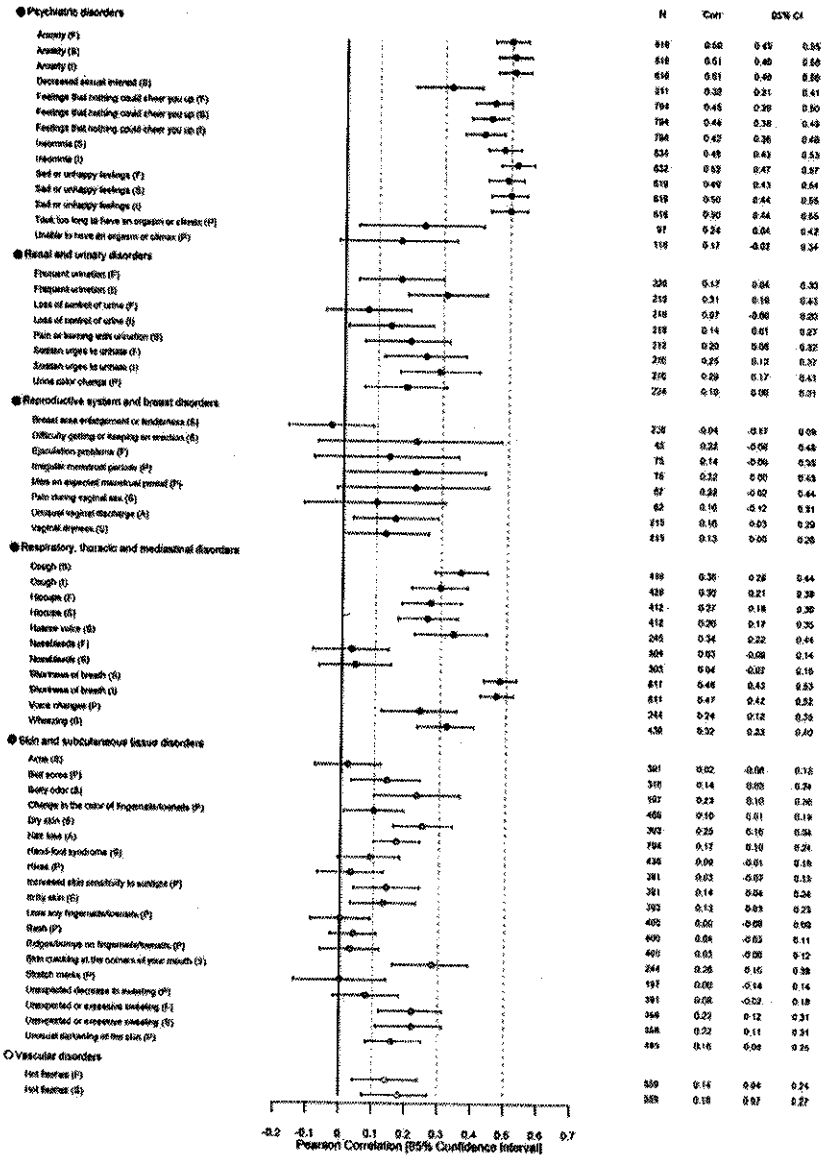



**Figure 1. Pearson Correlations between 124 PRO-CTCAE Item Scores and EORTC QLQ-C30 HRQOL Summary Score* at Visit 1**

Abbreviations: PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire; HRQOL, Health-related quality of life; CTCAE, Common Terminology Criteria for Adverse Events

*See eTable 3 for all computed Pearson correlations between PRO-CTCAE items and EORTC QLQ-C30 functioning, global, and symptom scales.

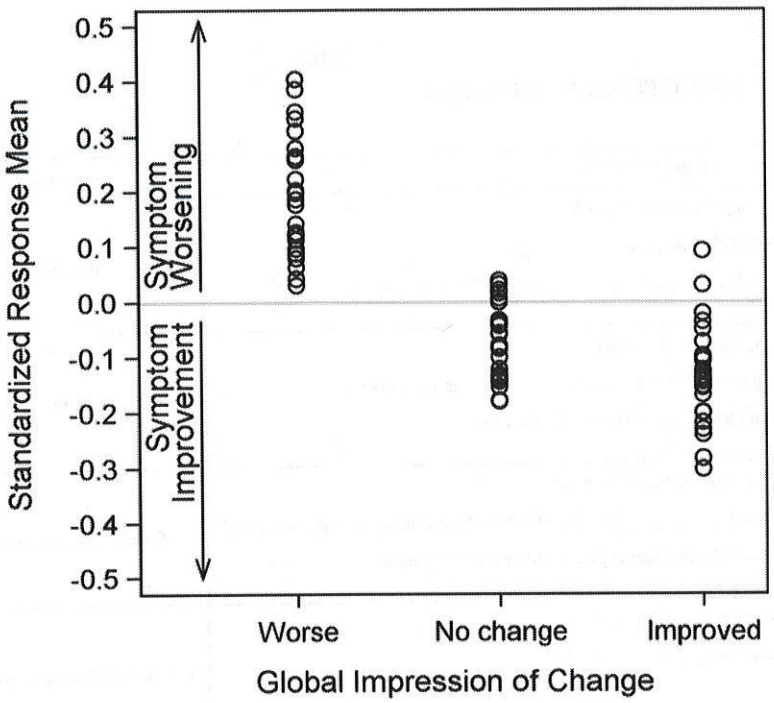



**Figure 2. Standardized Response Means across 27 PRO-CTCAE Items by Patient-Reported Global Impression of Change Category***

Abbreviation: PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events

*Figure 2 includes 27 frequency, severity, and interference items selected prior to initiation of the responsiveness analysis. The set of 20 "core" symptomatic AEs was reviewed and symptomatic AEs were selected if they had high potential to be meaningfully related to global changes in quality of life, physical condition, and/or emotional state (i.e., the Global Impression of Change items which were administered at the second visit). Of the 20 reviewed symptomatic AEs, 13 were included based on this criterion (see eTable 2). The symptomatic AEs which were excluded were felt to be related to initiation or changes in specific treatments (dry mouth, problems with tasting food/drink, rash) so may not exhibit change in a heterogeneously treated sample of patients; require a longer duration of follow-up to exhibit change (arm/leg swelling, hair loss); or be related to cognitive condition (headache, problems with concentration) which was not assessed in the Global Impression of Change items.

## Table 1

### PRO-CTCAE Item Formats*

| Please think back over the past 7 days: | Example |
|---|---|
| **Frequency** (25 symptomatic AE terms):<br>How **OFTEN** did you have _____?<br>Never / Rarely / Occasionally / Frequently / Almost constantly | Vomiting |
| **Severity** (51 symptomatic AE terms):<br>What was the **SEVERITY** of your _____ at its WORST?<br>None / Mild / Moderate / Severe / Very severe | Pain |
| **Interference** (25 symptomatic AE terms):<br>How much did _____ **INTERFERE** with your usual or daily activities?<br>Not at all / A little bit / Somewhat / Quite a bit / Very much | Sudden urges to urinate |
| **Presence** (21 symptomatic AE terms):<br>Did you have any _____?<br>No / Yes | Unusual darkening of the skin |
| **Amount** (2 symptomatic AE terms):<br>Did you have any _____?<br>Not at all / A little bit / Somewhat / Quite a bit / Very much | Hair loss |

Abbreviations: PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; AE, adverse event

*See Basch et al.[15] for a complete listing of PRO-CTCAE items.

**Table 2**

**Patient Characteristics (N=940)**

| Characteristic | No. | % |
|---|---|---|
| Age at enrollment | | |
| Median | 59 | |
| Range | 19 - 91 | |
| Age group | | |
| <30 | 23 | 2.5% |
| 30-64 | 597 | 63.5% |
| 65-74 | 235 | 25.0% |
| ≥75 | 85 | 9.0% |
| Gender | | |
| Female | 539 | 57.3% |
| Male | 401 | 42.7% |
| Ethnicity | | |
| Hispanic or Latino | 56 | 6.0% |
| Not Hispanic or Latino | 832 | 88.5% |
| Missing | 52 | 5.5% |
| Race | | |
| White | 675 | 71.8% |
| Black or African American | 203 | 21.6% |
| Asian | 42 | 4.5% |
| Other or multiple races reported | 8 | 0.9% |
| Missing | 12 | 1.3% |
| Education | | |
| High school or less | 305 | 32.4% |
| Some college | 199 | 21.2% |
| College graduate or more | 415 | 44.1% |
| Missing | 21 | 2.2% |
| Cancer type | | |
| Lung, head or neck | 329 | 35.0% |
| Breast | 260 | 27.7% |
| Genitourinary or gynecologic | 172 | 18.3% |
| Gastrointestinal | 95 | 10.1% |
| Hematologic | 47 | 5.0% |
| Other or unknown | 37 | 3.9% |
| ECOG Performance Status at first visit | | |
| 0-1 | 779 | 82.9% |
| 2-4 | 161 | 17.1% |

| Characteristic | No. | % |
|---|---|---|
| Cancer treatment in prior two weeks | | |
| Chemotherapy | 522 | 55.5% |
| Radiation | 424 | 45.1% |
| Surgery | 35 | 3.7% |

Abbreviation: ECOG, Eastern Cooperative Oncology Group

VOLUME 32 · NUMBER 14 · MAY 10 2014

JOURNAL OF CLINICAL ONCOLOGY

REVIEW ARTICLE

# What Is the Value of the Routine Use of Patient-Reported Outcome Measures Toward Improvement of Patient Outcomes, Processes of Care, and Health Service Outcomes in Cancer Care? A Systematic Review of Controlled Trials

*Grigorios Kotronoulas, Nora Kearney, Roma Maguire, Alison Harrow, David Di Domenico, Suzanne Croy, and Stephen MacGillivray*

Grigorios Kotronoulas, Nora Kearney, Roma Maguire, University of Surrey, Guildford; Alison Harrow, Dundee Cancer Centre; David Di Domenico, Stephen MacGillivray, University of Dundee, Dundee; Suzanne Croy, Dementia Services Development Centre, University of Stirling, Stirling, United Kingdom.

Published online ahead of print at www.jco.org on April 7, 2014.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Grigorios Kotronoulas, PhD, MSc, BSN, School of Health and Social Care, University of Surrey, Duke of Kent Building, Guildford, Surrey, GU2 7TE, United Kingdom; e-mail: g.kotronoulas@surrey.ac.uk.



0732-183X/14/3214w-1480w/$20.00

DOI: 10.1200/JCO.2013.53.5948

## ABSTRACT

### Purpose
The systematic use of patient-reported outcome measures (PROMs) has been advocated as an effective way to standardize cancer practice. Yet, the question of whether PROMs can lead to actual improvements in the quality of patient care remains under debate. This review examined whether inclusion of PROM in routine clinical practice is associated with improvements in patient outcomes, processes of care, and health service outcomes during active anticancer treatment.

### Methods
A systematic review of five electronic databases (Medline, EMBASE, CINAHL [Cumulative Index to Nursing and Allied Health Literature], PsycINFO, and Psychology and Behavioral Sciences Collection [PBSC]) was conducted from database inception to May 2012 to locate randomized and nonrandomized controlled trials of patients receiving active anticancer treatment or supportive care irrespective of type of cancer.

### Results
Based on prespecified eligibility criteria, we included 26 articles that reported on 24 unique controlled trials. Wide variability in the design and use of interventions delivered, outcomes evaluated, and cancer- and modality-specific context was apparent. Health service outcomes were only scarcely included as end points. Overall, the number of statistically significant findings were limited and PROMs' intervention effect sizes were predominantly small-to-moderate.

### Conclusion
The routine use of PROMs increases the frequency of discussion of patient outcomes during consultations. In some studies, PROMs are associated with improved symptom control, increased supportive care measures, and patient satisfaction. Additional effort is required to ensure patient adherence, as well as additional support to clinicians who will respond to patient concerns and issues, with clear system guidelines in place to guide their responses. More research is required to support PROM cost-benefit in terms of patient safety, clinician burden, and health services usage.

*J Clin Oncol 32:1480-1501. © 2014 by American Society of Clinical Oncology*

## INTRODUCTION

Anticancer treatments have brought about definite advances in patient survival rates.[1] However, treatment is associated with significant toxicity that is potentially life-threatening,[1] and can often result in poor treatment adherence, impaired quality of life (QoL), and mortality.[2,3] Systematic monitoring is crucial to detect problems, to address needs of patients, and to plan care.[4] Using patient-reported outcome measures (PROMs), "measurements of any aspect of a patient's health status that come directly from the patient,"[5] facilitates a systematic and comprehensive approach to patient assessment and identifies problems that are often overlooked within routine practice. Regularly collecting PROM data is an effective way to standardize practice and improve patient management.[4] Nevertheless, the question of whether PROMs can improve the quality of patient care, and whether this relates both to health professional engagement with them and to the system guidelines in place to guide response, remains under debate. Given the costs associated with collecting PROMs, evidence of their effect on patient outcomes (POs), processes of care (PoCs), and/or health service outcomes (HSOs) is needed.

Previous reviews have concluded some clinically meaningful, but not always statistically significant, effects on the use of PROMs in clinical practice.[5-11] Only two of these reviews[9,11] were specific to cancer care and differed in terms of objectives, comprehensiveness, and quality. Taking into consideration the lack of clarity around the use of PROMs in cancer care, we conducted a comprehensive systematic review of all available controlled trials (CTs) to examine whether routine use of PROMs by health care professionals (HPs) can improve the quality of care patients receive during active anticancer treatment. The value of PROM use was examined through detection of positive effects on POs, PoCs, and HSOs, as suggested by statistical/clinical changes.

## METHODS

We searched five electronic databases (Medline, EMBASE, CINAHL [Cumulative Index to Nursing and Allied Health Literature], PsycINFO, and PBSC) from database inception to May 2012, using a systematic strategy that was devised and refined through an iterative process (Appendix Table A1 [online-only]). Additional articles were identified through previous topical reviews.[5-11] We also examined reference lists of the articles retained for any studies that might have been overlooked. Where necessary, we contacted study authors to provide clarification on characteristics of the study samples included. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines where applicable.[12]

### Study Selection Criteria

Trials were deemed eligible if they were primary or secondary reports of CTs testing PROM interventions in which PROM-generated feedback was made available to HPs or patients to improve quality of patient care; involved adult patients (> 18 years old) with cancer, irrespective of disease stage, who received any type of active anticancer treatment or supportive care, even if only part of the sample received active treatment/care but percentages were reported; were randomized CTs (RCTs) or non-RCTs; and were published in the English language with readily available abstracts. Trials were excluded if they evaluated PROMs as part of broader psychobehavioral interventions, in which PROMs were only used to evaluate intervention effectiveness; investigated the effects of a medicinal product; were conducted with survivors of cancer who were not actively receiving anticancer treatment; tested the psychometric properties of PROMs; or involved children with cancer, or survivors of childhood cancers.

### Study Selection and Data Extraction Procedures

Study selection involved two stages: an initial title and abstract screening with eligibility evaluation performed by two screening groups that independently screened the retrieved records against selection criteria, and retrieving potentially eligible full-text articles, which were independently evaluated for eligibility by five reviewers. Selection of the final sample of studies was discussed until a consensus was reached. Five reviewers extracted data using forms that were specifically developed for this review, pilot tested the forms on three randomly selected studies, and refined the forms accordingly.

### Risk of Bias and Methodologic Quality Evaluation

We used the Cochrane Collaboration Risk of Bias Tool[13] to evaluate six different domains of a CT: adequacy of sequence generation, concealment of allocation, blinding, completeness of follow-up, freedom from reporting bias, and other forms of bias. We evaluated each domain of bias as low risk, high risk, or unclear. Three reviewers assessed five articles each, and a fourth reviewer cross-checked the evaluations until a consensus was reached. Reviewers were not blinded to authors, institutions, or journals of publication.

### Outcome Evaluation

Based on previous topical reviews,[5-11] three major outcome categories were formed: POs (ie, health status/well-being/functioning; symptom burden/distress; health-related QoL; psychological distress), PoCs (ie, patient satisfaction with treatment/care/consultation; patient behaviors/actions/adherence; patient-HP communication; patient-HP concordance in assessments; HP engagement in assessment), and HSOs (ie, patient safety; cost-effectiveness; number of contacts with clinicians; patient resources/services use). We anticipated that not all CTs would report on every outcome category or on every outcome within a specific category.

### Synthesis of Results and Determination of Effect Sizes

Individual outcomes were classified according to prespecified major outcome categories, and findings were narratively synthesized. Prevalence (%) of studies examining each individual outcome and major categories was examined and plotted. Because of variability in the patient populations, outcomes assessed, outcome PROMs used, and reporting of results, we deemed a meta-analysis was not feasible. However, where enough data were available, effect sizes (ES; Cohen's $d$) and 95% CIs were estimated based on mean postintervention total scores of outcome measures or percentages of patients reporting specific outcomes based on specific formulas.[14,15] By convention, ES where $d \geq 0.2$ were considered small, $d \geq 0.5$ were moderate, and $d \geq 0.8$ were large.[16]

## RESULTS

### Search Results and Study Characteristics

Initial searches retrieved 4,997 references from electronic databases and 18 from previous published literature reviews.[5-11] Twenty-six articles[17-42] reporting on 24 unique CTs fulfilled eligibility criteria and were included in a qualitative synthesis (Fig 1). All but four trials[18,24,34,36] were RCTs, and 16 adopted a longitudinal study design (Table 1). Patient study samples varied widely in size (median, 194 individuals; range, 48 to 1,134 individuals; for a total of 6,279 individuals). HP samples varied similarly (median, 22 HPs; range, four to 262 HPs; total, n = 713), but they were reported in only 11 trials. Nine CTs tested interventions designed specifically for patients with breast,[20,22,26,27] lung,[20,29,30,33,34] or hematologic malignancies.[32] Seventeen CTs tested interventions delivered in the outpatient/ambulatory setting. Only two RCTs targeted patients with early-stage cancers.[19,22] Thirty-seven percent to 100% of patients were receiving active anticancer treatments during study participation, and these treatments were most frequently chemotherapy or radiotherapy.

In terms of intervention design, patients in the control group either received usual care only[19,21,28,34,36,41,42] or completed PROMs similar to that of the experimental group, but feedback remained unavailable to HP.[17,18,24,26,30-33,37,40] Only one three-arm RCT combined these two alternative conditions in the same design.[35,38,39] In the more diverse CTs, PROMs were completed at home by the experimental group but were not administered to patients in the control group[25,29]; were completed by all participants, but PROM summaries of the experimental group were only placed in the medical records or sent to HPs[20,23]; or were completed by patients in the experimental group only to direct further intervention based on distress expressed by a subset of the group.[20,22,27] In only five CTs did HPs follow specific guidelines to guide response to PROM feedback.[20,22,23,26,28]

Twenty-nine PROMs were administered in the reviewed trials to help deliver the interventions (Appendix Table A2). Eleven CTs relied on only one intervention PROM, seven incorporated two PROMs, and six CTs used three or more instruments.[17,18,23,24,28,42] The most frequently used PROM was the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC QLQ-C30; n = 11). Other PROMs focused on symptom prevalence and severity (n = 11), supportive care needs (n = 8), QoL issues (n = 5), or sources of distress (n = 3). The PROMs were administered on media including electronic platforms (n = 11), paper-and-pencil tools in clinic (n = 12), take-home log books (n = 3), and mailed assessments and/or telephone interviews (n = 7; Table 1).

### Risk of Bias Within and Across Studies

Two RCTs were rated as low risk in five of the seven bias categories.[26,29] Yet, bias in the design and/or reporting was present in all of the included trials (Table 2), regardless of whether patients were randomly assigned to the study



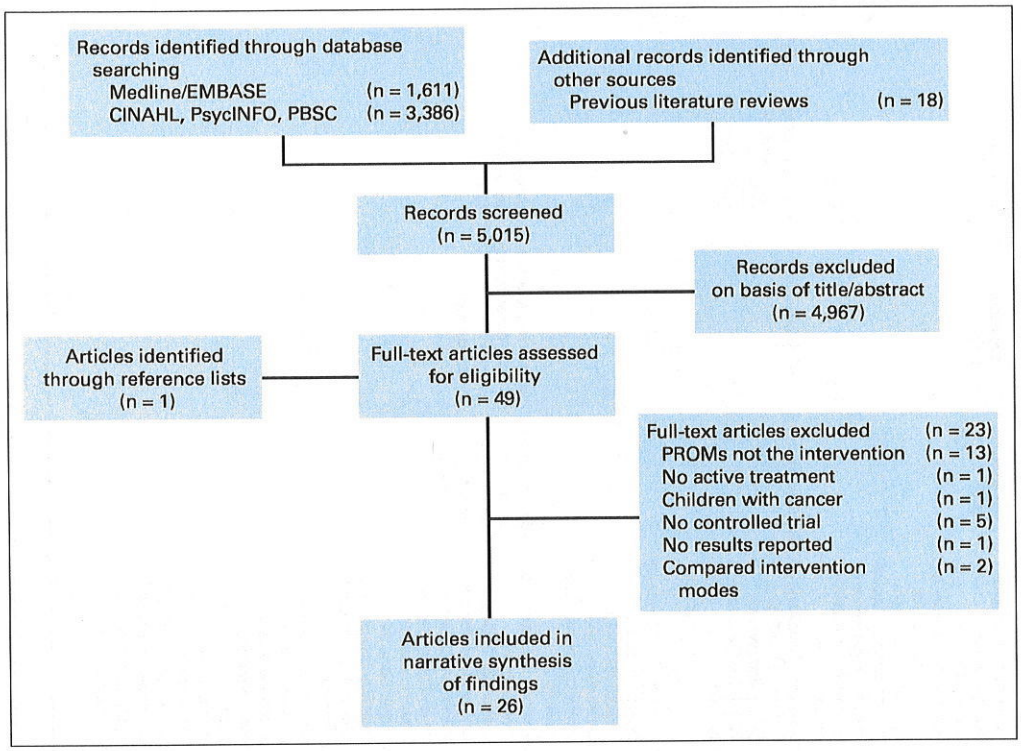



**Fig 1.** Diagram of the study selection process according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[12,43] CINAHL, Cumulative Index to Nursing and Allied Health Literature; PBSC, Psychology and Behavioral Sciences Collection; PROMs, patient-reported outcome measures.

condition. Only seven RCTs were rated as low risk on both the random-assignment generation process and allocation-concealment bias.[17,20,26-29,41] Conversely, all non-RCTs were consistently rated as high risk. With the exception of three RCTs,[21,22,40] performance bias was rated as high for all CTs given that blinding on the HP level was not feasible. With the exception of seven CTs,[18,25,28-30,40,41] risk of detection bias was also deemed high or unclear. Ten CTs were rated as high risk regarding attrition-related bias.[18-20,24,27,33,35,38-42] Selective outcome reporting bias was predominantly unclear (n = 18; 75%). Additional sources of bias interfered with 15 CTs. Most frequently, authors were unclear as to whether HPs who received patient feedback actually used it during consultations.

### Outcomes Evaluation

POs and/or PoCs were reported as primary outcomes in 21 CTs (87.5%) and 19 CTs (79.2%), respectively; however, intervention effects on HSOs were only scarcely investigated (Table 2 and Table 3).[20,22,27,30,42] Eighteen CTs evaluated the effects of interventions in the long term (> 8 weeks), with follow-up assessments ranging in number from two to four or more that were conducted for up to 12 months (but mainly ≤ 6 months) after baseline assessment.

### Patient Outcomes

*Physical symptoms.* Overall, positive effects with reduced symptom prevalence or severity were reported in seven CTs (six RCTs), mainly clinically and less frequently statistically significant. ES ranged widely and were mainly small-to-moderate in terms of intervention effects on physical symptom prevalence ($d$ = 0.01 to 0.75), physical symptom severity ($d$ = 0.0 to 0.44), psychological symptom prevalence ($d$ = 0.07 to 0.15), psychological symptom severity ($d$ = 0.01 to 0.30), or psychological symptom distress ($d$ = 0.09 to 0.42; Appendix Table A3). Across CTs, patients in the experimental group reported greater reductions in symptom-threshold events and symptom interference with functioning,[40] severity of menopausal symptoms and sexual dysfunction,[22] frequency of constipation and vomiting,[25] incidence of pain[37] or fatigue,[41] debilitating symptoms,[18] and distress associated with symptoms/problems[32,41] compared with those in the control group, irrespective of cancer type or stage.

*Quality of life.* Survivors of breast cancer,[22] patients with nonlocalized breast cancer or colorectal cancer,[23] and groups of patients with mixed cancer diagnoses at an advanced stage[21,31,42] or at various clinical stages[24,28] had no significant postintervention effects in nine CTs (Table 2; Appendix Table A3). In terms of overall QoL, ES ranged from 0.04 to 0.59, but were mainly small in magnitude. Nevertheless, rates of diseased QoL were reduced in women with breast cancer 6 months after surgery in the experimental group compared with the control group ($d$ = 0.35).[26] Among patients with lung cancer, QoL scores deteriorated in the experimental group more than in the standard-care group over the 16 weeks of observation.[29] Velikova et al[38] reported improvements in patient QoL scores at treatment initiation that were influenced by whether QoL was actually discussed during consultations.[38]

*Psychological symptoms.* Results were generally unsupportive of significant postintervention effects on anxiety and/or depression regardless of whether direct real-time[18,23] or indirect[20] patient feedback was made available to HPs. This was evident despite overall reductions in psychological distress over time.[27] Similarly, McLachlan et al[28] found no overall intervention effects on depression scores, but the subgroup of patients classified as moderately or severely depressed benefitted more from the intervention. Where significant improvements in anxiety or depression were reported,[42] these were small-to-moderate in magnitude ($d$ = 0.15 to 0.42) and not universal across all assessment PROMs.

*Supportive care needs.* Five CTs provided generally unclear evidence; despite some small-to-moderate ES ($d$ = 0.16 to 0.58) across domains of need, these were not always in favor of the experimental group (Appendix Table A3). The PROM intervention was no better than usual care in tackling needs of patients in two trials.[18,23] We found statistically significant between-group differences in 13 of 19 categories of perceived need[32] and sexual health concerns ($d$ = 0.49)[22] in favor of the experimental group among patients with hematologic malignancies[32] and breast cancer,[22] respectively. In a non-RCT, patients receiving routine psychological screening reported more psychological, information, and physical/daily living needs, but not sexuality needs, at 6 months postbaseline compared with the unscreened cohort.[36]

### Processes of Care

*Medical decisions made/advice given/changes in treatment/referrals made.* Despite being the outcomes most frequently investigated (Table 3), evidence

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Berry et al,[17] 2011 | Outpatient clinic, US | Mixed cancer diagnoses (type, stage, and time since diagnosis); starting a new medical or radiation treatment regimen; RR, 62%; AR, 20% | 47% MD; 23% RT; 30% SCT | 327 (I); 333 (C) | 262 | Two-arm RCT | Intervention: Completion of intervention PROMs through ESRA-C and summaries available to HPs before consultation. Control: Completion of intervention PROMs through ESRA-C but summaries unavailable to HPs. | Processes of care; health service outcomes | Electronic interactive tool | Short term (same day as consultation visit) | No |
| Boyes et al,[18] 2006 | Outpatient clinic, Australia | Mixed cancer diagnoses (type, stage, and time since diagnosis); attending clinic for the first time; RR, 65%; AR, 40% | 65% SRG; 11% RT; 6% CT; 3% HT; 4% other ATR | 42 (I); 38 (C) | 4 | Pilot longitudinal two-arm non-RCT | Intervention: Completion of touch-screen computer survey before consultation, and summaries available to consultants. Control: Completion of touch-screen computer survey before consultation but summaries unavailable to consultants. | Patient outcomes; processes of care | Electronic interactive tool; paper tool in clinic | Long term (three f/u visits) | No |
| Braeken et al,[19] 2011 | Outpatient RT clinic, the Netherlands | Mixed early-stage cancer diagnoses; before first consultation; scheduled to receive > 10 fractions of RT; RR, 51%; AR, NR | 100% RT | 268 (I); 300 (C) | 14 | Two-arm clustered RCT | Intervention: Completion of intervention PROM before first and last consultation, and reports available to radiotherapists involved in care; the radiotherapists discussed patient needs and referred patients to psychosocial care providers. Control: Care as usual. | Processes of care | Paper tool in clinic | Long term (three assessments after baseline within 12 m) | No |
| Carlson et al,[20] 2010 | Outpatient clinic, Canada | New diagnosis of breast (any stage) or lung cancer (any subtype or stage), attending clinic for the first time; RR, 89%; AR, 24% | 2% SRG; 25% CT; 40% RT; 15% HT; 38% SC | 391 (I); 378 (I); 365 (C) | NR | Longitudinal three-arm RCT | Intervention (full screening): Completion of intervention PROMs, summaries available to patient, and placed on the electronic medical record. Intervention (triage): Same as full screening, plus patients were invited to speak to a member of the psychosocial team; triage and referral options available to those requesting an appointment. Control (minimal screening): Completion of intervention PROM, but no summaries available to patients or placed on the medical record. | Patient outcomes; processes of care; health service outcomes | Electronic interactive tool; telephone or email f/u assessment | Long term (f/u assessment at 3 m) | Yes |

(continued on following page)

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment (continued)

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cleeland et al,[40] 2011 | Home, US | Mixed diagnoses (stage of disease) scheduled for thoracic surgery for primary lung cancer or lung metastases; RR, NR; AR, 21% | 100% SRG | 50 (I); 50 (C) | NR | Two-arm RCT | Intervention: Completion of intervention PROM at home twice a week through a telephone-based interactive voice response system; symptom information in the form of e-mail alert available to advanced practice nurse if symptoms met or exceeded preset severity alerts. Control: Completion of same intervention PROM at home, but no feedback available to clinicians. | Patient outcomes; processes of care | Electronic interactive tool at home; paper-based tool in clinic | Intermediate (seven f/u time points within 4-6 w) | No |
| Detmar et al,[21] 2002 | Outpatient clinic, the Netherlands | Mixed diagnoses of advanced cancer (type and time since diagnosis); having received at least two cycles of palliative CT; RR, 71%; AR, 22% | 100% CT | 114 (I)‡; 100 (C) | 10 | Longitudinal crossover two-arm RCT | Intervention: Completion of intervention PROM and summaries available to both patients and physicians during consultation. Control: Usual care. | Patient outcomes; processes of care | Paper tool in clinic | Long term (three f/u visits after baseline) | Yes |
| Ganz et al,[22] 2000 | Outpatient clinic, US | Breast cancer stage I or II diagnosed between 8 m and 5 y earlier; after completion of adjuvant CT or RT; RR, 77%; AR, 5% | 56% HT | 37 (I); 39 (C) | NR | Longitudinal two-arm RCT | Intervention: Daily completion of intervention PROM for 28 d before baseline; review of information and receipt of individualized intervention for three symptoms: hot flashes, vaginal dryness, and urinary incontinence; f/u assessment. Control: Daily completion of intervention PROM for 28 d before baseline; intervention was provided after f/u. | Patient outcomes; health service use | Take home paper tool/log book; paper tool in clinic | Long term (4-m f/u visit) | Yes |
| Girgis et al,[23] 2009 | Home, Australia | Nonlocalized breast or colorectal cancer within 6 m of initial diagnosis; RR, 32%; AR, 6% | 53% CT; 13% RT; 2% SRG; 31% other ATR | 120 (I); 119 (I); 117 (C) | 122§ | Longitudinal three-arm RCT | Intervention (TCW): CATI using intervention PROMs and summaries available to TCW. Intervention (O/GP): CATI using intervention PROMs and summaries available to O/GP. Control: Usual care. CATI but no summaries provided to HPs. | Patient outcomes; processes of care | Telephone based | Long term (3 and 6 m after baseline) | No |

(continued on following page)

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment (continued)

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Hilarius et al,[24] 2008 | Outpatient clinic, the Netherlands | Mixed cancer diagnoses (type and stage) at the start of CT treatment; RR, 83%; AR, 26.5% | 100% CT | 148 (I); 150 (C) | 10 | Longitudinal sequential two-arm cohort | Intervention: Completion of intervention PROM and summaries available to both patients and nurses during consultation. Control: Completion of intervention PROM, but summaries unavailable to nurses during consultation. | Patient outcomes; processes of care | Electronic interactive tool | Long term (four f/u visits) | Yes |
| Hoekstra et al,[25] 2006 | GP practice and home, the Netherlands | Advanced breast, lung, or GI cancer with a life expectancy of 1-12 m; RR, 89%; AR, 32% | 100% PSC | 69 (I)‡; 77 (C) | 89 | Longitudinal two-arm clustered RCT | Intervention: Weekly self-assessment of physical symptoms at home through use of the intervention PROM. Control: Standard care. | Patient outcomes | Take-home paper tool/log book | Long term (every other month) | NR |
| Kearney et al,[41] 2009 | Home and outpatient clinic, UK | Breast, lung, or colorectal cancer (any stage) at the initiation of a new course of CT treatment (any CT line); RR, NR; AR, 23% | 100% CT | 56 (I); 56 (C) | NR | Longitudinal two-arm RCT | Intervention: Completion of intervention PROM on mobile phone at home on days 1-14 post-CT administration; symptom information available to clinicians in real-time in the form of alerts (amber: mild/moderate severity; red: severe or life-threatening); clinicians contacted the patient within 48-72 h (amber) or 1 h (red). Control: Standard care (written and verbal information). | Patient outcomes | Electronic interactive tool at home; paper-based tool in clinic | Long term (four f/u time points within 12-16 w) | Yes |
| Klinkhammer-Schalke et al,[26] 2012 | Inpatient surgery clinics, Germany | Newly diagnosed breast cancer (any stage) at discharge after initial surgical treatment; RR, 82%; AR, 15% | 100% SRG | 99 (I); 100 (C) | 146 | Longitudinal two-arm RCT | Intervention: Completion of intervention PROM by patient and health status form by physician; profiles available to experts; expert opinion available to coordinating practitioners who arranged QoL therapy consisting of up to five standardized treatments. Control: Completion of intervention PROM, but profiles and expert opinions unavailable to practitioners. | Patient outcomes; processes of care | Electronic interactive tool | Long term (f/u assessments at 3, 6, 9, and 12 m after baseline) | No |

(continued on following page)

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment (continued)

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Kornblith et al,[42] 2006 | Home, US | Breast, colon, or prostate cancer (stages III or IV) within the first 2 m of active treatment; life expectancy of ≥ 12 m; RR, 82%; AR, 62% | 100% ATR | 96 (I); 93 (C) | NR | Longitudinal two-arm RCT | Intervention: Completion of intervention PROMs at home monthly for 6 m through TM in addition to EM; feedback available to oncology nurse if levels of distress above preset cut-off scores. Individualized discussion and treatment recommendation during f/u calls. Control: Standard care and EM only. | Patient outcomes; processes of care; health services outcomes | Telephone based | Long term (two f/u assessments at 6 and 9 m) | NR |
| Maunsell et al,[27] 1996 | Inpatient clinic, Canada | Newly diagnosed breast cancer (any stage) after initial surgical treatment; RR, 89%; AR, 10% | 67% RT; 30% CT; 46% HT | 130 (I); 131 (C) | NR | Longitudinal two-arm RCT | Intervention: Brief psychosocial intervention by social worker postsurgery and f/u screening for psychological distress with intervention PROM; further intervention for highly distressed patients. Control: Brief psychosocial intervention by social worker postsurgery but no f/u screening. | Patient outcomes; processes of care; health service use | Telephone based | Long term (3 and 12 m) | NR |
| McLachlan et al,[28] 2001 | Outpatient clinic, Australia | Mixed cancer diagnoses (type, stage, and time since diagnosis); having attended at least one consultation; RR, 59%; AR, 29% | 26% SC; 32% CT ± RT; 5% other ATR | 296 (I); 154 (C) | NR | Longitudinal two-arm RCT | Intervention: Assessment with intervention, PROM before consultation, and summary immediately available to consultants. Individualized management plan based on patient's responses. Control: Conventional clinical encounter and self-reported information unavailable to consultants. | Patient outcomes; processes of care | Electronic interactive tool | Long term (2 and 6 m after baseline) | NR |
| Mills et al,[29] 2009 | Inpatient clinic, UK | Inoperable lung cancer, any subtype; RR, 51%; AR, 50% | 61% CT; 17% RT; 16% CT plus RT; 6% SC | 57 (I); 58 (C) | NR | Longitudinal two-arm RCT | Intervention: Weekly completion of intervention PROM at home; patients were asked to share information with any HP involved in their care. Control: Usual care. | Patient outcomes; processes of care | Take-home paper tool | Long term (16 w) | NA |

(continued on following page)

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment (continued)

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Nicklasson et al,[30] 2013 | Outpatient clinic, Sweden | Incurable lung cancer (any subtype or stage) or mesothelioma with a life expectancy at the first clinic visit of ≥ 3 m; RR, 75%; AR, 1% | 78% CT; 42% RT; 9% SC | 85 (I); 88 (C) | 22 | Longitudinal two-arm RCT | Intervention: Completion of computerized intervention PROM before consultation; summaries were available to consulting physicians. Control: Completion of paper-and-pencil intervention PROM before consultation, but summaries were unavailable to consulting physicians. | Processes of care; health service outcomes | Electronic interactive tool; paper tool in clinic | Long term (8-12 w of f/u visits) | No |
| Rosenbloom et al,[31] 2007 | Outpatient clinic, US | Advanced breast, lung, or colorectal cancer with a life expectancy of ≥ 6 m during CT treatment; RR, NR, and AR: 28% | 100% CT | 69 (I); 71 (C); 73 (C) | NR | Longitudinal three-arm RCT | Intervention: Assessment with intervention PROM followed by structured interview with treating nurse about patient's responses. Assessment control: Assessment with intervention PROM followed by feedback to treating nurses, but no interview. Full control: Assessment with outcome PROM, but no interview with or feedback to treating nurses. | Patient outcomes; processes of care | Paper tool in clinic | Long term (four f/u visits) | No |
| Ruland et al,[32] 2010 | Inpatient and outpatient clinics, Norway | Newly diagnosed or recurrent hematologic malignancy at the start of treatment; RR, 90%; AR, 19% | 68% CT; 34% SCT | 75 (I); 70 (C) | NR | Longitudinal two-arm RCT | Intervention: Intervention PROM administered during inpatient, outpatient, and all f/u visits. Assessment summaries available to HPs. Control: Intervention PROM administered during inpatient, outpatient, and all f/u visits. Assessment summaries not available to HPs. | Patient outcomes; processes of care | Electronic interactive tool | Long term (≥ four follow-up visits) | NR |
| Sarna,[33] 1998 | Outpatient clinics, US | Advanced lung cancer (any subtype); newly diagnosed; RR, 83%; AR, 56% | 88% CT; 23% RT | 48¶ | NR | Longitudinal two-arm RCT | Intervention: Completion of intervention PROM and summaries available to staff nurses for discussion with the patient. Control: Completion of intervention PROM, but summaries unavailable to staff nurses. | Patient outcomes | Paper tool in clinic | Long term (six assessment points within a 6-m period) | No |

(continued on following page)

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment (continued)

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Taenzer et al,[34] 2000 | Outpatient clinic, Canada | Primary, secondary, or metastatic lung cancer of any stage; an average of 51 m postdiagnosis; RR, 70%; AR, NR | NR% SC; NR% ATR | 27 (I); 26 (C) | NR | Sequential pre- and postscreen, two-arm cohort | Intervention: Completion of intervention PROM before consultation and summaries provided to HPs. Control: Usual care. | Processes of care | Electronic interactive tool (I); paper tool in clinic (C) | Short term (same day as consultation) | No |
| Takeuchi et al,[35] 2011‖ | Outpatient clinic, UK | Mixed cancer diagnoses (type and stage) at the start of treatment; RR, 65%; AR, 37% | 100% ATR | 100 (I); 46 (C); 52 (C) | 28 | Longitudinal three-arm RCT | Intervention: Completion of touch-screen intervention PROMs before clinic visit and feedback available to physicians. Attention-control: Completion of intervention PROMs before clinic visit, but feedback unavailable to physicians. Control: Standard care. | Processes of care | Electronic interactive tool | Long term (four time points within 6 m) | No |
| Thewes et al,[36] 2009 | Rural outpatient clinics; home; Australia | Mixed cancer diagnoses (type and stage); newly diagnosed at the first clinic visit; RR, 81%; AR, 37% | 76% SRG; 66% CT; 53% RT; 33% HT | 43 (I); 40 (C) | NR | Sequential pre- and postscreen, two-arm cohort | Intervention: Completion of intervention PROM and feedback to nursing staff; patient assessment of problems and concerns if score above cutoff score. Control: Usual care; no intervention PROM administered. | Patient outcomes; processes of care | Paper tool in clinic; mailed f/u assessments | Long term (one f/u at 6 m after intervention) | NR |
| Trowbridge et al,[37] 1997 | Outpatient clinic, US | Mixed diagnoses of recurrent or metastatic solid or hematologic cancers or sarcomas; RR, NR; AR, NR | 100% ATR | 160 (I); 160 (C) | 13 | Two-arm RCT | Intervention: Completion of intervention PROM before consultation and summaries provided to consultant; discussion of self-reported information. Control: Completion of intervention PROM before consultation, but summaries unavailable to consultant. | Patient outcomes; processes of care | Paper tool in clinic; mailed assessments | Intermediate (4 w after intervention) | No |
| Velikova et al,[38] 2004‖ | Outpatient clinic, UK | Mixed cancer diagnoses (type and stage) at the start of treatment; RR, 65%; AR, 37% | 76% CT; 21% BT; 2% HT; 1% f/u | 144 (I); 70 (C); 72 (C) | 28 | Longitudinal three-arm RCT | Intervention: Completion of touch-screen intervention PROMs before clinic visit and feedback available to physicians. Attention-control: Completion of intervention PROMs before clinic visit, but feedback unavailable to physicians. Control: Standard care. | Patient outcomes; processes of care | Electronic interactive tool | Long term (four time points within 6 m) | No |

(continued on following page)

**Table 1.** Summaries of the Methodologic Characteristics of the 24 Studies (26 articles) Reporting on the Use of PROMs As Interventions in Patients With Cancer Receiving Active Anticancer Treatment (continued)

| Author and Year of Study | Setting/Location | Patient Population | Type of Treatment* | No. of Patients† | No. of HPs | Study Design | Intervention/Control | Outcomes Assessed | Method of Administration of PROM | Evaluation of Effects | Patient Received PROM Feedback |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Velikova et al,[39] 2010‖ | Outpatient clinic, UK | Mixed cancer diagnoses (type and stage) at the start of treatment; RR, 65%; AR, 37% | 76% CT; 21% BT; 2% HT; 1% f/u | 144 (I); 70 (C); 72 (C) | 28 | Longitudinal three-arm RCT | Intervention: Completion of touch-screen intervention PROMs before clinic visit and feedback of results available to physicians. Attention-control: Completion of intervention PROMs before clinic visit, but feedback unavailable to physicians. Control: Standard care. | Processes of care | Electronic interactive tool | Long term (four time points within 6 m) | No |

Abbreviations: AR, attrition rate; ATR, active treatment; BT, biological therapy; C, control; CATI, computer-assisted telephone interview; CT, chemotherapy; d, days; EM, educational materials; ESRA-C, Electronic Self-Report Assessment-Cancer; f/u, follow-up; GP, general practitioner; h, hours; HP, health care professional; HT, hormonal therapy; I, intervention; m, months; MD, medical; NA, not applicable; NR, not reported; O, oncologist; PROM, patient-reported outcome measure; PSC, palliative supportive care; QoL, quality of life; RCT, randomized controlled trial; RR, response rate; RT, radiotherapy; SC, supportive care; SCT, stem-cell transplantation; SRG, surgery; TCW, telephone caseworker; TM, telephone monitoring; UK, United Kingdom; US, United States; w, weeks; y, years.

*Percentages valid for total sample.

†Sample sizes of patients as randomly assigned (RCTs) and consented (non-RCTs) at baseline.

‡Physicians, rather than patients, were randomly assigned.

§Estimated as the total of 3 TCWs and 119 O/GPs.

¶Group sizes were not reported.

‖Articles are based on data from the same study; different sample sizes and outcomes are evaluated in each article.

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| RCTs | | | | | | | | | | | |
| Berry et al,[17] 2011 | Intervention effects depended on whether a symptom/QoL issue was reported at threshold ($P = .03$). When reported at threshold, the intervention resulted in a 29% increase in the odds of the issue being discussed compared with the CG. This was evident for concentration, cognitive function, impact on sexual activities and interest, and social function. | No EG/CG differences ($P = .35$) for the average length of clinic visits. Clinicians agreed that the intervention was useful in identifying appropriate symptom/QoL issues (67.8%), guiding the interview (64.3%), promoting communication (50%), and identifying appropriate areas for referral (53.6%). | — | No | Low | Low | High | Unclear | Low | High | High |
| Braeken et al,[19] 2011 | — | No significant intervention effects were observed for the total No. of patients referred to psychosocial care providers at 3 ($P = .32$), 9 ($P = .22$), or 12 m ($P = .44$). More patients in the EG brought up their need for psychosocial care during consultation ($P = .04$). EG were referred to social workers at an earlier stage than CG ($P < .01$). No significant intervention effect on improving patient-clinician communication about psychosocial problems; no effect on patients' satisfaction with communication with clinicians. | — | No | Unclear | Unclear | High | High | High | High | Low |
| Carlson et al,[20] 2010 | Only a marginally significant main effect of study condition at follow-up for distress scores ($P = .09$). Significantly fewer patients in the triage group (36%) exceeded the distress cut off *v* 46% and 48.7% in full screening and minimal screening groups, respectively ($P = .005$). No EG/CG differences in anxiety or depression scores at 3 m overall or within either the lung or breast groups. | No differences between study conditions in referrals made to psychosocial care ($P = .05$) before or after follow-up. Receiving a referral was linked to less improvement on the distress score. | No differences between full screening and minimal screening in patient self-referrals (14.3% *v* 10.3%). | Yes | Low | Low | High | High | High | Unclear | Low |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| Cleeland et al,[40] 2011 | EG significantly greater reduction in overall symptom threshold events during the 4-week trial period (19% *v* 8%; $P = .003$). Symptom threshold events for pain, distress, disturbed sleep, shortness of breath, and constipation were more in the CG at week 4. No EG/CG differences in mean symptom severity changes at the end of 4 weeks. Greater reduction in mean symptom interference over time in EG ($P = .02$). | EG significantly more comfortable with intervention ($P < .03$) and more likely to rate the intervention as easy to use ($P < .01$) compared with CG. Both groups expressed satisfaction with the intervention and agreed for it to be used in routine clinical practice. | — | No | Unclear | Low | Low | Low | High | Unclear | High |
| Detmar et al,[21] 2002 | No EG/CG differences at the fourth visit for any of the QoL scales. A significantly greater percentage of patients in the EG *v* the CG exhibited improvement over time in mental health (43% *v* 30%; $P = .04$) and role functioning (22% *v* 11%; $P = .05$). | Ten of 12 QoL issues were discussed more frequently in the EG, especially social functioning, fatigue, and dyspnea ($P < .05$). No EG/CG differences in exact or global physician-patient agreement, or in mean number of QoL-related patient management actions taken per patient. Patient and physician satisfaction was high in both groups. No differences in mean duration of visits. In the EG, the QoL summary profile provided an accurate picture of patient functioning and well-being (97%), and it would be useful as a standard part of the outpatient clinic procedure (87%). | — | No | Unclear | Unclear | Unclear | High | Low | Unclear | High |
| Ganz et al,[22] 2000 | Change scores for menopausal symptoms ($P < .001$) and sexual functioning ($P = .02$) differed significantly between groups, with EG reporting fewer severe symptoms and better sexual functioning at follow-up. No EG/CG differences in terms of vitality ($P = .77$). | — | Women in both the EG and CG sought out additional information about their symptoms, at about the same rate. A similar percentage of women in each group received some form of psychological referral. Women in the EG used medications more frequently. | Yes | Unclear | Unclear | Unclear | Unclear | Low | Unclear | High |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias n the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| Girgis et al,[23] 2009 | No overall intervention effect was observed. Physical functioning was significantly improved at the third telephone interview for participants in the telephone caseworker group ($P = .01$) and there was a trend toward fewer participants with unmet needs ($P = .07$). | Patients in the telephone caseworker group were more likely to have indicated issues of need discussed ($P < .001$), referrals made ($P < .001$), and strong agreement that the intervention improved communication with their health care team ($P < .001$). | — | Yes | Low | Unclear | High | High | Low | Unclear | High |
| Hoekstra et al,[25] 2006 | At the 2-m follow-up, the prevalence of symptoms was lower in the EG (prevalent differences 2.1%–24.3%) for nine of 10 symptoms (except coughing). Constipation and vomiting were significantly less prevalent in EG. Severity of fatigue, lack of appetite, shortness of breath, and nausea was lower in the EG (not significant). No EG/CG differences in severity of pain, coughing, sleeplessness, and diarrhea). | — | — | No | High | Unclear | High | Low | Low | Unclear | High |
| Kearney et al,[41] 2009 | CG had significantly more reports of fatigue ($P = .04$) and significantly fewer reports of hand-foot syndrome ($P = .03$) than EG. No EG/CG differences in reports of vomiting/nausea, diarrhea, or sore mouth/throat. No EG/CG differences in severity and distress of symptoms, with the exception of higher severity ($P = .03$) and distress ($P = .03$) of hand-foot syndrome in EG. | — | — | No | Low | Low | High | Low | High | Low | High |
| Klinkhammer-Schalke et al,[26] 2012 | At 6 m, 71% of patients in CG showed diseased QoL in at least one dimension. In the EG, this occurred in 56% of patients ($P = .048$). Relative risk was reduced 21% (95% CI, 0 to 37) and absolute risk was reduced 15% (95% CI, 0.3 to 29). The No. of diseased QoL dimensions per patient was lower in the EG at 6 m ($P = .035$). The percent of patients with zero QoL in at least one dimension at 6 m was 15% in the EG and 25% in the CG ($P = .124$). | At 3 m, coping strategies were applied more often but not significantly more in the EG than the CG ($P = .055$). Significantly more psychotherapy was given to women in the EG ($P = .005$) but the opposite was true for physiotherapy in the CG. At 6 m, the results were much more similar in the EG and CG. | — | Yes | Low | Low | High | High | Low | Low | Low |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| Kornblith et al,[42] 2006 | EG had significantly lower anxiety and depression at 6 m ($P < .001$). No differences on psychological distress, QoL, or comorbidities interfering with functioning. Significantly more patients in the EG had scores above cut off for depression/anxiety at 6 m (42% *v* 24%; $P = .041$). | No significant EG/CG differences in percent of physical symptom alerts. No differences in overall satisfaction with intervention (good/excellent, 88% *v* 74%). Significantly fewer patients in the EG rated the intervention very/extremely helpful in coping with an important problem ($P = .018$). | No EG/CG differences in use of mental health services at 6 m (9% *v* 12%). | No | Unclear | Unclear | High | High | High | Unclear | Unclear |
| Maunsell et al,[27] 1996 | Participants' psychological distress levels decreased over the study period ($P < .001$), but no EG/CG differences. No EG/CG differences in physical health, functional status, social and leisure activities, return to work, or marital satisfaction. | No EG/CG differences in the mean No. of social worker contacts. CG and EG were very similar in total intervention time, proportion of contacts conducted in person, and mean duration of contacts conducted in person and by telephone during the baseline period. Use of psychosocial services, medical consultations, or other patient initiatives that might improve quality of life did not differ between groups. | The mean No. of visits was 2.4 and 6.1 among CG and EG patients, respectively, representing 48.9 and 119.6 min of social worker contact. | No | Low | Low | High | Unclear | High | Unclear | High |
| McLachlan et al,[28] 2001 | No EG/CG differences in changes in psychological or health information needs, QoL, or psychosocial functioning between the baseline and follow-up assessments. For the subgroup of moderately/severely depressed patients, there was a significant reduction in depression for the EG relative to the CG at the 6-m assessment ($P = .001$). | No EG/CG differences ($P = .36$) in consultation times (17.7 min *v* 16.4 min) or levels of satisfaction ($P > .05$). For CG *v* EG patients, the percent of patients indicating their level of satisfaction was 95% *v* 98% for nursing care, 98% *v* 98% for medical care, 91% *v* 96% for information received about their illness and treatment, and 98% *v* 99% for overall satisfaction with the care received. | — | Yes | Low | Low | High | Low | Low | Unclear | High |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Selection Bias: Random Sequence Generation | Selection Bias: Allocation Concealment | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| Mills et al,[29] 2009 | Only a small but consistent difference in QoL was found between EG and CG. The EG had a poorer QoL in many domains. Two different QoL summary scores indicated a statistically significant between-group difference. | Only 23% of the diary group stated that they had shared their diary with any health professional. No intervention effects in communication, satisfaction with care, or the discussion of patient problems. EG discussed fewer topics with health professionals than the CG (not significant). Both groups reported high levels of satisfaction with their care, with no significant associations identified. | — | No | Low | Low | High | Low | Low | Unclear | Low |
| Nicklasson et al,[30] 2013 | — | Issues regarding emotional functioning were more frequently discussed in the EG by doctors or patients taken together ($P = .015$). No EG/CG differences in physical/role, social, cognitive functioning, or global health. Pain, dyspnea, fatigue, and anorexia were somewhat more frequently discussed in the EG (not significant). Medical/technical statements were more frequently raised in the CG ($P < .05$). Length of doctor-patient conversations was similar in the EG and CG ($P = .77$). No. of diagnostic and therapeutic interventions for emotional and social concerns was higher in the EG. | Planned outpatient visits were similar between EG and CG (327 *v* 323) | No | Unclear | High | High | Low | Low | Unclear | Low |
| Rosenbloom et al,[31] 2007 | No statistically significant differences across the three study conditions in QoL over time ($P > .05$). For all patients, QoL essentially did not change over the course of the study. | No significant differences across the three study conditions in general satisfaction and satisfaction with communication over time ($P > .05$). For all patients, satisfaction essentially did not change over the course of the study. No significant group differences ($P > .05$) in clinical treatment changes between the three conditions. | — | No | Unclear | Unclear | High | High | Low | Unclear | High |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| Ruland et al,[32] 2010 | Symptom distress in the EG decreased significantly over time in 11 (58%) of 19 symptom/problem categories *v* two (10%) for the CG. | Significantly more symptoms were addressed in the EG patient charts *v* those of the CG. Need for symptom management support over time also decreased significantly more for the EG than the CG in 13 (68%) symptom categories. | — | No | Low | High | High | High | Low | Unclear | High |
| Sarna,[33] 1998 | Symptom distress scores of the CG were higher than scores of the EG ($P < .001$). Chemotherapy status and group assignment were both strong predictors of distress scores. The no-chemotherapy subgroup showed greater levels of distress than the chemotherapy subgroup with the CG and EG groups. | — | — | No | Unclear | Unclear | High | Unclear | High | High | High |
| Takeuchi et al,[35] 2011; Velikova et al,[38] 2004; and Velikova et al,[39] 2010‡ | Patients in the EG and attention-control group had better QoL than the CG ($P = .006$ and $P = .01$, respectively), but the EG and attention-control groups were not significantly different ($P = .80$). A larger proportion of intervention patients showed clinically meaningful improvement in QoL. | More frequent discussion of chronic nonspecific symptoms ($P = .03$) in the EG, without prolonging encounters. No effect on patient management ($P = .60$). Discussion topics were predominantly raised by patients/relatives, regardless of group. Clinic discussions were associated with severity of reported symptoms, but not with patient-reported functional concerns. EG patients rated their continuity of care as better than the CG in terms of communication ($P = .03$). Patients' evaluations of the intervention were positive. | — | No | Unclear | Unclear | High | Unclear | High | Unclear | High |
| Trowbridge et al,[37] 1997 | No significant EG/CG differences in assessments of pain, pain regimens, and relief received at the 4-week follow-up. | Significant EG/CG differences in physicians' patterns of prescribing analgesics (25% *v* 14%; $P = .016$). No significant differences in the percentage of patients undertreated for pain (38% *v* 35%; $P > .05$). | — | No | High | High | High | High | Unclear | High | High |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| Non-RCTs | | | | | | | | | | | |
| Boyes et al,[18] 2006 | Patients in the EG with a debilitating symptom at visit 2 were less likely to report a debilitating symptom at visit 3 compared with CG ($P$ = .04). No EG/CG differences in change in anxiety ($P$ = .09) and depression scores ($P$ = .20). No significant EG/CG differences in change in average No. of moderate or high psychological needs reported over time ($P$ = .82). | For patients, the intervention was easy to complete, and they would be willing to complete the survey each time they visited the oncologist. Only three EG patients reported that their oncologist discussed the feedback report with them. Half of the medical oncologists (n = 2) reported that they discussed the feedback directly with their patients. | — | No | High | High | High | Low | High | Unclear | Low |
| Hilarius et al,[24] 2008 | No significant effects were found in changes in QoL over time. | QoL-related topics discussed more frequently in the EG ($P$ = .02). Nurses' awareness of patients' levels of daily activity, pain, and overall QoL was significantly better in the EG. The mean No. of QoL-related notations in the medical records was higher in the EG ($P$ < .05). Modest effects were observed in patient management; no significant effects in patient satisfaction over time. | — | No | High | High | High | High | High | Unclear | Unclear |
| Taenzer et al,[34] 2000 | — | In the EG, more QoL issues identified by the patient were addressed during the clinic appointment than in the CG ($P$ = .01). Marginally more categories were charted and a trend toward more actions being taken was recorded in the EG. Patients reported being equally and highly satisfied regardless of study group. | — | No | High | High | High | Unclear | Low | Unclear | High |

(continued on following page)

**Table 2.** Main Findings and Assessment of Risk of Bias in the 20 RCTs and Four Non-RCTs Identified for This Review (continued)

| Author and Year | Main Study Findings* | | | Guidelines Were Used to Guide Clinician Response | Risk of Bias† | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Selection Bias | | Performance Bias: Blinding Participants and Personnel | Detection Bias: Blinding of Outcome Assessment | Attrition Bias: Incomplete Outcome Data | Reporting Bias: Selective Outcome Reporting | Other Sources of Bias |
| | Patient Outcomes | Processes of Care | Health Service Outcomes | | Random Sequence Generation | Allocation Concealment | | | | | |
| Thewes et al,[36] 2009 | Participants in the screened cohort reported significantly higher levels of overall unmet needs ($P < .001$), psychological needs ($P = .02$), information needs ($P = .02$), and physical and daily living needs ($P = .04$) compared with the unscreened cohort. No differences on sexuality needs. | Screening did not significantly increase the rate of referrals to psychosocial staff of distressed individuals, but reduced time to referral. | — | No | High | Unclear | High | Unclear | Unclear | Unclear | Unclear |

Abbreviations: CG, control group; EG, experimental group; m, months; min, minutes; QoL, quality of life; RCT, randomized controlled trial.
*Information on effect sizes (where calculation was permitted by availability of data) is available in Appendix Tables A2, A3, and A4 (online only).
†Specific explanations for all ratings are available from the authors.
‡Articles are based on data from the same study; different sample sizes and outcomes are evaluated in each article.

**Table 3.** Classification of Study Outcomes According to the Three Prespecified Outcome Categories (n = 24)

| Patient Outcomes | | | Processes of Care | | | Health Service Outcomes | | |
|---|---|---|---|---|---|---|---|---|
| Classification | No. of Studies | % | Classification | No. of Studies | % | Classification | No. of Studies | % |
| Physical symptoms: prevalence and/or severity[18,22,25,32,37,40,41] | 7 | 29.2 | Patient actual use of the intervention PROM[25,29] | 2 | 8.3 | Health services use/ self-referrals[20,22,42] | 3 | 12.5 |
| QoL[21-24,26,28,29,31,38,42] | 10 | 41.7 | Duration of contacts with HPs[17-19,27,28,30,38] | 7 | 29.2 | Contact with HPs[27,30] | 2 | 8.3 |
| Psychological symptoms[18,20,23,27,28,42] | 6 | 25 | Patient engagement in self-care actions[27] | 1 | 4.2 | | | |
| Supportive care needs[18,22,23,32,36] | 5 | 20.8 | Patient outcomes discussed during consultation[17,19,21,24,29,30,34,35,38] | 8 | 33.3 | | | |
| Overall distress[20,31,33] | 3 | 12.5 | HP acceptability/evaluation of intervention[17-19,21,24,38,39] | 6 | 25.0 | | | |
| Overall physical health[27,42] | 2 | 8.3 | Patient satisfaction with care/communication with treating team[19,21,23,24,28,29,31,34,37,39,40] | 11 | 45.8 | | | |
| Working hours[27] | 1 | 4.2 | Patient outcomes addressed in patient records[32,34] | 2 | 8.3 | | | |
| Social support[27] | 1 | 4.2 | Medical decisions made/advice given/changes in treatment/referrals made[19-21,23,24,26,30,31,34,36,37] | 11 | 45.8 | | | |
| Social activity[27] | 1 | 4.2 | HP use of PROM information[38,39] | 1 | 4.2 | | | |
| Physical activity[27] | 1 | 4.2 | HP satisfaction with encounter with the patient[21] | 1 | 4.2 | | | |
| Marital satisfaction[27] | 1 | 4.2 | HP awareness of patient outcomes[21,24] | 2 | 8.3 | | | |
| | | | Patient satisfaction with intervention[19,21,24,36,39,40,42] | 7 | 29.2 | | | |
| | | | Impact of referrals on patient outcomes[20] | 1 | 4.2 | | | |
| | | | Perceived continuity and coordination of care[39] | 1 | 4.2 | | | |
| | | | Timing of referrals[19,36] | 2 | 8.3 | | | |

Abbreviations: HP, health professional; PROM, patient-reported outcome measure; QoL, quality of life.

of intervention effects on actions taken as a result of PROM feedback becoming available to clinicians remains generally ambiguous (Appendix Table A4). No significant intervention effects were reported in the number of patients referred to psychosocial care[19,20,36] or in clinical actions taken.[21,24,31] Although at 3 months after the intervention women with breast cancer in the experimental group were offered counseling and psychotherapy services more often, at 6 months this difference disappeared.[26] When PROMs were used to increase physician awareness of patients' levels of pain, a significant change ($d = 0.41$) in analgesic prescription patterns was found to favor the experimental group.[37] During treatment for chest malignancies, significantly more patients in the experimental group received diagnostic and therapeutic services for emotional and social concerns,[30] but numbers of QoL-related actions taken per patient were similar across study groups.[34]

*Patient satisfaction with care and/or communication with team.* Regardless of study condition, patient remarks on satisfaction with care and/or communication with HPs were generally positive.[19,21,24,28,29,31,34,39,40] Though eight CTs[19,24,28,29,31,34,40] failed to show significant intervention effects (Appendix Table A4). In the studies in which postintervention gains were reported, the positive effects referred to greater satisfaction with emotional support in the palliative chemotherapy context,[21] greater satisfaction with patients receiving follow-up from oncology nurses rather than general practitioners (though differences from usual care were not examined),[23] and enhanced communication with physicians in the outpatient setting compared with standard care.[39]

*Patient outcomes discussed during consultation.* Regardless of patients' cancer type, significant postintervention increases over time in the frequency of discussions pertinent to patient outcomes during consultations were recorded.[35,38] The odds of such outcomes being discussed seemed to depend on whether these were reported at a level indicating a problem.[17] Though emotional problems tend to be discussed more often during consultations in the experimental group,[19] social and sexual functioning issues may be those on which the intervention proves most effective.[17] Still, the overall patient-physician communication may not significantly improve.[19] In the lung cancer population, significantly more symptoms were discussed and addressed during consultations,[34] but intervention effects on QoL discussions fell short of significance. Much greater intervention effects were reported in the context of palliative chemotherapy (Appendix Table A4),[21] regarding overall communication about dyspnea ($d = 0.40$ to $0.77$)[21,24]; social functioning ($d = 0.49$) and fatigue ($d = 0.38$)[21]; and sleep problems ($d = 0.66$), constipation ($d = 0.40$), diarrhea ($d = 0.67$), and cognitive functioning ($d = 0.66$).[24]

*HP acceptability/evaluation of intervention.* Where addressed, intervention acceptability was moderate to high across all CTs (Table 2), with rates of perceived usefulness ranging from less than 50% to 68%. HPs felt obtaining an overall assessment of the patient was more helpful[21,38,39] to identify issues of concern[17,19,21,38] and to guide discussions with patients[17-19,24] rather than in communicating with patients[17,19] and in managing and enhancing the care provided.[18,38] Yet, in two similar CTs, all physicians[21] and nurses[24] agreed that the intervention facilitated patient-clinician communication. The ability of HPs to identify psychosocial concerns[19,21] and address difficult subjects such as sexuality issues[24] was also enhanced. Although actual changes in HP communication styles may not be seen even following the intervention,[19] physicians[21,39] and nurses[24] seem willing to continue using the PROM summary in everyday practice. Nurses significantly more frequently found PROM

interventions beneficial[17] and felt that use of relevant information resulted in more efficient use of their time.[24]

*Patient satisfaction with intervention.* Overall satisfaction with intervention was evident for at least 80% of patients.[40,42] The PROM interventions were seen as easy to use[40] and a useful way for patients to describe their situation[39] and communicate important information to HPs.[19] Patients expressed their willingness to continue using it in routine care.[39,40] However, in the Kornblith et al[42] CT, percentages of patients rating the PROM intervention as very or extremely helpful in coping with an important problem were notably low and favored the control rather than the experimental group (37% *v* 14%; $d = 0.69$). More than 83% of patients regarded the PROM content important for them and its use necessary for all patients receiving treatment.[19,36] Moreover, almost all patients (93%) appreciated having been asked about their emotional well-being during treatment.[36] In the palliative care setting, patients agreed that the summary profile enhanced their physician's or nurse's awareness of their health problems (79% to 89%), and that it would be useful as a standard part of their consultations (87% to 99%).[21,24]

*HP awareness of patient outcomes.* In the context of palliative chemotherapy, no intervention effects were reported on the magnitude of patient-physician agreement about patients' physical, emotional, and social well-being and daily activities ($d = 0.09$ to 0.50; Appendix Table A4).[21] The only exception was greater agreement in ratings of social functioning in the experimental group, but this applied only to the subgroup of patients who reported moderate-to-severe problems.[21] Oncology nurses' awareness of daily activities, pain, and QoL was significantly higher in the experimental group during the fourth patient visit.[24] Positive intervention effects were reported in patient care documentation in the medical records of patients being treated for hematologic malignancies[32] and in the number of QoL issues charted in records of patients with lung cancer.[34]

*Timing of referrals.* One RCT revealed that PROM feedback resulted in significantly earlier postconsultation referral of patients in the experimental versus the control group by an average of three weeks.[19] In a sequential cohort trial of patient-distress screening, average time to referral in the unscreened cohort was 14 days compared with a considerably earlier referral of only 5 days in the screened cohort.[36]

### Health Services Outcomes

Only five CTs explored the effects of the routine use of PROMs on HSOs (Table 3; Appendix Table A5), namely, numbers of patients making use of health services[20,22,42] and frequency of contacts with health professionals.[27,30] Ganz et al[22] reported only minimal use of services after referral to psychosocial care in women with breast cancer; whereas prevalence of cases in which patients sought professional help was similar irrespective of study group among newly diagnosed patients with lung cancer and breast cancer.[20] Among patients with advanced breast, colorectal, or prostate cancer, use of mental health services at 6 months after intervention was equally minimal regardless of study condition ($P = .34$).[42] In terms of frequency of patient-HP contacts, positive intervention effects were found among women with breast cancer[27] but not among patients with chest malignancies.[30]

## DISCUSSION

We found only tentative evidence regarding the effectiveness of PROM interventions to improve the quality of care provided to patients receiving active anticancer treatments. We used strict systematic methods during identification[12] and risk-of-bias appraisal[13] of all trials included here. We included 24 CTs, which investigated a wide range of outcomes, thus producing a disparate set of data and indicating lack of consensus around the role of PROMs and the range of outcome measures in clinical practice. Evidence suggests that, irrespective of the context of chronic illness, the impact of PROMs on POs is weak.[9,44] Where possible, we calculated ES in an attempt to quantify the magnitude of these effects, and our findings indicate inconsistencies in the overall significance (statistical or clinical) and low-to-moderate intervention effectiveness. Importantly, efficacy of the CTs reviewed seems low, confirming findings from previous reviews.[5,9,44]

Contrary to the limited evaluation of HSOs, PoCs were the most frequently investigated outcomes in our sample of trials. Mixed findings emerged regarding medical decisions made or actions taken by HPs as a result of the availability of PROM data. Changes in HP practices fell short of significance and, where such changes were documented,[30,37] the associated ES were still small. It is unclear whether limited referral options, additional subjective HP assessments, or other health care–related factors influenced the use of PROMs in practice. Patient satisfaction with care did not improve significantly, possibly owing to the presence of ceiling effects. Moreover, achievable improvements in patient communication with HPs, especially regarding emotional health issues, were documented, but ES were quite small. Somewhat greater ES can be proposed with regard to the actual discussion of POs during consultations, particularly physical symptoms, but not necessarily around supportive care needs.[19]

Fewer than 30% of the CTs addressed the important question of whether the use of PROM interventions appeals to patients and HPs. Though HPs may view PROMs as useful toward a more comprehensive or systematic assessment, communication is not always enhanced. In addition, there is still limited (albeit positive) evidence about whether HPs wish PROMs to become routine practice. Whether patients can comply with the systematic use of PROMs during treatment and encounters with the clinical team is equally unclear. Despite limited evidence, including electronic systems to enhance data collection and management, as well as use of clinical algorithms to support clinicians in the management of identified areas for intervention, might potentially increase adherence to and acceptability of PROM-enhanced clinical assessments.

Current data also suggest that patient physical symptoms and distress may be more amenable to improvement after PROM interventions than QoL, supportive care needs, or psychological symptoms. Even with the exception of the few studies that examined the use of health services by patients or contacts with HPs, important aspects of an intervention's applicability, such as patient safety or cost-effectiveness and cost-efficiency, are yet to be included as potential end points to encourage policy makers to consider making changes in the way cancer care is provided. Despite this lack of evidence, the Department of Health in England is aiming to extend the use of PROMs in a wider range of conditions in that country's National Health Service,[45] which would include cancer care.

Finally, measurement bias interfering with the effects of PROM interventions documented in this review should also be considered. Arguably, not all tools used in the delivery of interventions were originally developed as PROMs, which might have affected the reliability of reported outcomes and their subsequent interpretation. In addition, the psychometric robustness of the PROMs used to deliver and/or evaluate intervention effects is questionable and might have interfered with its ability to capture the actual magnitude of such effects. Similar comments can be made regarding sources of bias, such as absence of randomization or uncertainty about whether clinicians did use information generated by PROMs during consultations, which may have further affected the trials' internal and external validity and adversely affected credibility of available evidence.

Our search strategy was purposefully inclusive, with an aim to include all relevant literature. However, it was limited to the most common bibliographic databases, as well as to peer-reviewed articles



and reports published in the English language only. In addition, the gray literature was not searched. Owing to the vast heterogeneity in the studies included, a meta-analytic synthesis was not feasible. Unavailability of data also prevented us from calculating ES for some of the included studies. However, such cases were equally distributed across statistically significant and nonsignificant findings or across the different outcome categories; hence, we are confident that the associated reporting bias has not greatly affected our conclusions.

More research is necessary on the effects of PROM interventions on health outcomes across different types of cancers and treatment modalities. The use of PROMs in clinical practice seems to be most effective in increasing patient satisfaction with communication about emotional concerns. Discussion of POs during consultations may increase and, in some studies, is associated with improved symptom control, increased supportive care measures, and patient satisfaction. Additional patient-related outcomes could be usefully addressed in future trials, including perceived self-care self-efficacy, social activity, work limitations, or survival. Patients and HPs are willing to engage in the routine use of PROMs during anticancer treatment. However, it is paramount that PROM intervention implementation is effective and incorporates strategies that increase patient adherence to the actual use of PROMs and HP engagement in the active incorporation of PROM feedback during encounters with patients.[44] Consensus is also required on the standardization of PROMs to be used in future trials. Finally, dedicated research is required to support the cost-effective use of PROMs in clinical practice regarding patient safety, clinician burden, and health-services usage. This is an important area of consideration, particularly in times of increasing demands on health care.

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

*Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.*

**Employment or Leadership Position:** None **Consultant or Advisory Role:** None **Stock Ownership:** None **Honoraria:** None **Research Funding:** Nora Kearney, Philips HealthCare **Expert Testimony:** None **Patents, Royalties, and Licenses:** None **Other Remuneration:** None

## AUTHOR CONTRIBUTIONS

**Conception and design:** Grigorios Kotronoulas, Nora Kearney, Roma Maguire
**Collection and assembly of data:** Grigorios Kotronoulas, Roma Maguire, Alison Harrow, David Di Domenico, Suzanne Croy, Stephen MacGillivray
**Data analysis and interpretation:** All authors
**Manuscript writing:** All authors
**Final approval of manuscript:** All authors

## REFERENCES


1. Siegel R, DeSantis C, Virgo K, et al: Cancer treatment and survivorship statistics, 2012. CA Cancer J Clin 62:220-241, 2012
2. Cheng KK, Yeung RM: Impact of mood disturbance, sleep disturbance, fatigue and pain among patients receiving cancer therapy. Eur J Cancer Care (Engl) 22:70-78, 2013
3. Cleeland CS: Symptom burden: Multiple symptoms and their impact as patient-reported outcomes. J Natl Cancer Inst Monogr 16-21, 2007
4. Trotti A, Colevas AD, Setser A, et al: Patient-reported outcomes and the evolution of adverse event reporting in oncology. J Clin Oncol 25:5121-5127, 2007
5. Valderas JM, Alonso J: Patient reported outcome measures: A model-based classification system for research and clinical practice. Qual Life Res 17:1125-1135, 2008
6. Espallargues M, Valderas JM, Alonso J: Provision of feedback on perceived health status to health care professionals: A systematic review of its impact. Med Care 38:175-186, 2000
7. Gilbody SM, House AO, Sheldon T: Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome: A systematic review. Psychol Med 32:1345-1356, 2002
8. Greenhalgh J, Meadows K: The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: A literature review. J Eval Clin Pract 5:401-416, 1999
9. Luckett T, Butow PN, King MT: Improving patient outcomes through the routine use of patient-reported data in cancer clinics: Future directions. Psychooncology 18:1129-1138, 2009
10. Marshall S, Haywood K, Fitzpatrick R: Impact of patient-reported outcome measures on routine practice: A structured review. J Eval Clin Pract 12:559-568, 2006
11. Mitchell AJ, Waller A, Carlson LE: Implementing a screening programme for distress in cancer settings: Science and practice. Psicooncologia 9:259-275, 2012
12. Moher D, Liberati A, Tetzlaff J, et al: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med 151:264-269, W64, 2009
13. Higgins JP, Altman DG, Gotzsche PC, et al: The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 343:d5928, 2011
14. Lipsey MW, Wilson D: Practical Meta-Analysis (ed 1). Thousand Oaks, CA, Sage Publications, 2000
15. Wilson BD: Practical Meta-Analysis Effect Size Calculator. http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD-main.php
16. Cohen J: Statistical power analysis for the behavioral sciences (ed 2). Hillsdale, NJ, Lawrence Earlbaum Associates, 1988
17. Berry DL, Blumenstein BA, Halpenny B, et al: Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. J Clin Oncol 29:1029-1035, 2011
18. Boyes A, Newell S, Girgis A, et al: Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl) 15:163-171, 2006
19. Braeken AP, Kempen GI, Eekers D, et al: The usefulness and feasibility of a screening instrument to identify psychosocial problems in patients receiving curative radiotherapy: A process evaluation. BMC Cancer 11:479, 2011
20. Carlson LE, Groff SL, Maciejewski O, et al: Screening for distress in lung and breast cancer outpatients: A randomized controlled trial. J Clin Oncol 28:4884-4891, 2010
21. Detmar SB, Muller MJ, Schornagel JH, et al: Health-related quality-of-life assessments and patient-physician communication: A randomized controlled trial. JAMA 288:3027-3034, 2002
22. Ganz PA, Greendale GA, Petersen L, et al: Managing menopausal symptoms in breast cancer survivors: Results of a randomized controlled trial. J Natl Cancer Inst 92:1054-1064, 2000
23. Girgis A, Breen S, Stacey F, et al: Impact of two supportive care interventions on anxiety, depression, quality of life, and unmet needs in patients with nonlocalized breast and colorectal cancers. J Clin Oncol 27:6180-6190, 2009
24. Hilarius DL, Kloeg PH, Gundy CM, et al: Use of health-related quality-of-life assessments in daily clinical oncology nursing practice: A community hospital-based intervention study. Cancer 113:628-637, 2008
25. Hoekstra J, de Vos R, van Duijn NP, et al: Using the symptom monitor in a randomized controlled trial: The effect on symptom prevalence and severity. J Pain Symptom Manage 31:22-30, 2006
26. Klinkhammer-Schalke M, Koller M, Steinger B, et al: Direct improvement of quality of life using a tailored quality of life diagnosis and therapy pathway: Randomised trial in 200 women with breast cancer. Br J Cancer 106:826-838, 2012

**27.** Maunsell E, Brisson J, Deschênes L, et al: Randomized trial of a psychologic distress screening program after breast cancer: Effects on quality of life. J Clin Oncol 14:2747-2755, 1996

**28.** McLachlan SA, Allenby A, Matthews J, et al: Randomized trial of coordinated psychosocial interventions based on patient self-assessments versus standard care to improve the psychosocial functioning of patients with cancer. J Clin Oncol 19:4117-4125, 2001

**29.** Mills ME, Murray LJ, Johnston BT, et al: Does a patient-held quality-of-life diary benefit patients with inoperable lung cancer? J Clin Oncol 27:70-77, 2009

**30.** Nicklasson M, Elfström ML, Olofson J, et al: The impact of individual quality of life assessment on psychosocial attention in patients with chest malignancies: A randomized study. Support Care Cancer 21:87-95, 2013

**31.** Rosenbloom SK, Victorson DE, Hahn EA, et al: Assessment is not enough: A randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. Psychooncology 16:1069-1079, 2007

**32.** Ruland CM, Holte HH, Røislien J, et al: Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: A randomized clinical trial. J Am Med Inform Assoc 17:403-410, 2010

**33.** Sarna L: Effectiveness of structured nursing assessment of symptom distress in advanced lung cancer. Oncol Nurs Forum 25:1041-1048, 1998

**34.** Taenzer P, Bultz BD, Carlson LE, et al: Impact of computerized quality of life screening on physician behaviour and patient satisfaction in lung cancer outpatients. Psychooncology 9:203-213, 2000

**35.** Takeuchi EE, Keding A, Awad N, et al: Impact of patient-reported outcomes in oncology: A longitudinal analysis of patient-physician communication. J Clin Oncol 29:2910-2917, 2011

**36.** Thewes B, Butow P, Stuart-Harris R: Does routine psychological screening of newly diagnosed rural cancer patients lead to better patient outcomes? Results of a pilot study. Aust J Rural Health 17:298-304, 2009

**37.** Trowbridge R, Dugan W, Jay SJ, et al: Determining the effectiveness of a clinical-practice intervention in improving the control of pain in outpatients with cancer. Acad Med 72:798-800, 1997

**38.** Velikova G, Booth L, Smith AB, et al: Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. J Clin Oncol 22:714-724, 2004

**39.** Velikova G, Keding A, Harley C, et al: Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: Secondary outcomes of a randomised controlled trial. Eur J Cancer 46:2381-2388, 2010

**40.** Cleeland CS, Wang XS, Shi Q, et al: Automated symptom alerts reduce postoperative symptom severity after cancer surgery: A randomized controlled clinical trial. J Clin Oncol 29:994-1000, 2011

**41.** Kearney N, McCann L, Norrie J, et al: Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 17:437-444, 2009

**42.** Kornblith AB, Dowell JM, Herndon JE II, et al: Telephone monitoring of distress in patients aged 65 years or older with advanced stage cancer: A cancer and leukemia group B study. Cancer 107: 2706-2714, 2006

**43.** Moher D, Hopewell S, Schulz KF, et al: CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 63:e1-e37, 2010

**44.** Boyce MB, Browne JP: Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. Qual Life Res 22:2265-2278, 2013

**45.** Devlin N, Appleby J: Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making. London, United Kingdom, The King's Fund, 2010. https://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf



## Appendix

**Table A1.** Electronic Databases Searched and Search Terms Used

| Electronic Databases | Search Terms Used |
|---|---|
| Medline (1946 to May 2012) | 1. exp controlled clinical trial/ |
| EMBASE (1974 to May 2012) | 2. exp randomized controlled trial/ |
| CINAHL (inception to May 2012) | 3. 1 OR 2 |
| PsycINFO (inception to May 2012) | 4. exp neoplasms/OR cancer*.mp. OR neoplasm*.mp. OR carcinoma*.mp. OR oncol*.mp. OR malignan*.mp. OR tumor*.mp. OR tumour*.mp. OR leukemia*.mp. OR leukaemia*.mp. OR sarcoma*.mp. OR lymphoma*.mp. OR melanoma*.mp. OR blastoma*.mp. |
| PBSC (inception to May 2012) | 5. 3 AND 4 |
| | 6. (patient reported outcomes OR patient reported outcome OR patient based outcome OR patient reported outcome measure$).mp. |
| | 7. inventory.ti. OR inventory.ab. |
| | 8. instrument*.ti. OR instrument*.ab. |
| | 9. measure*.ti. |
| | 10. self report*.ti. OR self report*.ab. |
| | 11. 7 OR 8 OR 9 OR 10 |
| | 12. 6 OR 11 |
| | 13. 5 AND 12 |
| | 14. Remove duplicates from 13 |
| | 15. Limit 14 to abstracts |
| | 16. Limit 15 to English language |

NOTE: Search strategy as conducted in Ovid Medline.
Abbreviations: ab, abstract; CINAHL, Cumulative Index to Nursing and Allied Health Literature; exp, term explosion; mp, free text search for a term; PBSC, Psychology and Behavioral Sciences Collection; ti, title.

**Table A2.** Intervention and Outcome Assessment PROMs Used in the 24 Studies Reviewed (n = 26 articles)

| Author and Publication Year | Intervention PROM(s) | Outcome Assessment PROM(s)* | Same Intervention/Outcome PROM(s) |
|---|---|---|---|
| Berry et al,[17] 2011 | SDS | Audio-recorded consultations | No |
| | EORTC QLQ-C30 | Author-developed questionnaire regarding clinic visit duration; clinician evaluation of the intervention | |
| | Pain scale | | |
| | PHQ-9 | | |
| | SSS | | |
| Boyes et al,[18] 2006 | Physical symptoms scales | Physical symptoms scales | Yes, plus additional PROMs |
| | HADS | HADS | |
| | SCNS-SF31 | SCNS-SF31 | |
| | | Patient/clinician acceptability survey | |
| Braeken et al,[19] 2011 | SIPP | Medical records | No |
| | | Intervention evaluation inventories | |
| Carlson et al,[20] 2010 | DT and problem list | DT and problem list | Yes |
| | PSSCAN part C | PSSCAN part C | |
| Cleeland et al,[40] 2011 | MDASI | MDASI | Yes |
| | | Author-developed form for patient evaluation of the intervention | |
| Detmar et al,[21] 2002 | EORTC QLQ-C30 | Audio-recorded consultations | No |
| | | COOP | |
| | | WONCA | |
| | | Medical records | |
| | | Author-developed fatigue scale | |
| | | Patient Satisfaction Questionnaire C | |
| | | Physician satisfaction with communication | |
| | | SF-36 | |
| | | Patient/physician evaluation of the intervention survey | |
| Ganz et al,[22] 2000 | Daily diary symptom cards | Daily diary symptom cards | Yes, plus additional PROMs |
| | CARES (sexual summary scale) | CARES (sexual summary scale) | |
| | | RAND Vitality Scale | |
| Girgis et al,[23] 2009 | HADS | HADS | Yes, plus additional PROMs |
| | EORTC QLQ-C30 | EORTC QLQ-C30 | |
| | SCNS-SF34 | SCNS-SF34 | |
| | NA-ACP | NA-ACP | |
| | | Patient perceptions of improved communication | |
| Hilarius et al,[24] 2008 | EORTC QLC-C30 | Self-report communication questionnaire | No |
| | EORTC LC13 | COOP | |
| | EORTC BR23 | WONCA | |
| | EORTC CR38 | Chart audit | |
| | | PSQ-II | |
| | | SF-36 | |
| | | FACT-L/C/BCS | |
| | | Patient/nurse evaluation of the intervention questionnaire | |
| Hoekstra et al,[25] 2006 | The Symptom Monitor | The Symptom Monitor | Yes |
| Kearney et al,[41] 2009 | Author-developed symptom questionnaire integrating the CTCAE grading system and the CSAS (electronic version) | Author-developed symptom questionnaire integrating the CTCAE grading system and the CSAS (paper-based version) | Yes |
| Klinkhammer-Schalke et al,[26] 2012 | EORTC QLC-C30 | EORTC QLC-C30 | Yes, plus additional PROMs |
| | EORTC BR23 | EORTC BR23 | |
| | | Medical records | |
| Kornblith et al,[42] 2006 | HADS | HADS | Yes, plus additional PROMs |
| | EORTC QLQ-C30 | EORTC QLQ-C30 | |
| | MOS-SSS | MOS-SSS | |
| | | GDS-SF | |
| | | OARSQ, physical health subscale | |
| | | Utilization of mental health and psychosocial services scale | |
| | | GSRE | |
| | | Patient satisfaction with research program | |

(continued on following page)

**Table A2.** Intervention and Outcome Assessment PROMs Used in the 24 Studies Reviewed (n = 26 articles) (continued)

| Author and Publication Year | Intervention PROM(s) | Outcome Assessment PROM(s)* | Same Intervention/Outcome PROM(s) |
|---|---|---|---|
| Maunsell et al,[27] 1996 | GHQ-20 | GHQ-20<br>Social Support Questionnaire<br>LES<br>LWMAT<br>PSI<br>Perceptions of health and worries about health<br>Number of visits to HP<br>Medical records | Yes, plus additional PROMs |
| McLachlan et al,[28] 2001 | CNQ-SF<br>EORTC QLQ-C30<br>BDI-SF | CNQ-SF<br>EORTC QLQ-C30<br>BDI-SF<br>Patient satisfaction survey | Yes, plus additional PROMs |
| Mills et al,[29] 2009 | EORTC QLQ-C30<br>EORTC LC13 | FACT-L TOI subscale<br>PQLI<br>Utilization of diary<br>Patient/clinician communication checklist<br>Patient satisfaction with care survey | No |
| Nicklasson et al,[30] 2013 | EORTC QLQ-C30<br>EORTC LC13 | Audio-recorded consultations<br>Medical records | No |
| Rosenbloom et al,[31] 2007 | FACT-G | FLIC<br>Brief POMS-17<br>PSQ-III<br>Clinical treatment changes survey | No |
| Ruland et al,[32] 2010 | Choice ITPA | Choice ITPA<br>Chart audit | Yes, plus additional PROMs |
| Sarna,[33] 1998 | SDS | SDS | Yes |
| Taenzer et al,[34] 2000 | EORTC QLQ-C30 | PDIS<br>Exit interview<br>Medical record audit | No |
| Takeuchi et al,[35] 2011 | EORTC QLQ-C30<br>HADS | Audio-recorded consultations | No |
| Thewes et al,[36] 2009 | DT<br>SPHERE-Short | Medical records<br>SCNS-SF34<br>Satisfaction with intervention, Likert scales | No |
| Trowbridge et al,[37] 1997 | Pain inventories | Pain inventories<br>PMI<br>Chart audit | Yes, plus additional PROMs |
| Velikova et al,[38] 2004 | EORTC QLQ-C30<br>HADS | Audio-recorded consultations<br>FACT-G<br>Physician use of QoL information checklist | No |
| Velikova et al,[39] 2010 | EORTC QLQ-C30<br>HADS | MCQ<br>Satisfaction with care, single-item scales<br>Intervention evaluation questionnaires | No |

Abbreviations: BDI, Beck Depression Inventory; Brief POMS-17, Brief Profile of Mood States-17; BR-23, Breast Cancer 23 Module; CARES, Cancer Rehabilitation Evaluation System; CNQ-SF, Cancer Needs Questionnaire–Short Form; COOP, Dartmouth Primary Care Cooperative Information Functional Health Assessment; CR-38, Colorectal Cancer 38 Module; CSAS, Chemotherapy Symptom Assessment Scale; CTCAE, Common Toxicity Criteria Adverse Events; DT, Distress Thermometer; EORTC-LC13, European Organisation for Research and Treatment of Cancer–Lung Cancer Module 13; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer–Core Quality of Life Questionnaire, version 3.0; FACT-G, Functional Assessment of Cancer Therapy–General; FACT-L/C/BCS, Functional Assessment of Cancer Therapy–Lung/Colorectal/Breast Cancer Subscale; FLIC, Functional Living Index Cancer; GDS-SF, Geriatric Depression Scale–Short Form; GHQ, General Health Questionnaire; GSRE, Geriatric Schedule of Recent Experience; HADS, Hospital Anxiety and Depression Scale; HP, health professional; ITPA, interactive tailored patient assessments; LC-13, Lung Cancer 13 Module; LES, Life Experiences Survey; LWMAT, Lock-Wallace Marital Adjustment Test; MCQ, Medical Care Questionnaire; MDASI, MD Anderson Symptom Inventory; MOS-SSS, Medical Outcomes Study–Social Support Survey; NA-ACP, Needs Assessment for Advanced Cancer Patients; OARSQ-Physical Health, Older American Resources and Services Questionnaire–Physical Health; PDIS, Patient Satisfaction Questionnaire; PHQ-9, Patient Health Questionnaire–9; PMI, Pain Management Index; PQLI, Palliative Care Quality of Life Index; PROM, patient-reported outcome measure; PSI, Psychiatric Symptom Index; PSQ-III/II, Medical Outcomes Study–Patient Satisfaction Questionnaire III/II; PSSCAN Part C, Psychological Screen for Cancer–Part C; QoL, quality of life; RAND, Research and Development; SCNS-SF31, Supportive Care Needs Survey–Short Form 31; SCNS-SF34, Supportive Care Needs Survey–Short Form 34; SDS, Symptom Distress Scale; SF-36, Medical Outcomes Study 36-Item Short Form Health Survey; SIPP, Screening Inventory of Psychosocial Problems; SPHERE-Short, Somatic and Psychological Health Report–Short form; SSS, Subject Significance Scale; TOI, Trial Outcome Index; WONCA, World Organization Project of National Colleges and Academics.

*If no specific PROM was used, method of assessment is reported instead.

**Table A3.** Evaluation of PROM Intervention Effects on Patient Outcomes

| Outcome | ES (*d*) | 95% CI*† | Effect Characterization‡ |
|---|---|---|---|
| Menopausal symptom distress | −1.18 | −1.68 to −0.67[22] | + |
| Prevalence | | | |
| Anxiety | −0.07 | −0.41 to 0.27[23] | ± |
| Depression | −0.15 | −0.73 to 0.43[23] | ± |
| Overall supportive care needs | −0.20 | −0.46 to 0.06[23] | ± |
| | 0.58[36] | | + |
| Need for help | | | |
| Psychological needs | −0.16 | −0.73 to 0.40[18] | ± |
| | 0.50[36] | | + |
| Information needs | −0.29 | −0.86 to 0.28[18] | ± |
| | 0.53[36] | | + |
| Patient care and support | −0.47 | −1.05 to 0.10[18] | ± |
| Physical and daily living | −0.34 | −0.91 to 0.24[18] | ± |
| | 0.46[36] | | ± |
| Sexual functioning | −0.49 | −0.96 to −0.02[22] | + |
| QoL | | | |
| Role functioning | −0.04 | −0.26 to 0.19[23] | ± |
| | −0.12 | −0.40 to 0.16[21] | ± |
| Emotional/psychological functioning | −0.18 | −0.41 to 0.05[23] | ± |
| | −0.11[31] | | ± |
| | −0.20 | −0.48 to 0.07[21] | ± |
| | 0.10 | −0.25 to 0.44[42] | ± |
| Cognitive functioning | −0.05 | −0.27 to 0.18[23] | ± |
| Social functioning | −0.01 | −0.24 to 0.22[23] | ± |
| | −0.04[31] | | ± |
| | −0.07 | −0.35 to 0.21[21] | ± |
| Physical functioning | −0.16 | −0.39 to 0.01[23] | ± |
| | −0.12[31] | | ± |
| | −0.04 | −0.32 to 0.24[21] | ± |
| | −0.20 | −0.55 to 0.15[42] | ± |
| Physical and functional well-being | −0.41 | −0.95 to 0.14[29] | ± |
| Mental health | −0.10 | −0.38 to 0.18[21] | ± |
| Vitality | 0.08 | −0.38 to 0.54[22] | ± |
| | −0.08 | −0.36 to 0.20[21] | ± |
| Bodily pain | −0.07 | −0.35 to 0.21[21] | ± |
| Nausea | −0.16[31] | | ± |
| Hardship owing to cancer | −0.05[31] | | ± |
| Overall QoL | −0.05 | −0.28 to 0.17[23] | ± |
| | −0.14[31] | | ± |
| | −0.59 | −1.16 to −0.01[28] | + |
| | −0.35 | −0.70 to −0.001[26] | + |
| | −0.04 | −0.38 to 0.31[42] | ± |
| Severity | | | |
| Fatigue | −0.37 | −0.77 to 0.04[25] | ± |
| | −0.25 | −0.63 to 0.12[41] | ± |
| Pain | 0.04 | −0.36 to 0.44[25] | ± |
| Lack of appetite | −0.04 | −0.44 to 0.36[25] | ± |
| Shortness of breath | 0.05 | −0.35 to 0.45[25] | ± |
| Sore mouth/throat | 0.32 | −0.05 to 0.69[41] | ± |
| Coughing | −0.37 | −0.77 to 0.03[25] | ± |
| Sleeplessness | −0.31 | −0.71 to 0.09[25] | ± |
| Hand-foot syndrome | 0.42 | 0.05 to 0.79[41] | - |
| Nausea | −0.44 | −0.84 to 0.04[25] | ± |
| | −0.18 | −0.55 to 0.20[41] | ± |
| Constipation | 0.24 | −0.16 to 0.64[25] | ± |
| Diarrhea | 0.0 | −0.40 to 0.40[25] | ± |
| | 0.06 | −0.32 to 0.43[41] | ± |
| Vomiting | 0.33 | −0.07 to 0.73[25] | ± |
| | 0.01 | −0.36 to 0.38[41] | ± |
| Anxiety | −0.09 | −0.65 to 0.48[18] | ± |
| | −0.05[20] | | ± |
| | −0.30 | −0.65 to 0.04[42] | ± |

(continued on following page)

**Table A3.** Evaluation of PROM Intervention Effects on Patient Outcomes (continued)

| Outcome | ES (*d*) | 95% CI*† | Effect Characterization‡ |
|---|---|---|---|
| Depression | 0.08 | −0.49 to 0.64[18] | ± |
| | −0.01[20] | | ± |
| | −0.15 | −0.49 to 0.20[42] | ± |
| Psychological distress | −0.09 | −0.34 to 0.16[27] | ± |
| | −0.42 | −0.76 to −0.07[42] | + |
| Prevalence | | | |
| Fatigue | −0.07 | −0.62 to 0.47[25] | ± |
| | −0.29 | −0.60 to 0.02[18] | ± |
| | −0.20[41] | | + |
| Pain | −0.33 | −0.78 to 0.12[25] | ± |
| Lack of appetite | −0.29 | −0.74 to 0.15[25] | ± |
| | −0.19 | −0.55 to 0.18[18] | ± |
| Shortness of breath | −0.06 | −0.50 to 0.38[25] | ± |
| Coughing | 0.34 | −0.11 to 0.79[25] | ± |
| Sleeplessness | −0.40 | −0.85 to 0.04[25] | ± |
| Nausea | −0.10 | −0.57 to 0.37[25] | ± |
| | −0.06 | −0.79 to 0.67[18] | ± |
| | −0.10[41] | | ± |
| Constipation | −0.73 | −1.29 to −0.17[25] | + |
| | −0.06 | −1.60 to 1.49[18] | ± |
| Diarrhea | −0.32 | −0.90 to 0.27[25] | ± |
| | −0.88 | −2.10 to 0.37[18] | ± |
| | 0.01[41] | | ± |
| Vomiting | −0.98 | −1.83 to −0.13[25] | + |
| | −0.05[41] | | ± |
| Skin rash | −0.06 | −1.60 to 1.49[18] | ± |
| Sore mouth | 0.25 | −0.58 to 1.08[18] | ± |
| | 0.06[41] | | ± |
| Metallic taste | −0.06 | −1.17 to 1.05[18] | ± |
| Hot flashes | 0.75 | −0.48 to 1.98[18] | ± |
| Hand-foot syndrome | 0.23[41] | | + |
| Overall distress | −0.15[20] | | ± |
| Distress | | | |
| Vomiting | 0.05 | −0.32 to 0.42[41] | ± |
| Nausea | −0.15 | −0.52 to 0.22[41] | ± |
| Diarrhea | 0.0 | −0.37 to 0.37[41] | ± |
| Hand-foot syndrome | 0.35 | −0.02 to 0.72[41] | + |
| Sore mouth/throat | 0.33 | −0.05 to 0.70[41] | ± |
| Fatigue | −0.31 | −0.69 to 0.06[41] | ± |
| Overall | −0.02[31] | | ± |
| | −0.16[20] | | ± |
| Health status | −0.01 | −0.34 to 0.33[27] | ± |
| | 0.0 | −0.34 to 0.34[42] | |
| Worry about health | −0.10 | −0.39 to 0.20[27] | ± |
| Working during assessment | 0.01 | −0.28 to 0.30[27] | ± |
| Hours worked per week | −0.05 | −0.30 to 0.20[27] | ± |
| Household activities performed | −0.08 | −0.33 to 0.17[27] | ± |
| Engagement in social activities | −0.23 | −0.48 to 0.02[27] | ± |
| Engagement in leisure activities | 0.14 | −0.11 to 0.39[27] | ± |
| Engagement in physical activities | −0.02 | −0.27 to 0.23[27] | ± |
| Marital satisfaction | 0.0 | −0.25 to 0.25[27] | ± |

NOTE. Negative ES denote more favorable outcomes (eg, less severity or better scores) for the intervention group, and vice versa. ES were not calculated for controlled trials that reported pre-intervention between-group differences in the outcome in question, or where no relevant data were available. Where data were available, but no such baseline comparisons were performed/stated, baseline scores/percentages were compared using two-tailed independent sample *t* tests, thus ensuring that postintervention scores were not a result of preintervention differences. When studies reported results at more than one time point, the final time point was used, thus ensuring independence of data; hence, each study contributed no more than one ES for a specific outcome.[14] For studies with more than one experimental group, separate ES were calculated if different intervention PROMs were used; however, if the same intervention PROM was used, one ES was calculated based on pooled experimental versus control effects. If a study indicated that the effect was nonsignificant but no statistics were provided, ES was entered as zero.

Abbreviations: ES, effect sizes; PROM, patient-reported outcome measure; QoL, quality of life.

*ES calculations were performed only in those studies for which enough data were available.

†Where no 95% CIs are reported, not enough data were available to calculate them.

‡Based on *P* values ($P < .05$) and direction; + favors the intervention group ($P < .05$); − favors the control group ($P < .05$); ± represents $P \geq .05$.

**Table A4.** Evaluation of PROM Intervention Effects on Processes of Care

| Outcome | ES (*d*) | 95% CI*† | Effect Characterization‡ |
|---|---|---|---|
| Action | | | |
| Enrolled onto medical trial | −0.15 | −0.53 to 0.22[27] | ± |
| Met with other survivors | −0.14 | −0.41 to 0.14[27] | ± |
| Participated in patient-support group | −0.02 | −0.47 to 0.43[27] | ± |
| Consulted treating oncologist | −0.22 | −0.52 to 0.09[24] | ± |
| Consulted family physician | −0.002 | −0.33 to 0.32[27] | ± |
| Consulted other physician | −0.10 | −0.38 to 0.18[27] | ± |
| Had consultation for CAM therapies | −0.18 | −0.54 to 0.19[27] | ± |
| Had psychiatric/psychological consultation | −0.02 | −0.44 to 0.40[27] | ± |
| Sought help because of feeling depressed/sad | −0.11 | −0.41 to 0.19[27] | ± |
| Had a confidant | −0.27 | −0.67 to 0.13[27] | ± |
| Participated in relaxation activities | −0.05 | −0.43 to 0.32[27] | ± |
| Made dietary changes | 0.13 | −0.14 to 0.41[27] | ± |
| Discussed | | | |
| Nausea/vomiting | −0.06 | −0.26 to 0.14[17] | ± |
| | 0.22 | −0.08 to 0.52[24] | ± |
| | 0.02[38] | | ± |
| Appetite | −0.07 | −0.41 to 0.27[21] | ± |
| | −0.06 | −0.24 to 0.13[17] | ± |
| | −0.09 | −0.41 to 0.22[24] | ± |
| | −0.40[38] | | ± |
| Insomnia/sleep problems | −0.34 | −0.65 to −0.03[30] | + |
| | 0.06 | −0.25 to 0.37[21] | ± |
| | −0.05 | −0.23 to 0.13[17] | ± |
| | −0.66 | −1.00 to −0.32[24] | + |
| | −0.64[38] | | + |
| Pain | −0.13 | −0.50 to 0.24[21] | ± |
| | 0.02 | −0.16 to 0.19[17] | ± |
| | −0.10 | −0.39 to 0.20[24] | ± |
| | −0.01[38] | | ± |
| Fatigue | −0.05 | −0.36 to 0.25[30] | ± |
| | −0.30 | −0.62 to 0.03[21] | ± |
| | 0.0 | −0.19 to 0.19[17] | ± |
| | −0.13 | −0.43 to 0.17[24] | ± |
| | −0.34[38] | | ± |
| Bowel pattern | −0.06 | −0.36 to 0.25[30] | ± |
| Constipation | −0.38 | −0.69 to −0.07[21] | + |
| Diarrhea | 0.14 | −0.05 to 0.33[17] | ± |
| Concentration | −0.40 | −0.72 to −0.08[24] | + |
| Appearance | −0.67 | −1.04 to −0.30[24] | + |
| Impact on sex | −0.29 | −0.64 to 0.07[17] | ± |
| Breathing/dyspnea | 0.19 | −0.07 to 0.45[17] | ± |
| | −0.58 | −0.99 to −0.17[17] | + |
| | 0.01 | −0.18 to 0.19[17] | ± |
| | −0.77 | −1.22 to −0.33[24] | + |
| | −0.15[38] | | ± |
| Outlook | −0.18 | −0.48 to 0.13[30] | ± |
| Cough | −0.40 | −0.82 to 0.02[21] | ± |
| Fever/chills | −0.05 | −0.24 to 0.15[17] | ± |
| Depression | −0.05 | −0.24 to 0.14[17] | ± |
| Suicidal ideation | −0.03 | −0.21 to 0.15[17] | ± |
| Symptoms of illness | −0.12 | −0.36 to 0.13[17] | ± |
| | −0.26 | −0.89 to 0.36[17] | ± |
| Physical functioning | −0.07 | −0.38 to 0.23[30] | ± |
| | 0.02 | −0.52 to 0.56[29] | ± |
| | 0.03 | −0.15 to 0.21[17] | ± |
| | 0.26 | −0.10 to 0.63[24] | ± |
| | −0.21[38] | | ± |
| | −0.18 | −0.48 to 0.13[30] | ± |
| | −0.98 | −1.31 to −0.64[21] | + |
| | −0.05 | −0.26 to 0.17[19] | ± |

(continued on following page)

**Table A4.** Evaluation of PROM Intervention Effects on Processes of Care (continued)

| Outcome | ES (*d*) | 95% CI*† | Effect Characterization‡ |
|---|---|---|---|
| Emotional functioning | −0.11 | −0.28 to 0.07[17] | ± |
| | 0.05 | −0.28 to 0.37[24] | ± |
| | −0.24[38] | | ± |
| | −0.44 | −0.75 to −0.13[30] | + |
| | −0.17 | −0.48 to 0.14[21] | ± |
| | 0.26 | −0.29 to 0.81[29] | ± |
| | −0.19 | −0.38 to −0.01[19] | + |
| Social functioning | −0.14 | −0.37 to 0.08[17] | ± |
| | −0.18 | −0.59 to 0.23[24] | ± |
| | 0.19[38] | | ± |
| | −0.21 | −0.51 to 0.10[30] | ± |
| | 0.05 | −0.49 to 0.62[29] | ± |
| | −0.49 | −0.93 to −0.04[21] | + |
| | −0.16 | −0.38 to 0.06[19] | ± |
| Cognitive functioning | −0.08 | −0.35 to 0.18[17] | ± |
| | −0.66 | −1.19 to −0.12[24] | + |
| | −0.33[38] | | ± |
| | 0.0 | −0.31 to 0.31[30] | ± |
| | −0.36 | −0.97 to 0.25[21] | ± |
| Daily functioning | 0.14 | −0.22 to 0.50[24] | ± |
| | 0.38 | −0.19 to 0.94[29] | ± |
| Role functioning | 0.01 | −0.18 to 0.20[17] | ± |
| | −0.15[38] | | ± |
| | 0.33 | −0.23 to 0.90[29] | ± |
| | 0.70 | 0.37 to 1.03[21] | - |
| Sexual problems | 0.06 | −0.16 to 0.28[19] | ± |
| Impact on family relationships | 0.30 | −0.25 to 0.85[29] | ± |
| Existential issues | 0.0 | −0.31 to 0.31[30] | ± |
| Financial issues | 0.10 | −0.20 to 0.41[30] | ± |
| | 0.02 | −0.53 to 0.58[29] | ± |
| Medical/technical issues/effects of treatment | 0.27 | −0.04 to 0.57[30] | ± |
| | 0.23 | −0.33 to 0.78[29] | ± |
| Overall condition | 0.37 | −0.20 to 0.95[29] | ± |
| Global QoL | −0.01 | −0.24 to 0.21[17] | ± |
| | −0.45 | −0.76 to −0.14[30] | + |
| | 0.10 | −0.21 to 0.41[30] | ± |
| No. of concerns/symptoms discussed during consultations | −1.09 | −1.67 to −0.52[34] | + |
| | −0.41[38] | | + |
| | −0.38 | −0.66 to −0.10[21] | + |
| No. of concerns/issues charted on patient records by nurses | −0.54 | −0.81 to −0.27[24] | + |
| | −0.68[32] | | + |
| No. of concerns/issues charted on patient records by physicians | −0.33[32] | | + |
| No. of concerns/issues charted on patient records by health professionals, mixed sample | −0.49 | −1.04 to 0.05[34] | ± |
| Average duration of contact | 0.18 | −0.07 to 0.43[27] | ± |
| | −0.08 | −0.24 to 0.09[17] | ± |
| | 0.12 | −0.13 to 0.37[28] | ± |
| | 0.09[38] | | ± |
| | 0.03 | −0.27 to 0.33[30] | ± |
| | 0.09 | −0.19 to 0.37[21] | ± |
| Satisfaction with nursing care | −0.56 | −1.40 to 0.28[28] | ± |
| Satisfaction with medical care | −0.16 | −1.16 to 0.84[28] | ± |
| Satisfaction with information received | −0.50 | −1.12 to 0.12[28] | ± |
| | 0.18 | −0.36 to 0.72[34] | ± |
| | 0.03 | −0.53 to 0.60[29] | ± |
| Satisfaction with support/rapport/communication | 0.0 | −0.54 to 0.54[34] | ± |
| | −0.07[31] | | ± |
| | −0.04 | −0.61 to 0.53[29] | ± |
| | −0.37 | −0.65 to −0.09[21] | + |
| | 0.13 | −0.05 to 0.31[19] | ± |

(continued on following page)

**Table A4.** Evaluation of PROM Intervention Effects on Processes of Care (continued)

| Outcome | ES (*d*) | 95% CI*† | Effect Characterization‡ |
|---|---|---|---|
| Satisfaction with help received about important problems | 0.69 | 0.20 to 1.17[42] | + |
| Satisfaction with involvement in decision-making | 0.14 | −0.42 to 0.71[29] | ± |
| Satisfaction with HPs addressing patient needs | −0.35 | −0.90 to 0.19[34] | ± |
| Overall satisfaction with care | 0.13 | 0.44 to 0.69[28] | ± |
| | −0.39 | −1.92 to 1.15[28] | ± |
| | −0.08[39] | | ± |
| | 0.33[31] | | ± |
| Overall satisfaction with intervention | 0.13 | −0.44 to 0.69[29] | ± |
| Intervention acceptability, comfortable with using the system | −0.52 | −1.03 to −0.01[42] | + |
| Intervention acceptability, system easy to use | −0.49 | −0.94 to −0.04[40] | + |
| HP satisfaction with clinical encounter | −0.59 | −0.14 to −1.05[40] | + |
| HP action | 0.0[21] | | ± |
| No. of actions taken/medical decisions made per patient | −0.40 | −0.94 to 0.15[34] | ± |
| | 0.16[38] | | ± |
| | 0.02[31] | | ± |
| Referred to psychosocial care or other provider | −0.32 | −0.62 to −0.02[30] | + |
| | 0.08 | −0.27 to 0.42[19] | ± |
| | −0.31 | −0.87 to 0.26[36] | ± |
| | −0.01 | −0.31 to 0.28[24] | ± |
| | 0.11 | −0.22 to 0.43[26] | ± |
| | −0.32[20] | | ± |
| Prescription of medication | 0.04 | −0.18 to 0.26[19] | ± |
| | 0.26 | −0.07 to 0.60[24] | ± |
| Ordering tests | −0.41 | −0.72 to −0.09[37] | + |
| Changing/stopping chemotherapy | −0.11 | −0.45 to 0.22[24] | ± |
| Offering counseling on managing health problems | −0.05 | −0.38 to 0.27[24] | ± |
| HP awareness of patient outcomes | −0.26 | −0.65 to 0.14[21] | ± |
| Physical | −0.13 | −0.40 to 0.14[24] | ± |
| Feelings | −0.21 | −0.69 to 0.27[21] | ± |
| | −0.16 | −0.43 to 0.11[24] | ± |
| Daily activities | −0.13 | −0.66 to 0.39[21] | ± |
| | −0.28 | −0.55 to −0.01[24] | + |
| Social activities | 0.09 | −0.41 to 0.59[21] | ± |
| | −0.09 | −0.35 to 0.18[24] | ± |
| Overall health | −0.50 | −1.05 to 0.05[21] | ± |
| | −0.20 | −0.47 to 0.07[24] | ± |
| Pain | 0.19 | −0.27 to 0.64[21] | ± |
| | −0.54 | −0.82 to −0.27[24] | + |
| Fatigue | 0.20 | −0.34 to 0.74[21] | ± |
| | −0.15 | −0.41 to 0.12[24] | ± |
| QoL | −0.18 | −0.58 to 0.23[21] | ± |
| | −0.27 | −0.54 to 0.0[24] | ± |
| Prevalence of patients undertreated for pain | −0.07 | −0.33 to 0.18[37] | ± |

NOTE. Negative ES denote more favorable outcomes (ie, more frequent discussion or better communication) for the intervention group and vice versa. ES were not calculated for controlled trials that reported preintervention between-group differences in the outcome in question or where no relevant data were available. Where data were available but no such baseline comparisons were performed/stated, baseline scores/percentages were compared using two-tailed independent sample *t* tests, thus ensuring that postintervention scores were not because of preintervention differences. When studies reported results at more than one time point, the final time point was used, thus ensuring independence of data. Hence, each study contributed no more than one ES for a specific outcome.[14] For studies with more than one experimental group, separate ES were calculated if different intervention PROMs were used; however, if the same intervention PROM was used, one ES was calculated based on pooled experimental versus control effects. If a study indicated that the effect was not significant but no statistics were provided, ES was entered as zero.

Abbreviations: CAM, complementary/alternative medicine; ES, effect sizes; HP, health professional; PROM, patient-reported outcome measure; QoL, quality of life.

*ES calculations were performed only in those studies for which enough data were available.

†Where no 95% CIs are reported, not enough data were available to calculate them.

‡Based on *P* value ($P < .05$) and direction; + favors the intervention group ($P < .05$); − favors the control group ($P < .05$); ± represents $P \geq .05$.

**Table A5.** Evaluation of PROM Intervention Effects on Health Service Outcomes

| Outcome | ES (*d*) | 95% CI* | Effect Characterization† |
|---|---|---|---|
| Patient use of psychological referrals | −0.10 | −1.02 to 0.82[22] | ± |
| Self-referrals | −0.20 | −0.44 to 0.04[20] | ± |
| Patient contacts with health professional | −0.85 | −1.10 to −0.59[27] | + |
| | −0.15 | −0.45 to 0.15[30] | ± |
| Patient use of mental health services | 0.18 | −0.45 to 0.82[42] | ± |

NOTE. Negative effect sizes denote more favorable outcomes (eg, more frequent use of service or more contacts) for the intervention group and vice versa. ES were not calculated for controlled trials that reported preintervention between-group differences in the outcome in question or where no relevant data were available. Where data were available but no such baseline comparisons were performed or stated, baseline scores/percentages were compared using two-tailed independent sample *t* tests, thus ensuring that postintervention scores were not because of preintervention differences. When studies reported results at more than one time point, the final time point was used, thus ensuring independence of data. Hence, each study contributed no more than one ES for a specific outcome.[14] For studies with more than one experimental group, separate ES were calculated if different intervention PROMs were used; however, if the same intervention PROM was used, one ES was calculated based on pooled experimental versus control effects. If a study indicated that the effect was nonsignificant but no statistics were provided, ES was entered as zero.

Abbreviations: ES, effect size; PROM, patient-reported outcome measure.

*ES calculations were performed only in those studies for which enough data were available.

†Based on *P* value ($P < .05$) and direction; + ($P < .05$ favors intervention group); − $P < .05$ favors control group); ± ($P \geq .05$).

# THE ADDITION OF MOOD AND ANXIETY DOMAINS TO THE UNIVERSITY OF WASHINGTON QUALITY OF LIFE SCALE

**Simon N. Rogers, FDS, RCS, FRCS, MD,[1] Suzanne Gwanne,[2] Derek Lowe, MSc, C.Stat,[3] Gerry Humphris, PhD, MClin Psychol, CPsychol Reader,[4] Beven Yueh, MD,[5] Ernest A. Weymuller, Jr, MD[6]**

[1] Consultant and Honorary Reader, Regional Maxillofacial Unit, University Hospital Aintree, Fazakerley, Liverpool L9 1AL United Kingdom. E-mail: snrogers@globalnet.co.uk
[2] Medical student, Liverpool University, United Kingdom
[3] Medical Statistician, Mossley Cheshire, United Kingdom
[4] Department of Clinical Psychology, School of Psychiatry and Behavioural Sciences, Rawnsley Building, Manchester University, United Kingdom
[5] Department of Otolaryngology–Head and Neck Surgery, University of Washington School of Medicine, Seattle, Washington, and Health Services Research and Development Service, Veterans Affairs Puget Sound Health Care System, Seattle, Washington
[6] Department of Otolaryngology–Head and Neck Surgery, University of Washington School of Medicine, Seattle, Washington

*Accepted 2 January 2002*
*Published online 8 May 2002 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hed.10106*

**Abstract:** *Background.* There are numerous head and neck specific quality of life questionnaires, each having its own merits and disadvantages. The University of Washington questionnaire has been widely used and is notable by the inclusion of a shoulder dysfunction domain, domain importance ratings, and patient free text. It is short, simple to process, and provides clinically relevant information. However, it has lacked any psychological dimension of quality of life. The aim of this study was to report the inclusion of two psychological domains (mood, anxiety) to the most recent refinement of the questionnaire (version 3).

*Method.* A cross-sectional survey was performed in April 2000. Questionnaires were sent to 183 patients alive and disease free after surgery for oral and oro-pharyngeal malignancy. Replies were received from 145 patients (79% response rate).

*Results.* The new domains (mood and anxiety) correlated significantly with the emotional functioning domains from the EORTC C30 and with the pain and appearance domains of UW-QOL. There were also significant correlations between the "global quality of life" item and the two new domains. Mood ($p = .005$) and anxiety ($p < .001$) scores were associated with patient age but with no other clinicodemographic variable.

*Conclusion.* The addition of mood and anxiety domains makes the UW-QOL version 4 a single broad measure suitable for effective health-related quality of life evaluation in the routine clinical setting. © 2002 Wiley Periodicals, Inc. *Head Neck* **24:** 521–529, 2002

*Correspondence to:* S. N. Rogers

Ethical approval was given for this study by the South Sefton Ethical Committee. Dr. Yueh is supported by a Career Development Award (#RCD-98318) from the Health Services Research & Development Service of the Veterans Health Administration, Department of Human Affairs. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

**Keywords:** health-related quality of life; head and neck cancer; mood; anxiety; UW-QOL

There is a growing awareness of the importance of health-related quality of life (HRQOL) and the merit of its inclusion as an outcome parameter in patients with head and neck cancer. There are many questionnaires available,[1,2] and in the past they have tended to be used for research rather than routine clinical practice. The benefit of a short questionnaire is that it can be self-completed by the patient. Sadura and coworkers[3] suggest that a self-completed questionnaire needs to be understandable and take less than 10 minutes to complete. Simplicity is paramount, because HRQOL is best measured longitudinally. Thus, each patient will complete a questionnaire on several occasions from baseline (pretreatment) to 1 year and annually thereafter. The administration of questionnaires adds an additional burden to clinical resources[4] (ease of processing is an important feature in questionnaire selection). Despite their brevity, questionnaires can give clinically useful information,[5–7] and their routine inclusion as an outcome measure in head and neck units is to be encouraged.

The University of Washington questionnaire (UW-QOL) has an established place in the evaluation of HRQOL in patients with head and neck cancer.[1,2] One of its most appealing features is its simplicity. In the original description, Hassan and Weymuller[8] stated that "the advantages of the head and neck questionnaire are that (1) it is brief and self-administered, (2) it is multifactorial, allowing sufficient detail to identify subtle change, (3) it provides questions specific to head and neck cancer, and (4) it allows no input from the health provider, thus reflecting the QOL as indicated by the patient". The questionnaire has undergone two major revisions since it was first published. In version 2, each of the nine original domains was followed by an importance-rating scale, and three new single-item "quality-of-life" questions were also added.[9] In version 3, two new domains (taste, saliva) were added, and the employment domain was dropped (Table 1).[10,11] And, rather than asking patients to rank order the importance of each individual domain, version 3 just asks patients to indicate which three domains have been most important in the last 7 days. These changes have served to address several shortcomings,[10] but version 3 still did not include an emotional domain. Because HRQOL refers to the physical, emotional, and social impact of diseases and their treatments on patients' lives,[12,13] the emotional domain is an essential component of a broad quality-of-life outcome measure. Mood was chosen as an appropriate domain to capture depressive morbidity. Previous work by Allen and colleagues[14] has shown that "depressed mood in the last month" had the strongest item-total correlation to a brief depression assessment scale developed for elderly people in medical and surgical inpatients. Anxiety was also selected as an essential additional domain describing the emotional component. Anxiety may independently exist or covary with depressed mood.

**Table 1.** Summary of development of the UW-QOL.

| Domain | Version 1 | Version 2 | Version 3 | Version 4 |
|---|---|---|---|---|
| Pain | X | X | X | X |
| Appearance | X | X | X | X |
| Activity | X | X | X | X |
| Recreation | X | X | X | X |
| Swallowing | X | X | X | X |
| Chewing | X | X | X | X |
| Speech | X | X | X | X |
| Shoulder | X | X | X | X |
| Taste | | | X | X |
| Saliva | | | X | X |
| Mood | | | | X |
| Anxiety | | | | X |
| Employment | X | X | | |
| Global QOL items | | X | X | X |
| Free text | | X | X | X |
| Importance rating | | X | X | X |

The purpose of this cross-sectional study was to report version 3 with its two additional domains of taste and saliva and to report the inclusion of two new domains (mood and anxiety) in version 4. This is the first time that version 3 has been reported in a UK population and the first time version 4 has been described in the literature.

There is considerable value in a HRQOL questionnaire that is widely acceptable to head and neck cancer centers and units. It is possible that the addition of two new psychological domains (UW-QOL version 4) will make the questionnaire a realistic outcome measure in routine practice. It is hoped that this article will help to endorse this proposal.

## SUBJECTS AND METHODS

**Patients.** On the departmental oncology database between 1995–1999, there were 290 previously untreated patients with oral and oropharyngeal squamous cell carcinoma. Patients with a previous malignancy were excluded. All patients were treated by primary surgery with or without adjuvant radiotherapy, sixteen had failed to attend outpatients clinic in the preceding year and were therefore not sent a questionnaire. Of the remaining 274 patients, 84 were known to have died. A further seven were excluded from the study, because they were already completing questionnaires in a study using the UW-QOL v2. Thus on March 15, 2000, version 4 was sent to 183 surviving patients. If a reply had not been received within 3 weeks, one further request was made.

**Measures.** In version 4, the two new domains are mood and anxiety. Mood has a 5-point Likert scale; My mood is excellent and unaffected by my cancer, My mood is generally good and only occasionally affected by my cancer; I am neither in a good mood nor depressed about my cancer; I am somewhat depressed about my cancer; I am extremely depressed about my cancer. Anxiety has a 4-point Likert scale: I am not anxious about my cancer; I am a little anxious about my cancer. I am anxious about my cancer; I am very anxious about my cancer. The domains are scored on a scale ranging from 0 (worst) to 100 (best), consistent with the existing algorithms of the UW-QOL. The important-rating schema was modified to include mood and anxiety, thus patients were asked to tick up to 3 of 12 boxes.

A UW-QOL composite score from 0 to 100 was obtained by averaging the scores of the domains. When two or more domains were not answered, no composite score was calculated. Scoring is scaled, so that a score of 0 represents the worst quality of life, and a score of 100 represents the best quality of life.

In our study, the employment domain for the UW-QOL v1 was included, so that a composite UW-QOL score from version 1 could be compared. Patients were also sent the emotional functioning subscale from the EORTC C30[13] (Q 21–24) and dry mouth, sticky saliva, sense of smell items from the EORTC H&N35.[15] The EORTC items were included to allow comparison with the taste, saliva, mood, and anxiety domains in the newest version of the UW-QOL.

**Statistical Methods.** Internal consistency was measured with Cronbach's alpha. If an item fails to correlate well with the other items, we can expect to see the alpha value rise in its absence. Factor analysis was conducted to group individual questions with strong correlation into discrete clusters or constructs. A conventional varimax method of rotation was used.[16] The loading of an item on a factor reflects the correlation of that item to the mathematically derived "latent" factor. Patients were divided into roughly three equal groups on the basis of time from operation, the cut off points being chosen before any QOL analyses took place. The level of statistical significance was taken as $p < .05$, but care should be taken on the interpretation of borderline significance, because many statistical tests were done.

## RESULTS

Questionnaires were sent to 183 patients, and replies were received from 145 of these, giving a 79% response rate. The median time from operation was 830 days (2 years 3 months), with an interquartile range (IQR) of 372 days (1 year) to 1239 days (3 years 5 months). Patient characteristics are described in Table 2 for all patients and within each of three survival time periods. Four of 10 patients were aged 65 or older and 65% were men. Three of 10 had oropharyngeal cancers, whereas 4 of 10 had tumors larger than 4 cm. Half had had radiotherapy. Longer-term survivors were less likely to have radiotherapy. Otherwise, there were no obvious

**Table 2.** Patient characteristics.

| | Time from first operation | | | |
|---|---|---|---|---|
| | <500 days | 500–999 days | 1000+ days | All patients |
| Age 65+ | 40% (21/53) | 39% (22/57) | 39% (26/66) | 40% (71/178) |
| Male gender | 72% (39/54) | 63% (36/57) | 64% (42/66) | 65% (118/182) |
| Posterior site | 37% (19/51) | 21% (13/54) | 29% (19/65) | 30% (53/174) |
| Tumor size 3–4 | 33% (17/51) | 46% (26/56) | 41% (27/66) | 40% (71/177) |
| Flap surgery | 70% (35/50) | 80% (44/55) | 78% (50/64) | 76% (132/173) |
| Radiotherapy | 60% (28/47) | 47% (26/55) | 43% (28/65) | 49% (84/171) |

*Table gives % (number of cases).*

trends of these variables with time from operation. The response rate varied most by type of surgery, with 90% response (37 of 41) from patients having had primary closure, laser or split skin graft (ssg) and 77% response (102 of 132) for those having had flap surgery. There were no other clear differences in response rate by age group, gender, tumor site, tumor size, or radiotherapy.

There was little correlation between length of time from operation and the UW-QOL v4 domain scores (Spearman correlations: median 0.04; IQR, −0.02–0.06; ranges, −0.19–0.17). Subsequent analyses were therefore for the group as a whole. The distributions of domain scores for the UW-QOL v4 are summarized in Table 3. The distributions of scores for taste, saliva, mood, and anxiety show a pattern similar to most of the other domains in that there was no overly strong ceiling or floor effects.

The taste, saliva, mood, and anxiety domains correlated with the relevant items from the EORTC questionnaire (Table 4). The strongest correlations involving taste and saliva were with the UW-QOL domains of swallowing and chewing and with the dry mouth and sense of taste domains from the EORTC. The strongest correlations involving mood and anxiety were with the emotional functioning items and their domain score from the EORTC and with the pain and appearance domains of the UW-QOL. Both HRQOL and overall QOL measures were associated with mood and anxiety. For both these global measures, the strongest correlations were with pain, activity, recreation, and mood (data not shown).

Radiotherapy was strongly associated with taste (MW, $p < .001$), the mean taste score being 55 (SE 4) for patients given treatment and 77 (4) for those not requiring treatment. Best taste

**Table 3.** Distribution of domain scores.

| | UW-QOL scores | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 25 | 30 | 50 | 70 | 75 | 100 | Mean | SE | % Best |
| UW-QOL | | | | | | | | | | |
| Pain | 3 | 4 | | 29 | | 45 | 59 | 77 | 2 | 42 |
| Appearance | 1 | 8 | | 28 | | 71 | 35 | 73 | 2 | 24 |
| Activity | 3 | 4 | | 59 | | 33 | 44 | 69 | 2 | 31 |
| Recreation | – | 13 | | 33 | | 54 | 42 | 72 | 2 | 30 |
| Swallowing | 8 | | 14 | | 66 | | 55 | 74 | 2 | 38 |
| Chewing | 18 | | | 83 | | | 42 | 58 | 3 | 29 |
| Speech | 3 | | 14 | | 71 | | 50 | 75 | 2 | 36 |
| Shoulder | 8 | | 15 | | 28 | | 86 | 80 | 3 | 63 |
| Taste | 11 | | 35 | | 35 | | 59 | 67 | 3 | 42 |
| Saliva | 9 | | 29 | | 37 | | 62 | 71 | 3 | 45 |
| Other | | | | | | | | | | |
| Mood | 3 | 19 | | 19 | | 50 | 49 | 72 | 2 | 35 |
| Anxiety | 9 | | 12 | | 67 | | 51 | 73 | 2 | 37 |

*For each domain the table gives the number of patients with each score, the mean and SE of patient scores, and the percentage of patients selecting the best response possible (100). The shaded area denotes values that do not exist for that domain.*

**Table 4.** Spearman correlation coefficients involving taste, saliva, mood and anxiety.

| | Taste | Saliva | Mood | Anxiety |
|---|---|---|---|---|
| UW-QOL | | | | |
| Pain | **0.29** | 0.25 | **0.45** | **0.41** |
| Appearance | **0.40** | **0.37** | **0.48** | **0.34** |
| Activity | **0.38** | **0.40** | **0.41** | 0.28 |
| Recreation | **0.39** | **0.45** | **0.39** | 0.27 |
| Swallowing | **0.49** | **0.60** | **0.36** | 0.24 |
| Chewing | **0.51** | **0.51** | 0.23 | 0.13 |
| Speech | **0.35** | **0.30** | **0.29** | 0.15 |
| Shoulder | 0.18 | 0.15 | 0.12 | 0.09 |
| Taste | | **0.56** | **0.30** | 0.19 |
| Saliva | **0.56** | | **0.31** | 0.16 |
| Mood | **0.30** | **0.31** | | **0.53** |
| Anxiety | 0.19 | 0.16 | **0.53** | |
| HRQOL, compared with month before cancer | 0.26 | 0.28 | 0.26 | 0.25 |
| In general, HRQOL during the past 7 days | **0.34** | **0.31** | **0.54** | **0.43** |
| Overall QOL during past 7 days | **0.30** | 0.22 | **0.50** | **0.34** |
| EORTC | | | | |
| Have you had a dry mouth? | **0.45** | **0.78** | 0.19 | 0.10 |
| Have you had sticky saliva? | **0.35** | **0.36** | 0.24 | 0.14 |
| Have you had problems with sense of taste? | **0.84** | **0.54** | **0.41** | 0.28 |
| Did you feel tense? | 0.16 | 0.11 | **0.59** | **0.51** |
| Did you worry? | 0.26 | 0.18 | **0.57** | **0.62** |
| Did you feel irritable? | 0.15 | 0.11 | **0.41** | **0.32** |
| Did you feel depressed? | 0.22 | 0.15 | **0.55** | **0.48** |
| Emotional function (EF) | 0.23 | 0.17 | **0.61** | **0.54** |

*Bold highlighting denotes $p < 0.001$.*
*Range of N of patients: UW-QOL v4: 137–143; Mood: 141; Anxiety: 140; EORTC: 123–130.*

scores of 100 were reported for 28% (18 of 64) and 57% (39 of 69), respectively, similarly for saliva ($p < .001$), with mean saliva scores of 58 (4) and 81 (3), and best score rates of 27% (17 of 64) and 61% (40 of 66). Weaker associations of taste and saliva with tumor site were also found ($.01 < p < .05$), with posterior tumour patients tending to have the worse scores. Flap surgery also gave greater problems with taste than other types of surgery ($p = .002$). Mood ($p = .005$) and anxiety ($p = .001$) were associated with age group but not with any other variable. Mean mood scores were 67 (3) for those aged less than 65 and 79 (3) for older patients. The percentage of patients indicating maximum mood scores ("best scores") were 24% (20 of 82) and 50% (28 of 56), respectively. Mean anxiety scores were 65 (3) and 84 (3), whereas best anxiety scores were found for 22% (18 of 82) and 58% (32 of 55), respectively.

The 12 domains were considered together in the analysis of internal consistency for the composite score. All the interdomain correlations were positive and are shown for each domain in Figure 1. The shoulder domain correlated least with other domains. Cronbach's alpha coefficient was 0.86 for a composite score. The loss of any single domain did not change the alpha coefficient to any great extent (range, 0.84–0.86). For the 16 UW-QOL v3 domains, the alpha coefficient was 0.85.

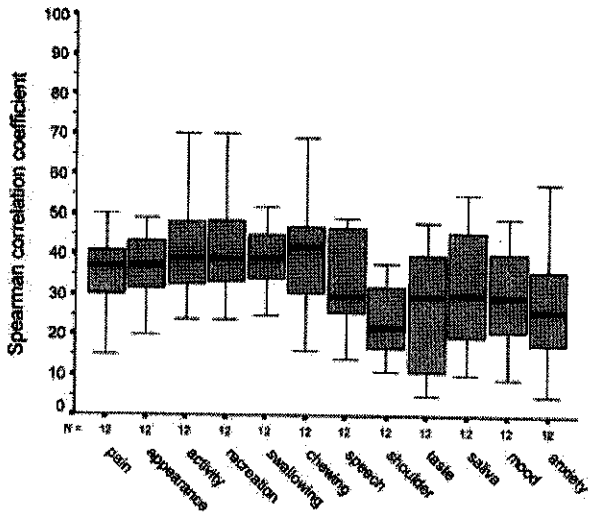



**FIGURE 1.** Interdomain correlations for the UW-QOL-R, mood, anxiety, employment. Each box and whisker is a summary of 12 correlations of that domain with the other 11 domain.

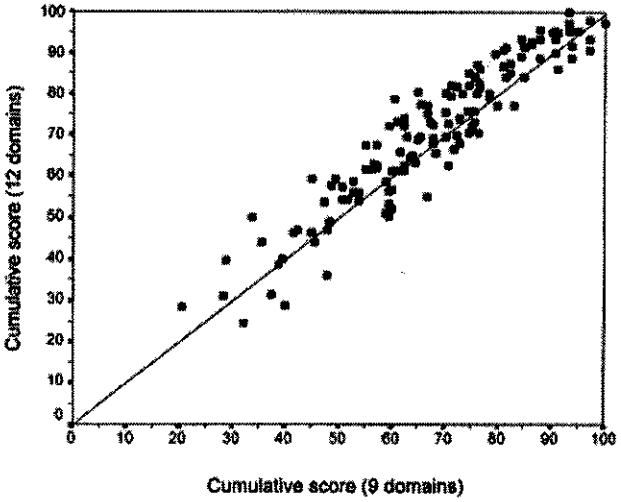



**FIGURE 2.** Scatterplot of version 4 (12 domain) composite score against version 1 and 2 (9 domain) composite score. The change from 9 to 12 domains was due to the dropping of the employment domain and the addition of the taste, saliva, mood, and anxiety domains.

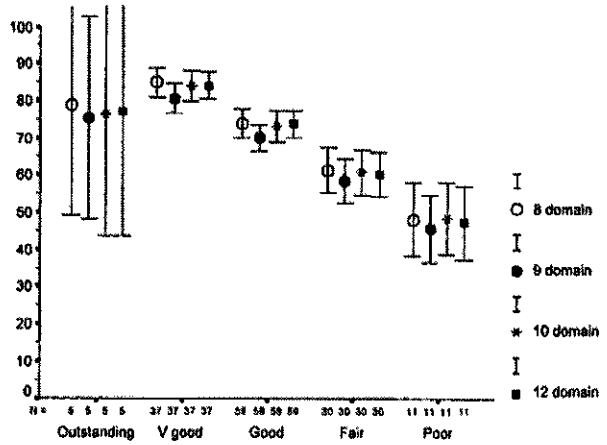



**FIGURE 3.** 95% confidence intervals for mean composite score, for the different UWQOL versions, respectively, by UW-QOL-R global QOL question.

An exploratory factor analysis on the 12 domains produced one dominant and three other main factors. The analysis explained 71% of total variation, with the first factor accounting for 41% and the other factors 11%, 10%, and 8%, respectively. The UW-QOL domains that loaded more strongly (loadings of 0.40 and above) onto the first factor were appearance (0.45), swallowing (0.65), chewing (0.77), and speech (0.56). Pain (0.30) and taste (0.36) had slightly weaker loadings on this first factor. The second factor was compared of pain (0.45), activity (0.76), recreation (0.69), and shoulder (0.45), with pain (0.39) having a slightly weaker loading. Factor three was made up of mood (0.73), anxiety (0.71), and appearance (0.53), whereas factor four was made up of taste (0.45) and saliva (0.96).

We compared the composite version 4 (12 domain) scores to the version 1 and 2 (9-domain) composite scores. The correlation was very high ($r = .94$) with tight scatter (Figure 2). The version 4 scores tended to be higher than the version 1 scores, because the generally lower employment scores were omitted. The correlations between the composite scores making up 8 domains (versions 1 and 2 excluding employment), 9 domains (versions 1 and 2), 10 domains (version 3), and 12 domains (version 4), were all within the range 0.94–0.99.

The coefficients of correlation between the four composite scores and the HRQOL measure were 0.70, 0.72, 0.66 and 0.69. With the overall QOL measure, the correlations were 0.56, 0.56, 0.54, and 0.58. The relationship between overall QOL and the composite scores is shown in Figure 3. Tumor size, radiotherapy treatment, and type of surgery were strongly associated (all $p < .001$) with the composite scores of versions 1, 3, and 4. For example, the mean (SE) of the version 4 composite score was 65 (2) for those with larger tumor sizes (4+ cm) and 76 (2) for those with smaller tumors: with radiotherapy, 64 (2); without radiotherapy; 78 (2); flap surgery, 69 (2); primary closure or laser, 79 (3).

Chewing, speech, swallowing, and saliva were the issues that were the most important to patients in the previous 7 days (Table 5). Least important was recreation. This importance question was unanswered in 6% (8 of 145) of patients, and for one patient the relevant page was absent from the form. Patients within 500 days of operation and patients more than 1000 days from operation agreed on the same four important issues, with chewing ranked first and swallowing as second. For patients within 500 to 999 days of operation, their appearance (3) and activity (2) were also ranked as high as speech (1) chewing (4th equal) and swallowing (4th equal). Patients less than 65 were less likely to rate taste as important and more likely to rate mood (Table 5). Women rated appearance and taste more highly than men, whereas, men rated activity and speech more highly. Chewing was rated higher by patients with anterior tumors, whereas swallowing was rated higher by those with the larger tumors and by those having had radiotherapy.

**Table 5.** Most important issues to patient in previous 7 days. Patients were asked to choose up to three domains.*

| | N of patients choosing domain | Rank order | Most obvious associations with clinicodemographic variables (age-group, sex, tumor site, tumor size, type of surgery, and radiotherapy) |
|---|---|---|---|
| Pain | 23 | 7= | None |
| Appearance | 29 | 5 | Male: 14% (13/95), female: 31% $p = 0.03$ |
| Activity | 27 | 6 | Male: 24% (23/95), female: 6% (3/49), $p = 0.006$ |
| Recreation | 15 | 12 | None |
| Swallowing | 39 | 3 | T1–2: 21% (18/86), T3–4: 38% (21/55), $p = 0.03$<br>Radiotherapy: 40% (27/67), no radiotherapy: 16% (11/69), $p = 0.002$ |
| Chewing | 46 | 1 | Posterior: 20% (9/45), anterior: 39% (37/94), $p = 0.03$ |
| Speech | 40 | 2 | Male: 34% (32/95), female: 16% (8/49), $p = 0.03$ |
| Shoulder | 22 | 9= | None |
| Taste | 23 | 7= | <65yrs: 11% (9/84), 65 + yrs: 25% (14/57), $p = 0.04$<br>Male: 11% (10/95), female: 27% (13/49), $p = 0.02$ |
| Saliva | 37 | 4 | None |
| Mood | 22 | 9= | <65y: 20% (17/84), 65 + y: 7% (4/57), $p = 0.03$ |
| Anxiety | 20 | 11 | None |

**Three patients ticked 4 choices, one patient ticked five. These are included.*

The free text question was responded to by 45% (65 of 145) of patients, often at some length, giving unique insights into aspects of their lives not tapped by the specifics of a tick-box questionnaire.

## DISCUSSION

HRQOL refers to the physical, emotional, and social impact of diseases and their treatments on patients.[12,13] If the measure is for patient self-completion, it needs to be understandable and ideally should take less than 10 minutes to complete.[3] The UW-QOL questionnaire is a broad questionnaire that has the potential as a routine outcome measure in head and neck cancer centers and units. Criticisms of the earlier version of the UW-QOL were addressed recently by Weymuller and coworkers.[10] The UW-QOL instrument was revised and version 3 published. The authors conclude that the questionnaire meets the following desirable characteristics: short and rapid to complete, reproducible, reliable and valid in a population of head and neck cancer patients, does not require excessive training to administer, easy to interpret, and yields discriminative results (separation by site and stage).

The main deficiency of version 3 is that it lacks an emotional component of HRQOL. Although the psychological impact of disease and its treatment is reflected in the single QOL items of the UW-QOL, it is an imprecise marker of psychological outcome.[10,17] It was for this reason that after the first international head and neck quality of life workshop held in Liverpool in November 1999, consideration was given to the addition of two further domains; mood and anxiety. Care was taken in the wording of the two domains. Single questions of psychological dysfunction are well recognized especially in primary care settings.[18,19] Some items have been designed especially for terminally ill[26] and for elderly patients.[21] The authors wanted to avoid expanding each domain into multiple questions. Therefore, the design of the two additional questions was based previous attempts to assess emotional components with single items and remaining consistent with the terminology and scoring of the rest of the questionnaire.

In this study, version 4 has been compared with the EORTC C30 emotional functional subscale and three items (dry mouth, sticky saliva, sense of smell) from the EORTC H&N35. Because a postal survey was performed, the number of questionnaire items were kept to a minimum to promote an adequate response rate; therefore, full versions of the EORTC C30 and H&N 35 were not used. A more detailed comparison has been reported previously,[17] and it is expected that the correlations reported using version 1 will still stand in the newest version of the UW-QOL. It was not the intention of this study to critically appraise the UW-QOL against other commonly advocated head and neck cancer questionnaires.

The exploratory factor analysis suggests a couple of things: first that the 12 UW-QOL

domains might load onto four distinct subscales (factors) and, second, the clearly dominant first factor, comprising maybe half the UW-QOL domains, justifies the use of the composite score. It may be that both approaches can be adopted, namely subscale and composite score reporting of results. The relatively high Cronbach alpha value for the whole 12-item measure indicates that the scale could be reported as a total score. We very much emphasize the exploratory nature of our analyses to derive hypotheses that can be tested with larger numbers of patients. It will also be important to see whether these factors hold together according to the posttreatment interval.

The UW-QOL is backward compatible. There seems little lost when basing a composite score on the original 8 domains common to all versions or basing it on 9 (version 1) or 10 (version 3). Therefore, historical pooled data based on the original eight domains is still valuable, despite recent modifications to the questionnaire. However, we do expect the 12-domain composite score to add greater sensitivity to the discrimination between clinically distinct groups of patients. The composite score has been shown to be a useful indicator of HRQOL,[6] and expected associations with tumor size, radiotherapy, and type of surgery were confirmed in this study.

Importance weighting adds a very useful dimension to the UW-QOL questionnaire.[22,23] In this study, on average, patients after primary surgery for oral and oropharyngeal cancer seem to rate chewing, speech, and swallowing as most important. Although it may be difficult to include importance rating into an overall HRQOL score,[9] it seems reasonable to ask the patient which domains are most important, because this can act as a focus for treatment intervention. This technique was used by Deleyiannis and coworkers in the analysis of postlaryngectomy QOL.[22]

Several factors require attention when interpreting the findings of this study. The cross-sectional design has allowed for rapid assessment of the characteristics of version 4 but has not allowed us to explore the responsiveness of this version over time. Although a longitudinal study is underway, this study has shown interesting results. There is a clear link between subjective saliva and chewing dysfunction and adjuvant radiotherapy. In addition, this study emphasizes the association between swallowing, chewing, and dry mouth.

The emotional domain is of crucial importance in the evaluation of HRQOL outcome.[24] Hammerlid et al[25] reported that about one third of their 357 patients were possible or probable cases of a major mood disorder at each of six occasions sampled over a 1-year period. New cases of anxiety and depression were identified at each occasion. Longitudinal work suggests that a high level of depressive symptoms at baseline, that is before treatment, is a good predictor of symptom severity and functioning after treatment.[26] The new psychological domains of the UW-QOL correlated very well with the EORTC. To more comprehensively evaluate the two new domains (mood and anxiety), version 4 should be tested against specific psychological questionnaires of distress such as the Hospital Anxiety and Depression Scale (HAD), Centre for Epidemiologic Studies Depression CES-D, or Beck Depression Inventory (BDI). These later scales are well known. Other attempts to assess mood in medically ill patients have been reported that have been designed especially for the older patient (>55 years). These measures may serve as good tests of concurrent validity for the emotional domain included in the version 3 questionnaire.[27]

Free text is a valuable component of the UW-QOL. The hand-written comments gives information on a tremendous range of issues not often included as part of head and neck cancer questionnaires such as clinic waiting times, transport, other medical, or family events. Weymuller and coworkers[9] reported their experience from 549 prospectively evaluated patients. Their findings and those of ours both support the use of open-ended text.

## CONCLUSION

This study suggests that the UW-QOL is a suitable questionnaire for HRQOL evaluation after head and neck cancer treatment. In its current form, it provides a broad and rapidly applied measure that provides clinically relevant information in everyday practice. We commend its use in a minimum dataset, because it allows for a simple way of measuring HRQOL in routine head and neck cancer practice and is not limited to research applications.

## REFERENCES

1. Rogers SN, Fisher SE, Woolgar JA. A review of quality of life assessment in oral cancer. Int J Oral Maxillofac Surg 1999;28:99–117.

2. Ringash J, Bezjak A. A structured review of quality of life instruments for head and neck cancer patients. Head Neck 2001;23:201–213.
3. Sadura A, Pater J, Osoba D, Levine M, Palmer M, Bennett K. Quality of life assessment: patient compliance with questionnaire completion. J Natl Cancer Inst 1992;84: 1023–1026.
4. Weymuller EA, Yueh B, Deleyiannis FWB, Kuntz AL, Alsarraf R, Coltrera MD. Quality of life in head and neck cancer. Laryngoscope 2000;110 (Suppl 94):4–7.
5. Rogers SN, Lowe D, Brown JS, Vaughan ED. The University of Washington Head and Neck cancer measure as a predictor of outcome following primary surgery for oral cancer. Head Neck 1999;21:394–401.
6. Rogers SN, Lowe D, Humphris G. Distinct patient groups in oral cancer? A prospective study of perceived health status following primary surgery. Oral Oncology 2000; 36:529–538.
7. Rogers SN, Lowe D, Brown JS, Vaughan ED. The relationship between length of stay and health-related quality of life in patients treated by primary surgery for oral and oro-pharyngeal cancer. Int J Oral Maxillofac Surg 2001;30(3):209–215.
8. Hassan SJ, Weymuller EA. Assessment of quality of life in head and neck cancer patients. Head Neck 1993;15: 485–496.
9. Deleyiannis FW, Weymuller EA, Coltrera MD. Quality of life of disease-free survivors of advanced (stage III or IV) oropharyngeal cancer. Head Neck 1997;19: 466–473.
10. Weymuller EA, Yueh B, Deleyiannis FWB, Kuntz AL, Alsarraf R, Coltrera MD. Quality of life in patients with head and neck cancer: Lessons learned from 549 propectivity evaluated patients. Arch Otolaryngol Head Neck Surg 2000;126:329–335.
11. Weymuller EA, Alsarraf R, Yueh B, Deleyiannis FWB, Coltrera MD. Analysis of the performance characteristics of the University of Washington Quality of Life instrument and its modification (UW-QOL-R). Arch Otolaryngol Head Neck Surg 2001;127(S)489–493.
12. Weinberger M, Oddone EZ, Samsa GP, Landsman PB. Are health-related quality of life measures affected by the mode of administration? J Clin Epidemiol 1996;49:135–140.
13. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Nat Cancer Inst 1993;85:365–376.
14. Allen N, Ames D, Ashby D, Bennetts K, Tuckwell V, West C. A brief sensitive screening instrument for depression in late life. Age Ageing 1994;23(3):213–219.
15. Bjordal K, Ahlner-Elmqvist M, Tollesson E, et al. Development of a European Organisation for Research and Treatment of Cancer (EORTC) questionnaire module to be used in quality of life assessment in head and neck cancer. Acta Oncol 1994;33(8):879–885.
16. Cattell RB. The scientific use of factor analysis in behavioural and life sciences. New York: Plenum; 1978.
17. Rogers SN, Lowe D, Vaughan ED, Brown JS. A comparison of the University of Washington Head and Neck Measure with the SF-36, EORTC C33 and EORTC H&N. Oral Oncology 1998;34:361–372.
18. Williams JW Jr, Mulrow CD, Kroenke K, et al. Case-finding for depression primary care: a randomized trial. American J Medicine 1999;106(1):36–43.
19. Berwick DM, Murphy JM, Goldman PA, Ware JE Jr, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. Medicare 1991;29(2): 169–176.
20. Chochinov HM, Wilson KG, Enns M, Lander S. "Are you depressed?" Screening for depression in the terminally ill. A J Psychiatry 1997;154(5):674–676.
21. D'Arth P, Katona P, Mullan E, Evans S, Katona C. Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. Fam Prac 1994;11(3):260–266.
22. Deleyiannis FW, Weymuller EA, Coltrera MD, Futran N. Quality of life after laryngectomy: are functional disabilities important? Head Neck 1999;21:319–324.
23. Rogers SN, Laher SH, Overend L, Lowe D. Domain importance-rating using the UW-QOL questionnaire in patients treated by primary surgery for oral and oropharyngeal cancer. European Journal of Cranio-Maxillofacial Surgery (in press).
24. de Graeff A, de Leeuw JR, Ros WJ, Hordijk GJ, Blijham GH, Winnubst JA. Pretreatment factors predicting quality of life after treatment for head and neck cancer. Head Neck 2000;22(4):398–407.
25. Hammerlid E, Ahlner-Elmqvist M, Bjordal K, et al. A prospective multicentre study in Sweden and Norway of mental distress and psychiatric morbidity in head and neck cancer patients. Br J Cancer 1999;80:766–774.
26. Hammerlid E, Silander E, Hornestam L, Sullivan M. Health-related quality of life three years after diagnosis of head and neck cancer- a longitudinal study. Head Neck 2001;23:113–125.
27. Hammond MF. Rating depression severity in elderly physically ill patient: reliability and factor structure of the Hamilton and the Montgomery-Asberg depression rating scales Int J Geriatr Psychiatry 1998;13(4):257–261.

# The patients' account of outcome following primary surgery for oral and oropharyngeal cancer using a 'quality of life' questionnaire

S.N. ROGERS, PROFESSOR, FDSRCS, FRCS, MD CONSULTANT, *Regional Maxillofacial Unit, and Faculty of Health, Edge Hill University, University Hospital Aintree, Liverpool,* J. SCOTT, MEDICAL STUDENT, *Faculty of Medicine, Liverpool University, Liverpool,* A. CHAKRABATI, MSC, PGDE, RGN, RSCN, *Faculty of Health, Edge Hill University,* & D. LOWE, MSC, C.STAT. MEDICAL STATISTICIAN, *Regional Maxillofacial Unit, University Hospital Aintree, Liverpool, UK*

ROGERS S.N., SCOTT J., CHAKRABATI A. & LOWE D. (2008) *European Journal of Cancer Care* 17, 182–188

**The patients' account of outcome following primary surgery for oral and oropharyngeal cancer using a 'quality of life' questionnaire**

The purpose of this study was to collate all the University of Washington quality of life (UW-QOL) data collected in the Unit since 1995 and to summarize it in a patient-friendly format that can be used when discussing treatment outcomes. The sample of cancer patients consisted of 561 consecutive patients undergoing primary surgery for previously untreated oral and oropharyngeal squamous cell carcinoma presenting to the Regional Maxillofacial Unit Liverpool, between the years 1995 and 2004. Follow-up was on May 2006. Information sheets were constructed based on the UW-QOL, summarizing the patients at around 2 years (median 28 months). The response to each statement in all 12 domains, domains as important, and the general health-related QOL/QOL questions are given. Data has been summarized for six common groups based on the clinical presentation and treatment. The data show big differences by tumour size, free-flap surgery and adjuvant radiotherapy. Information presented in this format is potentially extremely useful when counselling patients and their families regarding the likely outcomes of treatment.

*Keywords:* questionnaires, information, health-related quality of life, UW-QOL, oral cancer, head and neck cancer.

## INTRODUCTION

Health-related quality of life (HRQOL) data is a valuable clinical outcome measure that can be used in conjunction with traditional parameters such as survival and recurrence rates. In head and neck cancer, validated questionnaires have emerged, which allow clinicians a better insight to the factors that predict HRQOL outcomes (Bjordal *et al.* 1999; Rogers *et al.* 2002; Terrell *et al.* 2004). Health-related quality of life data can influence treatment strategy. Examples include 'organ preservation' (chemoradiotherapy) rather than surgery for large oropharyngeal tumours (Cohen *et al.* 2006), and the choice of treatment for early laryngeal cancer (DiNardo *et al.* 1999; Stoeckli *et al.* 2003). Another potential use of HRQOL data is to feedback the patients perspective in a structured way and to help provide an opportunity for intervention.

Health-related quality of life data is often presented in the context of research, is difficult for health professionals to understand and is beyond most patients and carers. If presented in a meaningful way, HRQOL can give patients, carers and the clinical team a much better understanding of outcome following treatment. A summary of outcomes

Correspondence address: Simon N. Rogers, Consultant, Regional Maxillofacial Unit, University Hospital Aintree, Aintree Trust, L9 7LN Liverpool, UK (e-mail: snrogers@doctors.org.uk).

Ethical approval was given by Sefton Research Ethics Committee.

Accepted 9 June 2007
DOI: 10.1111/j.1365-2354.2007.00832.x

*European Journal of Cancer Care,* 2008, 17, 182–188

as reported by previous patients can improve the communication process between health professionals and new patients when discussing likely problems after treatment. Good communication between doctors and patients about their illness and treatment options is essential particularly in the area of cancer (Gamble 1998; Fallowfield & Jenkins 1999; Semple & McGowan 2002; Zielger *et al.* 2004). The research literature would support the premise that patients with head and neck cancer need information about treatments, illness and prognosis (Mesters *et al.* 2001; Newell *et al.* 2004; Llewellyn *et al.* 2006). Cancer patients want to be fully informed and share decision-making responsibility, but report not receiving sufficient information in all areas (Cox *et al.* 2006).

In head and neck cancer, there are several well-validated HRQOL questionnaires (Bjordal *et al.* 1999; Rogers *et al.* 1999a; Ringash & Bezjak 2001). One of the most commonly used in clinical practice in the UK is the University of Washington quality of life (UW-QOL) questionnaire (Kanatas & Rogers 2004). The questionnaire is relatively simple to complete and to collate. The UW-QOLv4 (Rogers *et al.* 2002) has four components: 12 domains (pain, appearance, activity, recreation, swallowing, chewing, speech, shoulder, taste, saliva, mood and anxiety), an importance-rating scale, more general 'QOL' questions and a 'free-text' section where patients can add their own comment. The domains are written in descriptive sentences such as for speech: *'I have difficulty saying some words but I can be understood over the phone'* or swallowing: *'I can only swallow liquid food'*. This allows discussion about the impact of treatment using examples that are easy for the patient to relate to. When data are presented as simple percentages using every day terminology, this gives clinical relevance and avoids patients needing to comprehend clinical anchors or meaningful changes in mean scores (Funk *et al.* 2004).

University of Washington quality of life data have already been published (Rogers *et al.* 1999b, 2002) and demonstrated a drop in scores at 3 and 6 months following surgery. Scores at 12 months give a good indication of longer-term outcome (Rogers *et al.* 1999c). For patients having simple surgery (laser or primary closure) without radiotherapy, their scores at 1 year are similar to those before operation. For patients with more advanced disease, requiring free tissue transfer and adjuvant radiotherapy, their scores at 1 year are lower than before operation, We have found that patients want an indication of what they will be like in the longer term and the key information seems to be scores at about their second year (findings of an earlier unpublished focus group at the Unit). To include scores at several time points tends to overwhelm the patient and family with too much information. Hence, the aim of this paper was to give UW-QOL data in a style that is easy for patient and health care professionals to understand within key (six) clinical groups at the 2-year point following operation.

## MATERIALS AND METHODS

This study comprised 561 consecutive patients undergoing surgery for previously untreated oral and oropharyngeal squamous cell carcinoma presenting to the Regional Maxillofacial Unit Liverpool from 1 January 1995 to 31 December 2004.

From 1995 to 1999, patients were asked to complete the UW-QOL questionnaire, at presentation and at about 6 and 12 months after surgery. From 2000, pre-treatment questionnaires were opportunistic and more systematic QOL data came from annual postal surveys of post-treatment survivors. The questionnaire represents the patient's own perception of their level of QOL. Many patients have completed several questionnaires well beyond 12 months from treatment, and for this study, the nearest available questionnaire to 24 months was taken to represent the view of the patient in the 'longer term'. Questionnaires earlier than 18 months were not considered.

Although the UW-QOL questionnaire has undergone several revisions, there are eight items common to all versions (pain, appearance, activity, recreation, swallowing, chewing, speech, shoulder function). Version 4, in use since 2000, also has taste, saliva, mood and anxiety domains, a question that asks patients to choose up to three domains of most importance to them and two global questions about their health-related and overall QOL. Overall, QOL includes not only physical and mental health, but also many other factors, such as family, friends, spirituality or personal leisure activities that are important to the enjoyment of life.

Data are presented for the whole group and for key patient subgroups according to size of tumour, site of tumour, surgery and adjuvant radiotherapy. The study does not include questionnaire data after recurrence. The period of follow-up was on May 2006, with mortality information being acquired through the Regional Unit's links to the Office of National Statistics.

Ethical approval was given by Sefton Research Ethics Committee.

## STATISTICAL METHODS

Presentation is primarily descriptive without reference to the scores and summary statistics usually associated with

research presentations of these data. The tables are meant as tables of reference to be used by health professionals with patients. The wording of each question is stated intentionally, in full. Tests of significance within the results (Kruskal–Wallis, chi-squared) provide evidence to support any associations we make between the 'types' of patient and their QOL. Kaplan–Meier survival methods were used to estimate survival.

## RESULTS

The cohort comprised 561 patients. Mean (standard deviation) age was 63 (12) years. Sixty-one per cent were men and 39% were women (Table 1). T3/T4 advanced tumours were present for 37%. Seventy per cent had free-flap surgery and 40% had adjuvant radiotherapy. The six key clinical groups, describing 96% (540/561) of the cohort, are summarized in Table 2.

Estimated all-causes mortality was 8% at 6 months, 16% at 12 months, 23% at 18 months and 27% at 24 months. A questionnaire after 18 months that was closest in time to 24 months after operation was available for 328 patients across the six clinical groups. For patients alive at 18 months, this represents an estimated 78% information yield. Estimates of yield ranged from 72% to 84% across the six clinical groups. The median (IQR) time from operation of the 328 questionnaires was 28 months (23–35 months), range 25–30 months for the medians across the six groups.

To investigate any overall trends with time the 328 patients were placed into three equally sized (tertile) groups and the median times from operation were 22, 28 and 39 months. We found no significant differences nor any discernible trends between these three groups for the UW-QOL domains nor for the importance of domains, but there was a trend for the later of the three groups to report slightly worse health-related ($P = 0.01$) and overall QOL ($P = 0.002$) (results not shown).

The tables of reference for use with patients are Tables 3–5. Each table gives the number of patients who have answered the questions and the percentage breakdown of their answers. As these are intended tables of reference, we have independently checked (triple checked) the transfer of percentages from statistical output into Tables 3–5. The tests of significance highlight differences between the six clinical groups, particularly in regards to oral function, less so in regard to mood, anxiety and single global questions about health-related and overall QOL.

Overall, when asked what their HRQOL had been like during the previous 7 days, 262 patients replied as: 5% outstanding, 26% very good, 39% good, 21% fair, 5% poor and 3% very poor. Overall, QOL not only includes physical and mental health but also many other important factors, and the 262 patients rated their overall QOL in the previous 7 days as: 4% outstanding, 32% very good, 35% good, 20% fair, 7% poor and 2% very poor.

Our previous work (Rogers *et al.* 1999b, 2002) has repeatedly demonstrated the influence of radiotherapy, surgery and tumour size on UW-QOL over time. The UW-QOL information for the longer term (Table 3)

**Table 1.** Characteristics of the 1995–2004 cohort of 561 patients

| | | % | *N* |
|---|---|---|---|
| Sex | Male | 61 | 340 |
| Age | <55 | 26 | 144 |
| | 55–64 | 32 | 179 |
| | 65–74 | 24 | 137 |
| | 75+ | 18 | 101 |
| Tumour site | Oral cavity | 88 | 493 |
| | Oropharynx | 12 | 68 |
| Clinical staging | T1 | 26 | 153 |
| | T2 | 36 | 200 |
| | T3 | 9 | 48 |
| | T4 | 29 | 160 |
| Surgery | Flap | 70 | 391 |
| | Laser/primary closure/ssg | 30 | 170 |
| Adjuvant radiotherapy (RT) | RT | 40 | 223 |

ssg, split skin graft.

**Table 2.** The six key clinical groups

| Group | T stage | Tumour site | Surgery | Radiotherapy (RT) | Patients in cohort | Questionnaires |
|---|---|---|---|---|---|---|
| A | T1, T2 | Oral | No Flap | No RT | 136 | 92 |
| B | T1, T2 | Oral | Flap | No RT | 93 | 63 |
| C | T1, T2 | Oral | Flap | RT | 71 | 41 |
| D | T3, T4 | Oral | Flap | No RT | 73 | 43 |
| E | T3, T4 | Oral | Flap | RT | 99 | 48 |
| F | Any | Oropharynx† | Any | Any | 68 | 41 |
| Others* | | | | | 21 | 9 |
| TOTAL | | | | | 561 | 337 |

*Others comprised 10 with T3-T4, oral, no flap, ± RT, 10 with T1-T2, oral, no flap, RT and one pedicle.
†Oropharynx group comprised 38% T3-T4, 82% with free-flap surgery, 57% with RT.

**Table 3.** Longer-term quality of life for patients according to tumour size, tumour site, type of surgery and adjuvant radiotherapy (RT)

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Pain | $N = 92$ | $N = 62$ | $N = 40$ | $N = 43$ | $N = 47$ | $N = 41$ |
| I have no pain. | 64 | 50 | 52 | 37 | 34 | 39 |
| There is mild pain not needing medication. | 20 | 24 | 22 | 35 | 30 | 24 |
| I have moderate pain – requires regular medication (e.g. paracetamol). | 13 | 23 | 20 | 23 | 26 | 24 |
| I have severe pain controlled only by prescription medicine (e.g. morphine). | 3 | 3 | 3 | 2 | 9 | 12 |
| I have severe pain, not controlled by medication. | 0 | 0 | 3 | 2 | 2 | 0 |
| Appearance | $N = 92$ | $N = 63$ | $N = 40$ | $N = 43$ | $N = 47$ | $N = 41$ |
| There is no change in my appearance. | 60 | 29 | 13 | 14 | 9 | 20 |
| The change in my appearance is minor. | 34 | 51 | 65 | 54 | 45 | 46 |
| My appearance bothers me but I remain active. | 7 | 16 | 18 | 23 | 32 | 24 |
| I feel significantly disfigured and limit my activities due to my appearance. | 0 | 5 | 5 | 9 | 15 | 7 |
| I cannot be with people due to my appearance. | 0 | 0 | 0 | 0 | 0 | 2 |
| Activity | $N = 91$ | $N = 63$ | $N = 40$ | $N = 43$ | $N = 47$ | $N = 41$ |
| I am as active as I have ever been. | 52 | 38 | 25 | 21 | 17 | 17 |
| There are times when I can't keep up my old pace, but not often. | 22 | 24 | 25 | 33 | 26 | 29 |
| I am often tired and have slowed down my activities although I still get out. | 24 | 35 | 48 | 40 | 47 | 51 |
| I don't go out because I don't have the strength. | 0 | 0 | 3 | 5 | 6 | 0 |
| I am usually in bed or chair and don't leave home. | 2 | 3 | 0 | 2 | 4 | 2 |
| Recreation | $N = 92$ | $N = 63$ | $N = 40$ | $N = 43$ | $N = 48$ | $N = 41$ |
| There are no limitations to recreation at home or away from home. | 54 | 35 | 23 | 21 | 13 | 17 |
| There are a few things I can't do but I still get out and enjoy life. | 27 | 37 | 50 | 44 | 54 | 54 |
| There are many times when I wish I could get out more, but I'm not up to it. | 15 | 24 | 23 | 19 | 19 | 20 |
| There are severe limitations to what I can do, mostly I stay at home. | 3 | 5 | 5 | 14 | 10 | 7 |
| I can't do anything enjoyable. | 0 | 0 | 0 | 2 | 4 | 2 |
| Swallowing | $N = 91$ | $N = 63$ | $N = 41$ | $N = 43$ | $N = 48$ | $N = 41$ |
| I can swallow as well as ever. | 80 | 49 | 22 | 33 | 23 | 27 |
| I cannot swallow certain solid foods. | 14 | 43 | 68 | 56 | 44 | 32 |
| I can only swallow liquid food. | 4 | 3 | 10 | 9 | 17 | 32 |
| I cannot swallow because it 'goes down the wrong way' and chokes me. | 1 | 5 | 0 | 2 | 17 | 10 |
| Chewing | $N = 92$ | $N = 63$ | $N = 41$ | $N = 43$ | $N = 48$ | $N = 40$ |
| I can chew as well as ever. | 66 | 37 | 17 | 26 | 8 | 18 |
| I can eat soft solids but cannot chew some foods. | 30 | 59 | 71 | 63 | 50 | 55 |
| I cannot even chew soft solids. | 3 | 5 | 12 | 12 | 42 | 28 |
| Speech | $N = 92$ | $N = 63$ | $N = 39$ | $N = 43$ | $N = 46$ | $N = 41$ |
| My speech is the same as always. | 67 | 41 | 21 | 33 | 20 | 24 |
| I have difficulty saying some words but I can be understood over the phone. | 28 | 52 | 62 | 60 | 57 | 66 |
| Only my family and friends can understand me. | 4 | 5 | 18 | 7 | 22 | 7 |
| I cannot be understood. | 0 | 2 | 0 | 0 | 2 | 2 |
| Shoulder | $N = 88$ | $N = 60$ | $N = 40$ | $N = 42$ | $N = 47$ | $N = 40$ |
| I have no problem with my shoulder. | 75 | 67 | 48 | 67 | 64 | 58 |
| My shoulder is stiff but it has not affected my activity or strength. | 16 | 13 | 28 | 17 | 17 | 23 |
| Pain or weakness in my shoulder has caused me to change my work/hobbies | 7 | 10 | 13 | 7 | 6 | 13 |
| I cannot work or do my hobbies due to problems with my shoulder. | 2 | 10 | 13 | 10 | 13 | 8 |
| Taste | $N = 84$ | $N = 53$ | $N = 32$ | $N = 36$ | $N = 40$ | $N = 33$ |
| I can taste food normally. | 63 | 55 | 34 | 56 | 15 | 30 |
| I can taste most foods normally. | 24 | 28 | 34 | 17 | 30 | 30 |
| I can taste some foods. | 11 | 15 | 25 | 25 | 35 | 27 |
| I cannot taste any foods. | 2 | 2 | 6 | 3 | 20 | 12 |
| Saliva | $N = 83$ | $N = 49$ | $N = 31$ | $N = 34$ | $N = 39$ | $N = 32$ |
| My saliva is of normal consistency. | 67 | 57 | 13 | 53 | 21 | 34 |
| I have less saliva than normal, but it is enough. | 27 | 33 | 45 | 26 | 41 | 19 |
| I have too little saliva. | 4 | 10 | 32 | 18 | 21 | 34 |
| I have no saliva. | 2 | 0 | 10 | 3 | 18 | 13 |
| Mood | $N = 83$ | $N = 51$ | $N = 31$ | $N = 35$ | $N = 38$ | $N = 34$ |
| My mood is excellent and unaffected by my cancer. | 54 | 37 | 45 | 34 | 32 | 32 |
| My mood is generally good and only occasionally affected by my cancer. | 27 | 35 | 26 | 31 | 47 | 47 |
| I am neither in a good mood nor depressed about my cancer. | 8 | 14 | 3 | 23 | 3 | 9 |
| I am somewhat depressed about my cancer. | 8 | 14 | 23 | 9 | 13 | 9 |
| I am extremely depressed about my cancer. | 2 | 0 | 3 | 3 | 5 | 3 |
| Anxiety | $N = 83$ | $N = 52$ | $N = 31$ | $N = 34$ | $N = 34$ | $N = 34$ |
| I am not anxious about my cancer. | 41 | 42 | 35 | 35 | 32 | 44 |
| I am a little anxious about my cancer. | 43 | 46 | 42 | 53 | 47 | 29 |
| I am anxious about my cancer. | 13 | 10 | 16 | 9 | 9 | 18 |
| I am very anxious about my cancer. | 2 | 2 | 7 | 3 | 12 | 9 |

Key to patient groups:
A. T1 T2 oral cancer no free flap no RT.
B. T1 T2 oral cancer free flap no RT.
C. T1 T2 oral cancer free flap + RT.
D. T3 T4 oral cancer free flap no RT.
E. T3 T4 oral cancer free flap + RT.
F. Oropharyngeal cancer.
The table gives the number of patients who have answered each question and the percentage breakdown of their answers.
Kruskal–Wallis test between the six groups: pain ($P = 0.004$), appearance ($P < 0.001$), activity ($P < 0.001$), recreation ($P < 0.001$), swallowing ($P < 0.001$), chewing ($P < 0.001$), speech ($P < 0.001$), shoulder ($P = 0.06$), taste ($P < 0.001$), saliva ($P < 0.001$), mood ($P = 0.24$) and anxiety ($P = 0.80$).

**Table 4.** Issues (domains) patients regard as important

| Importance | A $N=82$ | B $N=52$ | C $N=32$ | D $N=34$ | E $N=39$ | F $N=34$ |
|---|---|---|---|---|---|---|
| Pain | 10 | 19 | 13 | 17 | 18 | 18 |
| Appearance | 21 | 21 | 3 | 15 | 31 | 24 |
| Activity | 18 | 25 | 19 | 15 | 10 | 24 |
| Recreation | 13 | 6 | 6 | 15 | 3 | 15 |
| Swallowing | 13 | 25 | 44 | 24 | 41 | 44 |
| Chewing | 15 | 27 | 47 | 32 | 21 | 18 |
| Speech | 12 | 17 | 22 | 27 | 41 | 38 |
| Shoulder | 15 | 12 | 13 | 12 | 8 | 18 |
| Taste | 15 | 12 | 22 | 12 | 18 | 6 |
| Saliva | 16 | 17 | 47 | 32 | 39 | 50 |
| Mood | 18 | 21 | 16 | 21 | 10 | 21 |
| Anxiety | 21 | 25 | 16 | 18 | 10 | 15 |

Patients were asked 'Which issues have been the most important to you during the last 7 days?'. (Tick up to 3 boxes).
The table gives the number of patients and the percentage breakdown of their answers.
Chi-squared test between the six groups: pain ($P = 0.12$), appearance ($P < 0.001$), activity ($P = 0.002$), recreation ($P < 0.001$), swallowing ($P < 0.001$), chewing ($P < 0.001$), speech ($P < 0.001$), shoulder ($P = 0.42$), taste ($P < 0.001$), saliva ($P < 0.001$), mood ($P = 0.09$) and anxiety ($P = 0.64$).

**Table 5.** University of Washington health-related and overall quality of life

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| HRQOL* | $N=83$ | $N=49$ | $N=29$ | $N=31$ | $N=36$ | $N=33$ |
| Outstanding | 4 | 4 | 7 | 3 | 6 | 12 |
| Very good | 41 | 27 | 28 | 26 | 6 | 9 |
| Good | 35 | 35 | 38 | 35 | 53 | 48 |
| Fair | 16 | 27 | 21 | 26 | 22 | 18 |
| Poor | 5 | 8 | 3 | 6 | 6 | 3 |
| Very poor | 0 | 0 | 3 | 3 | 8 | 9 |
| Overall QOL† | $N=83$ | $N=50$ | $N=29$ | $N=31$ | $N=36$ | $N=33$ |
| Outstanding | 2 | 6 | 3 | 3 | 6 | 6 |
| Very good | 45 | 32 | 41 | 23 | 14 | 24 |
| Good | 31 | 28 | 31 | 35 | 44 | 45 |
| Fair | 17 | 26 | 17 | 26 | 22 | 12 |
| Poor | 5 | 8 | 7 | 6 | 11 | 3 |
| Very poor | 0 | 0 | 0 | 6 | 3 | 9 |

Patients were asked to rate their health-related quality of life (HRQOL) during the past 7 days.
*Kruskal–Wallis test between the six groups: $P = 0.04$.
†Patients were asked to consider everything in their lives that contributed to their personal well-being, and then to rate their overall quality of life (QOL) during the past 7 days. Kruskal–Wallis test between the six groups: $P = 0.09$.

indicates that the patients who do best are those with the smallest oral cancer tumours, not needing free-flap surgery and not having radiotherapy (group A). Those doing least well in respect of functional aspects of swallowing, chewing, speech, taste and saliva, were those who had had adjuvant radiotherapy – i.e groups C and E and most of the oropharyngeal group F.

## DISCUSSION

Head and neck cancer has a disproportionately large impact on HRQOL both as a result of the disease and its treatment (de Graeff *et al.* 2000). One of the reasons for this is that key functions for social interaction are affected such as appearance, chewing, saliva, speech and swallowing.

It is important that the impact of treatment and likely outcome are effectively discussed with the patient and family. Health-related quality of life data is a useful resource, and this paper has presented the data in such a way that patients and lay people can easily understand. This might encourage members of the head and neck multidisciplinary team to use the information in clinical practice. Incidentally, as part of an earlier focus group, we have found that patients are very keen to have their questionnaire data used for the benefit of new patients and expect this data to be used to improve awareness. Also, it is useful that the wider clinical team have an appreciation

of outcome from the patients' perspective. The tables give a clear indication of 'what will I be like?'. Whether this tabular format for presentation of information will be accepted as definitive remains to be seen; for example, patients might prefer statistics as bullet points within an information leaflet. It is hoped that simply presented HRQOL data might help meet some of the previously unmet needs of patients for some patients about to undergo treatment. By summarizing the key domains, it allows patients and carers the opportunity to reflect on the potential impact of treatment and seek clarification about domains of particular interest.

From previous focus groups undertaken at the Unit (unpublished data), patients seem to wish for information on 'what will I be like' in the longer term. They seemed resolved to overcome and cope with the acute side-effects and early problems encountered following treatment and looked towards future longer lasting outcomes. We have already shown that HRQOL outcome at 1–2 years reflects the long term (Rogers *et al.* 1999c), and data in this paper confirms this particularly for the 12 domains of the UW-QOL.

In this paper, six groups of patients with oral and oropharyngeal squamous cell carcinoma have been described. This was the maximum we felt the dataset could support. Any further refinement requires more data, and this is unlikely from a single unit within the near future. There is scope for multicentre collaboration. This breakdown does however, allow patients to be grouped according to the three main factors predicting HRQOL over time – tumour size, surgery and radiotherapy. Another way of presenting HRQOL data to gain more subtlety is to include more factors that predict HRQOL and to use regression methods; however, we are reduced again to either predicting a 'score' within a 0–100 range, or predicting a percentage. The latter has more intuitive appeal as it can relate directly to the language of the questionnaire though it is difficult to know which is the most appropriate percentage to take. Is it the percentage with the best response, or the percentage with the worse outcome? Focus group work will explore this further.

Health-related quality of life is reported as an end point in clinical treatment trials. To facilitate statistical comparison and to minimize publication space, QOL responses are often translated into unitless scores of 0 through 100 that span the full theoretical range of the measure. Health-related quality of life measures are validated on their ability to discern differences in clinically distinct groups using these scores. Mean domain scores are used as a way of distinguishing between distinct clinical groups of patients over time following treatment and for flagging up particular problems in certain groups at certain times. The language used in the presentation of such group statistics makes it difficult for health professionals to translate the results back and to use them with individual patients. The focus on using the data with individual patients is very much orientated to presenting the data verbatim as per questionnaire wording and comparing the individual response to an average for that clinical presentation. The patients' response can be used to trigger further assessment or intervention.

## CONCLUSIONS

In oncology practice, there is a need for HRQOL information to be presented in a clinically useful way that patients, carers and the clinical team readily understand. The simple presentation of HRQOL should help in the discussion of possible outcomes following treatment and is part of the continued progress towards incorporating HRQOL data into routine clinical practice.

## REFERENCES

Bjordal K., Hammerlid E., Ahlner-Elmqvist M., de Graeff A., Boysen M., Evensen J.F., Biorklund A., de Leeuw J.R., Fayers P.M., Jannert M., Westin T. & Kaasa S. (1999) Quality of life in head and neck cancer patients: validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-H & N35. *Journal of Clinical Oncology* **17**, 1008–1019.

Cohen E.E., Haraf D.J., List M.A., Kocherginsky M., Mittal B.B., Rosen F., Brockstein B., Williams R., Witt M.E., Stenson K.M., Kies M.S. & Vokes E.E. (2006) High survival and organ function rates after primary chemoradiotherapy for intermediate-stage squamous cell carcinoma of the head and neck treated in a multicenter phase II trial. *Journal of Clinical Oncology* **24**, 3438–3444.

Cox A., Jenkins V., Catt S., Langridge C. & Fallowfield L. (2006) Information needs and experiences: an audit of UK cancer patients. *European Journal of Oncology Nursing* **10**, 263–272.

DiNardo L.J., Kaylie D.M. & Isaacson J. (1999) Current treatment practices for early laryngeal carcinoma. *Otolaryngology-Head and Neck Surgery* **120**, 30–37.

Fallowfield L.J. & Jenkins V.A. (1999) Effective communication skills are the key to good cancer care. *European Journal of Cancer* **35**, 1592–1597.

Funk G.F., Karnell L.H., Smith R.B. & Christensen A.J. (2004) Clinical significance of health status assessment measures in head and neck cancer: what do quality-of-life scores mean? *Archives of Otolaryngology Head and Neck Surgery* **130**, 825–829.

Gamble K. (1998) Communication and information: the experience of radiotherapy patients. *European Journal of Cancer Care* **7**, 153–161.

de Graeff A., de Leeuw J.R., Ros W.J., Hordijk G.J., Blijham G.H. & Winnubst J.A. (2000) Pretreatment factors predicting quality of life after treatment for head and neck cancer. *Head and Neck* **22**, 398–407.

Kanatas A.N. & Rogers S.N. (2004) A national survey of health-related quality of life questionnaires in head and neck oncology. *Annals of the Royal College of Surgeons England* **86**, 6–10.

Llewellyn C.D., McGurk M. & Weinman J. (2006) How satisfied are head and neck cancer (HNC) patients with the information they receive pre-treatment? Results from the satisfaction with cancer information profile (SCIP). *Oral Oncology* **42**, 726–734.

Mesters I., van den Borne B., de Boer M. & Pruyn J. (2001) Measuring information needs among cancer patients. *Patient Education and Counselling* **43**, 253–262.

Newell R., Ziegler L., Stafford N. & Lewin R.J. (2004) The information needs of head and neck cancer patients prior to surgery. *Annals of the Royal College of Surgeons England* **86**, 407–410.

Ringash J. & Bezjak A. (2001) A structured review of quality of life instruments for head and neck cancer patients. *Head and Neck* **23**, 201–213.

Rogers S.N., Fisher S.E. & Woolgar J.A. (1999a) A review of quality of life assessment in oral cancer. *International Journal of Oral and Maxillofacial Surgery* **28**, 99–117.

Rogers S.N., Lowe D., Brown J.S. & Vaughan E.D. (1999b) The University of Washington head and neck cancer measure as a predictor of outcome following primary surgery for oral cancer. *Head and Neck* **21**, 394–401.

Rogers S.N., Hanna L., Lowe D. & Magennis P. (1999c) Long term QOL in patients 5–10 years after surgery for oral and oropharyngeal cancer. *Journal of Cranio-Maxillofacial Surgery* **3**, 187–191.

Rogers S.N., Gwane S., Lowe D., Humphris G., Yueh B. & Weymuller E.A. (2002) The addition of mood and anxiety domains to the University of Washington Quality of Life Scale. *Head Neck* **24**, 521–529.

Semple C.J. & McGowan B. (2002) Need for appropriate written information for patients, with particular reference to head and neck cancer. *Journal of Clinical Nursing* **11**, 585–593.

Stoeckli S.J., Schnieper I., Huguenin P. & Schmid S. (2003) Early glottic carcinoma: treatment according patient's preference? *Head Neck* **25**, 1051–1056.

Terrell J.E., Ronis D.L., Fowler K.E., Bradford C.R., Chepeha D.B., Prince M.E., Teknos T.N., Wolf G.T. & Duffy S.A. (2004) Clinical predictors of quality of life in patients with head and neck cancer. *Archives of Otolaryngology Head Neck Surgery* **130**, 401–408.

Zielger L., Newell R., Stafford N. & Lewin R. (2004) A literature review of head and neck cancer patients information needs, experiences and views regarding decision making. *European Journal of Cancer Care* **13**, 119–126.

ORIGINAL ARTICLE

# Screening for Dysfunction to Promote Multidisciplinary Intervention by Using the University of Washington Quality of Life Questionnaire

*Simon N. Rogers, FDS, RCS, FRCS, MD; Derek Lowe, MSC, CStat*

**Objective:** To investigate the potential of the University of Washington Quality of Life Questionnaire (UW-QOL) in routine clinics as a quick screening tool for possible dysfunction in patients after treatment of head and neck cancer.

**Design:** Retrospective analysis.

**Setting:** Regional Maxillofacial Unit, Aintree University Hospitals National Health Service Foundation Trust Liverpool, a National Health Service teaching hospital.

**Patients:** Consecutive disease-free patients with oral or oropharyngeal squamous cell carcinoma, who had undergone primary surgery with or without adjuvant radiotherapy, for whom UW-QOL version 4 data from 2000 to May 2006 were available in our research database; and consecutive patients from previous studies (4 postal surveys of disease-free patients with oral or oropharyngeal squamous cell carcinoma, 1 clinic-based study that targeted speech and swallowing in patients with oropharyngeal disease, 1 that evaluated shoulder function in patients with various diagnoses, and 1 that recruited patients without cancer attending a general dental practice).

**Main Outcome Measures:** Cutoff strategies for further evaluation/intervention derived from studies using the UW-QOL in parallel with 13 other established questionnaires. Effects of preferred cutoffs on trigger variation were assessed with the use of all available UW-QOL version 4 data (615 patients).

**Results:** Trigger rates for further intervention fell between 9% (recreation and speech) and 16% (swallowing). Eighty-one percent of patients with free-flap surgery and adjuvant therapy for T3 or T4 tumors met the trigger criteria at around 2 years, with 42% meeting the trigger on 3 or more domains.

**Conclusion:** The fourth version of the UW-QOL is suitable for routine screening in clinical practice.

*Arch Otolaryngol Head Neck Surg. 2009;135(4):369-375*

HEALTH-RELATED QUALITY of life evaluation after head and neck cancer treatment is recognized as an important patient-reported outcome measure. An ever-increasing number of articles have been published on this issue.[1] Validated questionnaires have emerged that allow clinicians a better insight into the factors that predict health-related quality-of-life outcomes.[2,3] They are useful as primary or secondary outcomes in clinical trials. When presented in a simple way, questionnaires can also provide additional information for patient, caregiver, and tumor board members.[4] They help provide feedback from the patient perspective in a structured way and encourage the opportunity for further evaluation and intervention for patients who are doing poorly.

A range of issues are affected by head and neck cancer, and it can be difficult to identify patients with problems in the clinical setting. This is made more problematic by the busy nature of outpatient clinics, where there are often considerable time pressures. Another hurdle is that patients tend to have low self-esteem and are reluctant to complain.[5] Many patients do not wish to be a burden and to take up time in the clinic on such matters. Hence, some patient problems will inevitably go unrecognized.[6] Questionnaires have a role in addressing this issue, and the use of touch-screen technology makes it achievable in routine practice.[7]

There is no one criterion standard questionnaire for quality of life.[8] One that is often used in clinical practice in the United Kingdom is the University of Washington Quality of Life Questionnaire (UW-QOL),[9]

Author Affiliations: Regional Maxillofacial Unit, University Hospital Aintree, Liverpool, and Evidence-Based Practice Research Centre, Faculty of Health, Edge Hill University, Ormskirk, England.

partly because of its simplicity. The fourth version of the UW-QOL[10] has 4 components: 12 domains (pain, appearance, activity, recreation, swallowing, chewing, speech, shoulder, taste, saliva, mood, and anxiety), an importance-rating scale, general quality-of-life questions, and a free-text section where patients can add their own comments. It would be just as appropriate to use other validated head and neck questionnaires to help screen patients in the clinic, but we have a track record for using the UW-QOL questionnaire over several years and it forms the basis of the touch-screen patient-derived outcome tool used in the Liverpool clinic.

The aims of this study were to collate data collected from several previous studies that used a variety of questionnaires in combination with the UW-QOL between 2000 and 2006; to postulate appropriate cutoffs in severity for UW-QOL domains as triggers for further evaluation and intervention; and to estimate how many patients the cutoffs would identify at various times in the cancer journey and in various clinical subgroups.

## METHODS

The 12 domains of the UW-QOL version 4 are scaled from 0 (worst) to 100 (best) according to the hierarchy of response. Patients are also asked which 3 domains have been most important to them in the past 7 days. We have used the UW-QOL questionnaire since 1995 (version 4 since 2000) and have maintained a research database of all UW-QOL questionnaires completed by the 1992 to 2005 cohort of patients with oral or oropharyngeal squamous cell carcinoma whose primary treatment was surgery with or without adjuvant radiotherapy.

All version 4 questionnaires from January 1, 2000, to May 31, 2006, were analyzed to identify patients with significant deficits in their quality of life who were, as such, potential candidates for intervention. For each UW-QOL domain, the domain scores were cross-tabulated with domain importance to derive risk groups. Our intention was to identify a group in which the trigger would select around 5% to 10% of the patients with the worst responses and a second group in which the trigger would select the next 5% to 10%. Other subgroups were identified to take the total at-risk selection to 25% to 35%.

We then analyzed data from a series of previous studies.[10-18] We calculated summary statistics for other measures collected concurrently with the UW-QOL for each of the risk groups identified for the UW-QOL domains. The other measures were the Center for Epidemiology Studies Depression Scale, Derriford Appearance Scale, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Head and Neck, EQ-5D EuroQol, Hospital Anxiety and Depression Scale, Liverpool Oral Rehabilitation Questionnaire version 3, M. D. Anderson Dysphagia Inventory, Neck Dissection Impairment Index, Shoulder Disability Questionnaire, Swallowing (Dysphagia) Specific Quality of Life, Voice Handicap Index, Voice-Related Quality of Life Measure, and Xerostomia-Related Quality of Life Questionnaire. These results formed the evidence base from which we then derived clinically relevant UW-QOL cutoff criteria in each domain as our triggers for intervention. We also took the strength of the wording of UW-QOL domain levels into consideration. Our aim was to derive domain criteria that would trigger intervention in no more than 1 in 5 patients. We recognize the subjectivity in our preferred criteria, and our results are presented as far as possible to allow others to vary their criteria for triggering more or fewer patients.

Using our preferred criteria, we then revisited the total UW-QOL database to explore the variation in trigger rates between main clinical groups at around 2 years after primary surgery.

All of the studies we included in this report recruited consecutive patients. Four were postal surveys of disease-free patients with oral or oropharyngeal squamous cell carcinoma treated by primary surgery with or without adjuvant radiotherapy. Another study was clinic based and investigated speech and swallowing issues, specifically studying patients with oropharyngeal tumors. Another clinic-based study measured shoulder function and included a wider range of diagnoses. Another study recruited subjects without cancer attending a general dental practice.

Ethical approval for each study was given by the Sefton Research Ethics Committee.

## RESULTS

Details of patient populations, response rate, relevant measures, and aims for each study are summarized in **Table 1**. The research database identified 758 patients in a consecutive 1992 through 2005 cohort having had primary surgery with or without adjuvant radiotherapy after a diagnosis of squamous cell oral or oropharyngeal cancer. Of these patients, 143 had died before January 2000, the time from which version 4 of the UW-QOL was used. From January 2000 to May 2006, 79% of patients (487 of 615) completed at least 1 UW-QOL questionnaire (median, 3; interquartile range, 1-5; range, 1-9). There were 1615 questionnaires in all, obtained at a median time from surgery of 36 months (interquartile range, 15-65 months; 10th-90th percentile, 5-101 months).

The database of 1615 UW-QOL questionnaires was used to ascertain the severity of response. For each UW-QOL domain, the domain scores were cross-tabulated with domain importance to derive the most at-risk patient groups. Application of the criteria described in the "Methods" section to each domain produced up to 4 risk subgroups (A through D) to reflect gradations in risk, with subgroup A having the worst UW-QOL responses (**Table 2**). For example, a UW-QOL pain score of either 0 or 25 defines the worst at-risk group for pain (subgroup A); a UW-QOL pain score of 50 together with the patient choosing pain as being an important issue defines subgroup B; a UW-QOL pain score of 50 together with the patient not choosing pain as an important issue defines subgroup C; and a UW-QOL pain score of 75 together with pain being important defines subgroup D. The other possible scores, 75, not important and 100, indicate less risk and are not shown in the table.

These UW-QOL groupings were then used to compare patient scores on other relevant measures collected concurrently (**Table 3** and **Table 4**) since 2000. Table 3 and Table 4 indicate the ability of the UW-QOL to distinguish between patients in regard to these other measures and were used to decide our most preferred trigger for each domain. We aimed to derive trigger criteria that would select not more than 1 in 5 patients. For example, in regard to swallowing, from variation in the M. D. Anderson Dysphagia Inventory and Swallowing (Dysphagia) Specific Quality of Life results between UW-QOL subgroups shown in Table 4, we decided to include UW-QOL swallowing subgroups A and B as trigger

**Table 1. Studies That Collected Other Data Together With UW-QOLv4[a]**

| Source | Study Population | Concurrent Measures of Relevance | Relevant UW-QOLv4 Domains |
|---|---|---|---|
| Rogers et al[10] | 1995-1999 Cohort of 290 previously untreated with H&N SCC, treated by primary surgery ± adjuvant radiotherapy; 183 surviving disease-free patients were sent questionnaire March 15, 2000, reminder 3 wk later; 79% (145) response | EORTC H&N 35 (dry mouth), EORTC H&N 35 (taste), EORTC C30 emotional function | Saliva, taste, mood, anxiety |
| Rogers et al[11] | 1992-2002 Cohort of 577 previously untreated with H&N SCC, treated by primary surgery ± adjuvant radiotherapy; 306 surviving disease-free patients were sent questionnaire April 2003; 64% (197) response | HADS, CES-D | Mood, anxiety |
| Rogers et al[12] | 1992-2003 Cohort of previously untreated with H&N SCC, treated by primary surgery ± adjuvant radiotherapy; 348 surviving disease-free patients were sent questionnaire February 2004; 64% (224) response | EQ-5D | Pain, activity, recreation, mood, anxiety, overall HRQOL |
| Katre et al[13] | 1992-2005 Cohort of previously untreated with H&N SCC, treated by primary surgery ± adjuvant radiotherapy; 383 surviving disease-free patients were sent questionnaire March 2006; 67% (258) response | DAS-24, XeQoLS | Appearance, saliva |
| Thomas et al[14,15] | 1999 to May 2005 cohort of 117 disease-free survivors with histologic diagnosis of SCC of oropharynx (subsites: tongue base, tonsil, lateral pharyngeal wall, soft palate); patients were invited to attend research clinic at their convenience May 1 to October 31, 2005; 66% (77) participated | V-RQOL, VHI, MDADI, SWAL-QOL | Speech, swallowing |
| Rogers et al[16] | Conducted in outpatient setting September 3, 2003, to July 13, 2004; all consecutive patients attending for reviews were included, only newly diagnosed patients excluded; all 100 patients approached agreed to participate | SDQ, NDII | Shoulder |
| Rogers et al[17] | Patients attending 6 Liverpool general dental practices; "normative" sample of patients without cancer aged 40-79 y, using quota sampling in 4 age-sex bands; 372 patients (349 routine and 23 emergency) | LORQv3 | Chewing, saliva |
| Scott et al[18] | 100 Consecutive patients attending Maxillofacial Oncology Clinic at University Hospital Aintree May-October 2006; no refusals; mouth opening was measured by senior physiotherapist | Mouth opening (mm) | Chewing |

Abbreviations: CES-D, Center for Epidemiology Studies Depression Scale; DAS-24, Derriford Appearance Scale; EORTC C30, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30; EORTC H&N 35, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Head and Neck 35; EQ-5D, EQ-5D EuroQol; H&N, head and neck; HADS, Hospital Anxiety and Depression Scale; HRQOL, health-related quality of life; LORQv3, Liverpool Oral Rehabilitation Questionnaire version 3; MDADI, M. D. Anderson Dysphagia Inventory; NDII, Neck Dissection Impairment Index; SCC, squamous cell carcinoma; SDQ, Shoulder Disability Questionnaire; SWAL-QOL, Swallowing (Dysphagia) Specific Quality of Life; UW-QOLv4, University of Washington Quality of Life Questionnaire version 4; VHI, Voice Handicap Index; V-RQOL, Voice-Related Quality of Life Measure; XeQoLS, Xerostomia-Related Quality of Life Questionnaire.

[a] The data sets of these studies were reanalyzed for this article to measure variation in these other data between relevant "at-risk" patient groups defined from all available UW-QOLv4 data (see Tables 3 and 4).

criteria. Table 2 indicates that subgroup A is defined by a UW-QOL swallowing score of 0 and subgroup B by a score of 30, and that by choosing subgroups A and B we might expect 16% of UW-QOL questionnaires (5% in subgroup A, 11% in subgroup B) to be captured by these criteria. In practice, the rates for our preferred triggers ranged between 9% (recreation and speech) and 16% (swallowing) of questionnaires.

Our preferred triggers were used to determine variation in trigger rates between the main clinical groups at around 2 years from surgery (**Table 5**). Of 615 patients alive in January 2000, 479 had 15 months or more of follow-up (to death or alive at last follow-up) to May 2006. Of these, 386 (81%) had quality-of-life data beyond 15 months; sometimes a patient had several questionnaires that qualified, and so the closest after 24 months was chosen. The median was 29 months and the interquartile range was 23 to 45 months for the 386 patients.

Table 5 highlights considerable differences in trigger rates at around 2 years after surgery according to the clini-

**Table 2. Subgroups of Patients According to Severity of UW-QOLv4 Response[a]**

| UW-QOLv4 Domain | No. of Questionnaires | Subgroup A | | Subgroup B | | Subgroup C | | Subgroup D | | Total Selected (A-D), No. (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Criteria | No. (%) | Criteria | No. (%) | Criteria | No. (%) | Criteria | No. (%) | |
| Pain | 1597 | 0 or 25 | 71 (4) | 50+IMP | 134 (8) | 50, Not IMP | 194 (12) | 75+IMP | 55 (3) | **454** (28) |
| Appearance | 1602 | 0 or 25 | 68 (4) | 50+IMP | 115 (7) | 50, Not IMP | 165 (10) | 75+IMP | 116 (7) | **464** (29) |
| Activity | 1599 | 0 or 25 | 72 (5) | 50+IMP | 118 (7) | 50, Not IMP | 431 (27) | None | | **621** (39) |
| Recreation | 1593 | 0 or 25 | 122 (8) | 50+IMP | 17 (1) | 50, Not IMP | 274 (17) | 75+IMP | 66 (4) | **479** (30) |
| Swallowing | 1593 | 0 | 76 (5) | 30 | 172 (11) | 70+IMP | 251 (16) | None | | **499** (31) |
| Chewing | 1591 | 0 | 224 (14) | 50+IMP | 304 (19) | None | | None | | **528** (33) |
| Speech | 1587 | 0 or 30 | 139 (9) | 70+IMP | 283 (18) | None | | None | | **422** (27) |
| Shoulder | 1554 | 0 | 72 (5) | 30+IMP | 77 (5) | 30, Not IMP | 80 (5) | 70+IMP | 88 (6) | **317** (20) |
| Taste | 1589 | 0 | 120 (8) | 30+IMP | 89 (6) | 30, Not IMP | 209 (13) | 70+IMP | 50 (3) | **468** (29) |
| Saliva | 1550 | 0 | 86 (6) | 30+IMP | 151 (10) | 30, Not IMP | 101 (7) | 70+IMP | 154 (10) | **492** (32) |
| Mood | 1577 | 0 or 25 | 198 (13) | 50+IMP | 38 (2) | 50, Not IMP | 116 (7) | 75+IMP | 81 (5) | **433** (27) |
| Anxiety | 1577 | 0 | 63 (4) | 30 | 180 (11) | 70+IMP | 124 (8) | None | | **367** (23) |

Abbreviations: IMP, Importance; UW-QOLv4, University of Washington Quality of Life Questionnaire version 4.
[a]Information is from questionnaires dated January 2000 to May 2006 from patients with oral/oropharyngeal cancer first treated in 1992 to 2005. The criteria are formed from numerical UW-QOLv4 domain scores (eg, 0 or 25) and from an amalgam of UW-QOLv4 domain scores and IMP, eg, 50+IMP means a score of 50 and the domain is important to the patient. Subgroup A includes those most at risk, and subgroup D, the least at risk.

**Table 3. Summary Statistics for Social-Emotional Measures of Pain, Activity, Recreation, Shoulder Function, Mood, and Anxiety Collected Concurrent With UW-QOLv4 by Patient At-Risk Subgroups Reflecting Gradation in UW-QOLv4 Response[a]**

| UW-QOLv4 Domain | Concurrent Measure[b] | Summary Statistic | Statistic (No.) | | | | |
|---|---|---|---|---|---|---|---|
| | | | Subgroup A | Subgroup B | Subgroup C | Subgroup D | Other Patients |
| Pain | EQ-5D | % Extreme or moderate pain/discomfort | 100 (6/6) | 94 (16/17) | 100 (25/25) | 75 (6/8) | 35 (57/165) |
| Activity | EQ-5D | % Some problems with or unable to perform usual activities | 100 (9/9) | 86 (18/21) | 80 (41/51) | None | 21 (30/140) |
| | EQ-5D | % Problems walking about or confined to bed | 89 (8/9) | 76 (16/21) | 69 (35/51) | None | 21 (29/140) |
| Recreation | EQ-5D | % Some problems with or unable to perform usual activities | 100 (16/16) | 100 (4/4) | 94 (32/34) | 56 (5/9) | 26 (41/158) |
| Shoulder | SDQ | Median [IQR] | 85 [58-91] (8) | 50 [39-84] (4) | 39 [32-80] (10) | 7, 40 (2) | 0 [0-14] (76) |
| | NDII | Median [IQR] | 83 [77-94] (8) | 54 [29-71] (4) | 44 [30-73] (10) | 8, 38 (2) | 6 [0-17] (76) |
| Mood | EORTC EF | Median [IQR] | 29 [8-65] (20) | 67 (1) | 67 [58-83] (17) | 63 [54-81] (6) | 83 [75-100] (81) |
| | EQ-5D | % Moderately or extremely anxious or depressed | 97 (28/29) | 60 (6/10) | 58 (7/12) | 44 (4/9) | 18 (29/161) |
| | HADS | % >11 Moderate/severe depression | 43 (6/14) | 0 (0/3) | 18 (3/17) | 30 (3/10) | 5 (7/148) |
| | CES-D | % ≥16 (Caseness) | 80 (12/15) | 100 (3/3) | 44 (7/16) | 45 (5/11) | 18 (25/142) |
| Anxiety | EORTC EF | Median [IQR] | 17 [4-33] (9) | 63 [33-77] (10) | 67 [58-75] (9) | None | 83 [67-98] (96) |
| | EQ-5D | % Moderately or extremely anxious or depressed | 100 (4/4) | 80 (16/20) | 100 (12/12) | None | 23 (42/186) |
| | HADS | % >11 Moderate/severe anxiety | 75 (3/4) | 63 (10/16) | 27 (3/11) | None | 7 (11/154) |

Abbreviations: IQR, interquartile range; see Table 1 for remaining definitions.
[a]Results come from various studies dated 2000 to 2006 (see Table 1). Subgroups A (highest risk) through D are domain dependent and are defined in Table 2.
[b]See Table 1 footnote for a more complete description.

cal characteristics of the patient, particularly adjuvant radiotherapy. A high percentage (81%) of patients who had undergone free-flap surgery with adjuvant therapy for T3 or T4 tumors met the trigger criteria, with 42% meeting criteria on 3 or more domains. At around 2 years there was also considerable variation in trigger rates by age and

**Table 4. Summary Statistics for Physical Measures of Appearance, Swallowing, Chewing, Speech, Taste, and Saliva Collected Concurrent With UW-QOLv4 by Patient At-Risk Subgroups Reflecting Gradation in UW-QOLv4 Response[a]**

| UW-QOLv4 Domain | Concurrent Measure[b] | Summary Statistic | Statistic (No.) | | | | |
|---|---|---|---|---|---|---|---|
| | | | Subgroup A | Subgroup B | Subgroup C | Subgroup D | Other Patients |
| Appearance | DAS-24 | Median [IQR] | 59 [42-70] (14) | 44 [32-51] (17) | 37 [30-47] (29) | 28 [22-33] (13) | 23 [18-28] (169) |
| Swallowing | MDADI | Median [IQR] | 39 [33-48] (8) | 38 [26-61] (12) | 53 [41-63] (25) | None | 78 [63-89] (27) |
| | SWAL-QOL | Median [IQR] | 33 [22-54] (9) | 45 [25-59] (13) | 57 [24-63] (25) | None | 84 [69-91] (27) |
| Chewing | LORQv3 | % Often or always difficulty chewing | 80 (4/5) | 44 (4/9) | None | None | 3 (12/354) |
| | SWAL-QOL | % Blended or tube-fed food | 94 (15/16) | 25 (3/12) | None | None | 14 (6/42) |
| | Mouth opening (mm) | Median [IQR] | 24 [20-34] (21) | 30 [25-38][c] (54)[c] | None | None | 40 [35-49] (25) |
| Speech | VHI | % Significant impairment (score of ≥61) | 63 (5/8) | 36 (5/14) | None | None | 10 (5/52) |
| | VHI | Median [IQR] | 65 [51-88] (8) | 53 [27-69] (14) | None | None | 14 [2-44] (52) |
| | V-RQOL | Median [IQR] | 53 [38-59] (9) | 36 [21-51] (14) | None | None | 10 [0-28] (54) |
| Taste | EORTC, H&N 14 | % Quite a bit or very much had problems with sense of taste | 100 (11/11) | 100 (9/9) | 65 (15/23) | 40 (2/5) | 6 (5/80) |
| Saliva | XeQoLS | Median [IQR] | 2.0 [1.5-3.6] (12) | 2.0 [1.1-2.8] (20) | 2.5 [1.2-3.4] (17) | 0.6 [0.3-1.2] (39) | 0.5 [0.1-1.1] (144) |
| | EORTC H&N 35 | % Quite a bit or very much had dry mouth | 100 (9/9) | 100 (13/13) | 79 (11/14) | 45 (5/11) | 20 (16/79) |
| | EORTC H&N 35 | % Quite a bit or very much with sticky saliva | 56 (5/9) | 42 (5/12) | 27 (4/15) | 20 (2/10) | 15 (12/82) |
| | LORQv3 | % Often or always had dry mouth | NA (0) | 75 (3/4) | 83 (5/6) | 80 (8/10) | 5 (19/348) |

Abbreviations: IQR, interquartile range; NA, not applicable; see Table 1 for remaining definitions.
[a] Results come from various studies dated 2000 to 2006 (see Table 1). Subgroups A (highest risk) through D are domain dependent and are defined in Table 2.
[b] See Table 1 footnote for a more complete description.
[c] This is for all patients scoring 50 on the UW-QOL chewing domain irrespective of importance. This particular study asked only the UW-QOL chewing domain question and did not ask about the importance of chewing.

sex for appearance and by age for mood and anxiety. The trigger rate for appearance was 13% to 20% in females and in males younger than 65 years compared with only 1% (1 of 88) in older males. The trigger rate for mood was 20% (40 of 197) in patients younger than 65 years and 7% (12 of 168) in older patients. For anxiety the trigger rate was 24% (46 of 194) in patients younger than 65 years and 7% (12 of 166) in older patients.

Our preferred triggers were applied to the 372 patients without cancer who attended general dental practice and who were asked UW-QOL questions in relation to their condition rather than to cancer. The trigger rates by domain for these patients without cancer (with trigger rates after 2 years for patients with cancer given in parentheses) were as follows: pain, 18% (12%); appearance, 5% (12%); activity, 9% (13%); recreation, 6% (10%); swallowing, 1% (14%); chewing, 1% (13%); speech, 2% (10%); shoulder, 5% (12%); taste, 0% (12%); saliva, 1% (16%); mood, 12% (15%); and anxiety, 12% (16%).

## COMMENT

Patient-reported outcomes are a vital component in our understanding of outcomes after treatment.[19] Health-related quality of life is an important facet of patient-reported outcomes. This is the first study of this type, to our knowledge, to postulate cutoff scores by means of the UW-QOL version 4 suitable for screening at-risk patients in a routine clinical setting. We have adopted a systematic approach in the data handling based on previous studies (Table 1). A large number of questionnaires have been brought together for analysis. However, we recognize that this includes relatively few patients with baseline data. This reflects the way the questionnaire has been used in our practice, where the focus has been on post-treatment dysfunction. It has already been recognized[20] that the main issues at baseline are pain, mood, and anxiety rather than the aspects of head and neck function. Surveys tend to take place annually, and there is inevitable loss of data at different time points. Nonresponses are liable to represent bias in some ways, but at least our data contain responses from around 80% of the cohort at some stage during their cancer process.

The idea of using a questionnaire to screen patients is not new. Although cutoffs can be calculated for other validated head and neck questionnaires, it is novel to use the UW-QOL version 4. The questionnaire briefly covers the range of common problems associated with cancer and its treatment. The single-item questions are a potentially suitable screening tool because they have been shown to correlate well with specific measures, namely the European Organization for Research and Treatment

**Table 5. Preferred Trigger Rates for Patients Approximately 2 Years After Surgery by Main Clinical Characteristics[a]**

| UW-QOLv4 Domain | Trigger Criteria[b] | Oral, % (No.) | | | | | Oropharynx, % (No.) (n=46) |
|---|---|---|---|---|---|---|---|
| | | T1-T2 | | | T3-T4, Flap | | |
| | | No Flap, No RT (n=101) | Flap | | | | |
| | | | No RT (n=82) | RT (n=42) | No RT (n=46) | RT (n=57) | |
| Pain | 0, 25, 50+IMP | 8 (8/100) | 16 (13/80) | 5 (2/40) | 11 (5/46) | 20 (11/56) | 13 (6/45) |
| Appearance | 0, 25, 50+IMP | 2 (2/101) | 11 (9/81) | 10 (4/40) | 15 (7/46) | 29 (16/56) | 13 (6/46) |
| Activity | 0, 25, 50+IMP | 9 (9/100) | 15 (12/80) | 8 (3/40) | 15 (7/46) | 18 (10/56) | 17 (8/46) |
| Recreation | 0, 25, 50+IMP | 5 (5/101) | 5 (4/81) | 3 (1/40) | 20 (9/46) | 16 (9/57) | 13 (6/46) |
| Swallowing | 0, 30 | 4 (4/100) | 8 (6/80) | 10 (4/42) | 9 (4/46) | 35 (20/57) | 35 (16/46) |
| Chewing | 0 | 4 (4/101) | 4 (3/81) | 12 (5/42) | 11 (5/46) | 39 (22/57) | 24 (11/45) |
| Speech | 0, 30 | 4 (4/101) | 7 (6/82) | 10 (4/40) | 11 (5/46) | 24 (13/54) | 11 (5/46) |
| Shoulder | 0, 30+IMP | 8 (8/96) | 17 (13/77) | 15 (6/41) | 9 (4/45) | 15 (8/55) | 11 (5/45) |
| Taste | 0, 30+IMP | 6 (6/101) | 6 (5/82) | 14 (6/42) | 9 (4/46) | 30 (17/57) | 13 (6/45) |
| Saliva | 0, 30+IMP | 7 (7/98) | 4 (3/78) | 32 (13/41) | 9 (4/43) | 30 (16/54) | 37 (16/43) |
| Mood | 0, 25, 50+IMP | 13 (13/100) | 13 (10/79) | 22 (9/41) | 9 (4/45) | 20 (11/54) | 11 (5/46) |
| Anxiety | 0 or 30 | 15 (15/100) | 9 (7/80) | 20 (8/41) | 14 (6/44) | 24 (12/50) | 22 (10/45) |
| ≥1 Domain triggered | | 42 (42/101) | 51 (42/82) | 64 (27/42) | 52 (24/46) | 81 (46/57) | 70 (32/46) |
| ≥3 Domains triggered | | 13 (13/101) | 17 (14/82) | 21 (9/42) | 17 (8/46) | 42 (24/57) | 35 (16/46) |

Abbreviations: IMP, importance; RT, radiotherapy; UW-QOLv4, University of Washington Quality of Life Questionnaire version 4.

[a]Time after surgery was a median of 29 months with an interquartile range of 23 to 45 months. Each patient is represented only once, with the follow-up closest to 24 months selected for analysis. The table excludes 12 other patients: 5 in the category T3-T4, oral, no flap; 5 T1-T2, oral, no flap; and 2 maxillary.

[b]These are our preferred trigger criteria. The criteria are formed from numerical UW-QOLv4 domain scores (eg, 0 or 25) and from an amalgam of domain scores and IMP, eg, 50 + IMP means a score of 50 and the domain is important to the patient.

of Cancer Quality of Life Questionnaire–Head and Neck (dry mouth),[10] EORTC H&N (taste),[10] European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30 (emotional),[10] Hospital Anxiety and Depression Scale and Center for Epidemiology Studies Depression Scale,[11] EQ-5D Euro-Qol,[12] Derriford Appearance Scale,[13] Xerostomia-Related Quality of Life Questionnaire (S.N.R. and D.L., unpublished data, 2008), M. D. Anderson Dysphagia Inventory and Swallowing (Dysphagia) Specific Quality of Life,[14] Voice-Related Quality of Life Measure and Voice Handicap Index,[15] Shoulder Disability Questionnaire and Neck Dissection Impairment Index,[16] Liverpool Oral Rehabilitation Questionnaire version 3,[17] and mouth opening.[18] The way the questions are worded in the UW-QOL helps to provide an indicator of outcome that patients, caregivers, and tumor board members can easily understand and apply in clinical practice.[4]

Applying cutoff criteria allows the patients reporting problems to be identified. This is part of the continued progress toward incorporating health-related quality-of-life data into routine clinical practice. The choice of where to set the cutoff is open to debate. We combined the patient score in each item together with whether it was one of the 3 important issues identified by the patient. The rationale for this was that, if there is mild dysfunction, it still might be appropriate to factor it in if the patient identifies it as an important issue. When looking at suitable cutoffs, we also considered the wording of the questionnaire. Finally, given the range of issues within the questionnaire and the likelihood of patients reporting problems in several of these, particularly after radical treatment for advanced disease, we took a pragmatic approach for clinical practice in not identifying too many patients. Our initial selection of at-risk subgroups A to D given in Table 2 only include one-quarter to one-third of patients for each domain. However, even this selection would result in a large proportion being identified overall. In Tables 3 and 4 we compared our groups A to D with concurrent data from a series of studies to further build up an evidence base from which we could refine our trigger criteria for identifying patients with problems. In practice, the rates for our final preferred triggers ranged between 9% (recreation and speech) and 16% (swallowing) of questionnaires.

The implications of how to use the algorithm in a busy clinic need careful thought. For example, with our preferred algorithm (Table 5), 42% of patients with early disease (T1 and T2, oral, no free flap, and no adjuvant radiotherapy) had 1 or more domains that met the trigger criteria compared with 81% with advanced disease (T3 and T4, oral, free flap, and adjuvant radiotherapy). The percentage that met trigger criteria for 3 or more domains was much lower but was still 42% in advanced disease.

Each domain was flagged at between 9% and 16%, which was part of our intention in setting the cutoff. Our final preferred algorithm was influenced by the concurrent questionnaires (Tables 3 and 4). Where we perceived there to be a big jump between groups (A to D) was where we thought it suitable to put a cutoff. We were also mindful to derive domain criteria that would trigger intervention in no more than 1 in 5 patients. In practice, the logical choice of cutoff for most domains seemed to be either at group B or C, but B was preferred to keep the trigger rate to a manageable proportion. The exceptions were chewing and speech, where trigger A was chosen because group B contained too many patients and it was impossible to subdivide group B any further. The choice of B in addition to A

would have selected 33% for chewing and 27% for speech. It can be argued that these are important domains and that patients with problems are better identified than missed. It is possible also that these percentages would be less for other head and neck cancer sites.

We recognize in general the subjectivity in our preferred criteria, and our results are presented as far as possible to allow others to vary their criteria for triggering more or fewer patients. We did not start out with any specific hypothesis as to which cutoffs would be useful, and we have let the data dictate our final selections. As such, this can be seen as hypothesis generation, and future research may confirm or refute the suitability of our choice of criteria. These criteria are not fixed and may change in the future as implications of the current preferred criteria are experienced in practice in a busy clinic setting. Even if there is no need to actively intervene, for some patients recognizing and discussing their issues may be beneficial in itself. Our preferred criteria have already been integrated into touch-screen technology as part of a wider patient concerns inventory, and the algorithm triggers form part of an output available to the consultant during clinic consultations. The results of this experience will be published in due course.

The ability to use touch-screen technology to acquire patient-derived data in real time in the clinic has tremendous potential to improve patient care. It is possible to use a questionnaire such as the UW-QOL version 4 to screen patients for dysfunction. To assist in the incorporation of these data into clinical practice, it is feasible to define criteria so that patients and their problem domains can be easily identified. This allows the clinician and members of the head and neck team a chance to discuss these issues with the patient and, where appropriate, arrange further support or intervention.

**Submitted for Publication:** July 1, 2008; final revision received November 18, 2008; accepted November 24, 2008.

**Correspondence:** Simon N. Rogers, FDS, RCS, FRCS, MD, Regional Maxillofacial Unit, Aintree University Hospitals NHS Foundation Trust, Aintree, Liverpool L9 7AL, England (snrogers@doctors.org.uk).

**Author Contributions:** Dr Rogers and Mr Lowe had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Rogers and Lowe. *Acquisition of data:* Rogers. *Analysis and interpretation of data:* Rogers and Lowe. *Drafting of the manuscript:* Rogers and Lowe. *Critical revision of the manuscript for important intellectual content:* Rogers and Lowe. *Statistical analysis:* Lowe. *Study supervision:* Rogers.

**Financial Disclosure:** None reported.

**Previous Presentation:** This study was presented at the Seventh International Conference on Head and Neck Cancer of the American Head and Neck Society; July 20, 2008; San Francisco, California.

## REFERENCES

1. Rogers SN, Ahad SA, Murphy AP. A structured review and theme analysis of papers published on "quality of life" in head and neck cancer: 2000 to 2005. *Oral Oncol.* 2007;43(9):843-868.
2. Bjordal K, Hammerlid E, Ahlner-Elmqvist M, et al. Quality of life in head and neck cancer patients: validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-H&N35. *J Clin Oncol.* 1999;17(3):1008-1019.
3. Terrell JE, Ronis DL, Fowler KE, et al. Clinical predictors of quality of life in patients with head and neck cancer. *Arch Otolaryngol Head Neck Surg.* 2004;130(4):401-408.
4. Rogers SN, Scott J, Chakrabati A, Lowe D. The patient's account of outcome following primary surgery for oral and oropharyngeal cancer using a "quality of life" questionnaire. *Eur J Cancer Care (Engl).* 2008;17(2):182-188.
5. Rogers SN, McNally D, Mahmood M, Chan M, Humphris GM. Psychologic response of the edentulous patient after primary surgery for oral cancer: a cross-sectional study. *J Prosthet Dent.* 1999;82(3):317-321.
6. Millsopp L, Brandom L, Humphris GM, Lowe D, Rogers SN. Facial appearance after operations for oral and oropharyngeal cancer: a comparison of casenotes and patient-completed questionnaire. *Br J Oral Maxillofac Surg.* 2006;44(5):358-363.
7. Millsopp L, Frackleton S, Lowe D, Rogers SN. A feasibility study of computer-assisted health-related quality of life data collection in patients with oral and oropharyngeal cancer. *Int J Oral Maxillofac Surg.* 2006;35(8):761-764.
8. Ringash J, Bezjak A. A structured review of quality of life instruments for head and neck cancer patients. *Head Neck.* 2001;23(3):201-213.
9. Kanatas AN, Rogers SN. A national survey of health-related quality of life questionnaires in head and neck oncology. *Ann R Coll Surg Engl.* 2004;86(1):6-10.
10. Rogers SN, Gwane S, Lowe D, Humphris G, Yueh B, Weymuller EA. The addition of mood and anxiety domains to the University of Washington Quality of Life Scale. *Head Neck.* 2002;24(6):521-529.
11. Rogers SN, Rajlawat B, Goru J, Lowe D, Humphris GM. Comparison of the domains of anxiety and mood of the University of Washington Head and Neck Cancer Questionnaire (UW-QOL V4) with the CES-D and HADS. *Head Neck.* 2006;28(8):697-704.
12. Rogers SN, Miller RD, Ali K, Minhas AB, Williams HF, Lowe D. Patients' perceived health status following primary surgery for oral and oropharyngeal cancer. *Int J Oral Maxillofac Surg.* 2006;35(10):913-919.
13. Katre C, Johnson IA, Humphris GM, Lowe D, Rogers SN. Assessment of problems with appearance following surgery for oral and oropharyngeal cancer using the University of Washington appearance domain and the Derriford Appearance Scale. *Oral Oncol.* 2008;44(10):927-934.
14. Thomas L, Jones T, Tandon S, Katre C, Lowe D, Rogers SN. An evaluation of the University of Washington Quality of Life swallowing domain following oropharyngeal cancer. *Eur Arch Otorhinolaryngol.* 2008;265(suppl 1):S29-S37.
15. Thomas L, Jones T, Tandon S, Carding P, Lowe D, Rogers SN. Speech and voice outcomes in oropharyngeal cancer and evaluation of the University of Washington Quality of Life speech domain. *Clin Otolaryngol Allied Sci.* In press.
16. Rogers SN, Scott B, Lowe D. An evaluation of the shoulder domain of the University of Washington quality of life scale. *Br J Oral Maxillofac Surg.* 2007;45(1):5-10.
17. Rogers SN, O'Donnell JP, Williams-Hewitt S, Christensen JC, Lowe D. Health-related quality of life measured by the UWQOL: reference values from a general dental practice. *Oral Oncol.* 2006;42(3):281-287.
18. Scott B, Butterworth C, Lowe D, Rogers SN. Factors associated with restricted mouth opening and its relationship to health-related quality of life in patients attending a maxillofacial oncology clinic. *Oral Oncol.* 2008;44(5):430-438.
19. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B; Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: conceptual issues. *Value Health.* 2007;10(suppl 2):S66-S75.
20. Rogers SN, Lowe D, Brown JS, Vaughan ED. The University of Washington head and neck cancer measure as a predictor of outcome following primary surgery for oral cancer. *Head Neck.* 1999;21(5):394-401.

# Measuring the Quality of Life of Cancer Patients: The Functional Living Index–Cancer: Development and Validation

By H. Schipper, J. Clinch, A. McMurray, and M. Levitt

**The classical criteria for the evaluation of clinical trials in cancer reflect alterations in physical well-being, but are insensitive to other important factors, such as psychosocial state, sociability, and somatic sensation that may play a critical role in determining the patients' functional response to their illness and its treatment. The Functional Living Index–Cancer is designed for easy, repeated patient self-administration. It is a 22-item questionnaire that has been validated on 837 patients in two cities over a three-year period. Criteria for validity include stability of factor analysis, concurrent validation studies against the Karnofsky, Beck Depression, Spielberger State and Trait Anxiety, and Katz Activities of Daily Living scales, as well as the scaled version of The General Health Questionnaire and The McGill/Melzack Pain Index. The index is uncontaminated by social desirability issues. The validation studies demonstrate the lack of correlation between traditional measures of patient response and other significant functional factors such as depression and anxiety ($r = 0.33$), sociability and family interaction, and nausea. These findings elucidate the frequently observed discrepancies between traditional assessments of clinical response and overall functional patient outcome. The index is proposed as an adjunct to clinical trials assessment and may provide additional patient functional information on which to analyse the outcome of clinical trials or offer specific advice to individual patients.**

ADVANCES in cancer treatment have brought about significant improvements in the overall survival expectation for many malignant diseases. In some instances, such as Hodgkin's disease, acute lymphoblastic leukemia of children, and testicular tumors the uniformly fatal outlook of a generation ago has been replaced by the reasonable expectation of cure in the majority of cases. Much of the progress made in the treatment of these illnesses can be attributed to the clinical trials approach, wherein advances are made in a step-wise fashion using the vehicle of comparative, usually randomized, clinical trials. Parameters usually taken into consideration in determining the efficacy of new treatments include survival, disease-free survival, response rate, remission-induction rate and duration, and treatment toxicity.[1] The overall treatment model basically adheres to an acute-disease approach, wherein the strategy is to prescribe a maximal intervention soon after diagnosis in an attempt to effect cure. Particularly for childhood acute lymphoblastic leukemia this approach is reasonable. However, it is increasingly recognized that most of the more common malignancies follow a more chronic pattern and our drug interventions are neither so successful nor as short term as in an acute-disease setting. For many cancers where cure is not achieved (for example, breast cancer) meaningful palliation, often with considerable prolongation of life can be achieved. In other conditions, such as lung cancer, the survival advantages achieved, though statistically significant, are small and may be of little human significance. Particularly in this setting, deficiencies in our usual assessment parameters become apparent. Our classical criteria seem acceptable in an acute-disease setting, but the data we interpret from them do not take into account issues such as patient morbidity and withdrawal from clinical trials, particularly where the disease acts in a chronic manner, and the treatment is administered over a prolonged period of time. Our usual measures of functional state such as the Karnofsky and Eastern Cooperative Oncology

*From the Manitoba Cancer Treatment and Research Foundation, University of Manitoba, Winnipeg, Canada.*

*Submitted May 16, 1983; accepted December 9, 1983.*

*Supported in part by the National Cancer Institute of Canada; St. Boniface General Hospital Research Foundation; Manitoba Medical Services Foundation Inc, and Manitoba Cancer Treatment and Research Foundation, and Health and Welfare (Canada). Dr. Schipper is a Clinical Trials Scholar of The National Cancer Institute of Canada.*

*Address reprint requests to H. Schipper, MD, Department of Medicine, St. Boniface General Hospital, 409 Tache Avenue, Winnipeg, Manitoba R2H 2A6.*



*0732-183X/84/0205-0020$3.00/0*

Group (Zubrod) scales, are not widely validated and are designed to reflect the medical physical aspects of illness, rather than the overall functional state of the patient.

Psychosocial considerations are being integrated into the assessment of cancer patients. For the most part, at present, investigators are concentrating on case histories, examining such factors as crisis intervention, depression, sexuality, and nausea and vomiting.[2–6] The nausea and vomiting may be viewed as a phenomenon representing the interface between traditional medical criteria and psychosocial factors.

Other contributory components to a global quality of life measure include freedom from pain, sociability, "impact" of illness, and satisfaction. There is in fact considerable debate as to whether "quality of life" is a distinct entity, different from measures such as the above, or whether it reflects a composite of factors that individuals view as important to the reality of functional living. Little wonder then that to date there are no validated and accepted tools that provide an opportunity to compare groups of patients vis-á-vis their overall functional response both to their cancer and to different treatment approaches.[7–9]

We have attempted to devise and validate a functional living index for cancer (FLIC) having the following properties and characteristics.

(1.) It is cancer specific, that is, the measure is specific enough to the cancer population to detect differences in functional state among cancer patients of a given disease group. In other words, unlike broad-based medical quality of life indices that are designed to measure the medical functional state of free-living populations, this test should take into account that patients have already been diagnosed as having malignant disease and should concentrate on distinguishing functional states within this population.

(2.) The index should be functionally oriented, addressing itself to those day-to-day living issues that represent the global construct of functional quality of life.

(3.) It should be designed for patient self-administration and not require the intervention of interviewers or health professionals for its administration.

(4.) The questions designed should be of general applicability, ease and consistency of interpretation, and of a number small enough to permit high compliance despite repeated administration.

(5.) It should be repeatable, in order that the patient's score derived can be followed over a period of time to elicit and evaluate trends both within patients and between groups.

(6.) It should be sensitive across the range of clinical practice being able to distinguish not only patients who are obviously well from those terminally ill, but more significantly, degrees of dysfunction between patients with varying extents of disease and intensities of therapeutic intervention.

(7.) The instrument designed should have adequately demonstrated face, content, construct, and concurrent validity as well as reliability.

## MATERIALS AND METHODS

### *A Preliminary Literature Review and Patient Interviews*

To define a "universe of concern" about the functional status of cancer patients, the medical, psychosocial, and philosophic literature was reviewed and a series of patient interviews was undertaken to define areas of day-to-day function that were deemed of consequence. From this process four principal areas of functional importance were ultimately defined: I, vocation/activity; II, affect/psychologic state; III, social interaction; and IV, somatic sensation. A subsidiary area specific to chemotherapy patients, "nausea," was also identified.

### *The Panel*

It was felt critically important to establish a questionnaire-design panel that adequately represented the "universe of interest" vis-á-vis cancer patients. The panel had to be small enough to be workable and to represent the input of all persons playing an active role in the functional existence of the cancer patient. Accordingly, a panel of 11 people was constructed. It consisted of a male patient, a female patient, and two patient spouses representing both the urban and rural populations; two physicians; a statistician/physiologist; an oncology nurse; a psychologist; a public health/Victorian Order of Nurses nurse assessor; and a clergyman. The panel was mandated to review the areas of concern identified and to formulate questions that from their own experience they thought might be meaningful to patients in the context of a functional quality of life questionnaire. The panelists were encouraged to devise questions suitable for scoring on analogue scales but they were not rigidly restricted as to format or number. In addition, the panelists were free to include questions that they felt represented other areas beyond those clearly falling within the defined groups.

### *The First-Generation Questionnaire*

Approximately 250 questions were provided by the panel which then reduced the number by eliminating duplication, un-

clear questions, and questions of limited applicability, to an initial bank of 92 items. This "first generation" questionnaire was administered to 175 patients seen in the outpatient clinics at the Manitoba Cancer Treatment and Research Foundation (MCTRF). Patients were 16 years of age and older. They were unselected, with the following exceptions: patients too ill to answer the questions, those unable to speak or understand English, or those who refused for any other reasons. One hundred seventy-five patients answered the questionnaire and the subsequent analysis altered or eliminated a large number of questions found to be unsuitable on the basis of deficient generalizability or clarity. In particular some demographic questions could not be answered by everyone. This led directly to the "second generation" questionnaire.

### *The Second-Generation Questionnaire*

Demographic items were removed from the body of the questionnaire, reconstructed in yes/no format, and appended as a frontispiece to the questionnaire. The purpose of the second-generation questionnaire was to undertake an initial factor analysis of the responses obtained with a view to reducing the number of questions. The questionnaire numbered 42 items and was administered to 312 patients, including both inpatients and outpatients, at the MCTRF. It was administered over a three-month period and each patient faced a single questionnaire exposure. Fewer than 5% of patients approached to complete the questionnaire were unable to do so. The results underwent factor analysis and those items loading most heavily on each of the factors were selected for the "third generation," 20-item document.

### *The Third-Generation Questionnaire: Winnipeg Validation*

The purpose of the third-generation questionnaire was externally to validate the quality of life questionnaire and its component factors against established and accepted measures of the content areas considered when constructing the questionnaire, and to test the stability of the factor analysis. This questionnaire numbered 20 items and was administered to 175 patients both in hospital and in the outpatient setting of the MCTRF and the Princess Elizabeth Chronic Care Hospital in Winnipeg. Multiple observers took part in each patient observation. At one sitting patients were asked to complete the following battery of tests: a Katz Activities of Daily Living Index[10] administered by the clinic or ward nurse; a Karnofsky index administered by the patient's physician;[11] a McGill-Melzack Pain questionnaire administered by study personnel;[12] the quality of life questionnaire, the Spielberger State and Trait anxiety tests,[13] the Beck Depression Scale,[14] and the scaled version of the General Health Questionnaire,[15] all patient self-administered. This battery of tests was selected to provide concurrent validation against established measures particularly in the areas of functional ability and psychosocial state. Factor analysis and correlation studies led to the development of the fourth-generation questionnaire.

### *The Fourth-Generation Questionnaire: Edmonton Validation*

The purpose of this run was to expose the questionnaire to a different patient population, and to further examine concurrent validity, stability of factor analysis, and compliance. Modifications of the questionnaire amplifying the issues of nausea and recreational activity were added, based on the analysis of the previous run. The McGill-Melzack Pain Index was deleted for logistic reasons and was replaced with the Jackson social desirability measure.[16–18] One hundred seventy-five patients at the W.W. Cross Cancer Institute in Edmonton, Alberta representing in- and outpatient populations completed the 26-item questionnaire. Over three weeks, the questionnaire was administered to each patient at one sitting.

### *Scoring*

Each item on the questionnaire was answerable in Likert format of range 1 to 7. Patients were instructed to answer all questions by marking with a vertical slash on the scoring line that point that best represented their response. For scoring, each interval was divided in half and responses were scored to the nearest even whole integer. The scale was reversed on certain items for the second Winnipeg and the Edmonton run so that a high score consistently represented a higher quality of life. Thus, scores on individual questions ranged from 1 to 7 and the overall score was derived by summing scores on all questions.

### *Statistical Methods*

*Factor analysis.* Factor analyses were performed using the SPSS (Statistical Package for the Social Sciences) Factor Analysis program. The method of factoring selected was an iterative principle factor method that factors the reduced correlation matrix (that is, the correlation matrix with the ones in the diagonal replaced by the squared multiple correlation of each variable with all other variables, which is a lower-bound estimate of the variables communality). In general, factors were retained that had eigenvalues greater than one, and the factors were rotated to a final solution using the orthogonal varimax criterion.[19]

*Correlation coefficients.* Pearson product moment correlation coefficients were calculated between all tests, subtests, factor scores, and individual test items using the SPSS Pearson Correlation Program. For the social desirability check the Differential Reliability Index was calculated. This is the square root of the difference between the squared correlation of an item with its own scale and the squared correlation of an item with the social desirability scale. A value above 0.30 indicates that an item is not contaminated by social desirability to the extent that it should be removed from the scale.[16]

*Comparison of means.* Mean FLIC scores for the stratification groups were compared using a one-way analysis of variance. Separate analyses were performed for the Winnipeg and Edmonton data.

## RESULTS

### *Factor Analysis*

Factor analysis of the second-run questionnaire revealed that demographic variables of age, sex, marital status, employment status, and place of residence had a factorial composition that did not overlap with that of the items in the main body of the questionnaire. Thus, demographic variables do not contribute to the dimensions resolved from the main questionnaire. Therefore, in subsequent factor analyses, the demographic variables were omitted. The factors that emerged

**Table 1. Factor Loadings of Questions on Physical Well Being and Ability Factor(s) for Each Data Set: Validation Studies**

| Questions | Winnipeg Run 1 | | Winnipeg Run 2, Factor 1 | Edmonton Factor 1 |
|---|---|---|---|---|
| | Factor 1 | Factor 4 | | |
| How much pain today? | 0.225 | 0.300 | ... | ... |
| Pain disrupting activity? | ... | ... | *0.719* | *0.790* |
| Cancer-related pain? | ... | ... | *0.591* | *0.662* |
| Uncomfortable? | ... | ... | *0.659* | 0.498 |
| Maintains leisure activities? | ... | ... | ... | *0.741* |
| Feel well today? | 0.384 | *0.601* | *0.826* | *0.735* |
| Well enough for meals or repairs? | *−0.542* | −0.424 | *0.801* | *0.560* |
| Satisfied with work? | *0.667* | 0.174 | *0.627* | *0.552* |
| Able to complete housework? | *0.641* | 0.222 | *0.754* | *0.552* |
| Thinks appears well? | 0.341 | *0.601* | *0.703* | *0.538* |
| Afraid of future? | 0.100 | 0.078 | 0.035 | −0.068 |
| Angry, frightened, or depressed? | −0.132 | −0.002 | 0.037 | ... |
| Afraid? | ... | ... | ... | 0.043 |
| Depressed? | ... | ... | ... | 0.176 |
| Angry? | ... | ... | ... | 0.087 |
| Discouraged? | 0.304 | 0.130 | 0.227 | 0.477 |
| Thinks about illness? | ... | ... | ... | 0.244 |
| Copes well with stress? | ... | ... | ... | 0.328 |
| "Family" hardship from cancer? | −0.315 | −0.067 | 0.425 | 0.270 |
| "Family" disruption from cancer? | −0.218 | −0.147 | *0.516* | 0.316 |
| Personal hardship from cancer? | 0.355 | 0.248 | *0.751* | *0.534* |
| How much nausea? | 0.130 | *0.539* | 0.345 | 0.297 |
| Nausea affecting activity? | ... | ... | ... | 0.271 |
| Spends time with "family"? | 0.090 | 0.144 | −0.011 | 0.017 |
| Spends time with friends? | 0.226 | 0.238 | 0.282 | 0.211 |
| Confident of medical staff? | −0.067 | 0.035 | 0.089 | 0.184 |
| Confident of treatment? | 0.058 | 0.187 | 0.158 | 0.346 |
| Any arguments today? | 0.008 | −0.031 | 0.070 | 0.100 |

NOTE. Tables 1 through 5 list the factor loadings of those questions in the three validation studies. Items with factor loadings > 0.500 are italic. In each of the factors, those questions loading heavily do so consistently across the studies. Likewise, items having little impact on a given factor achieve consistently low weightings in each validation run. In this table, the physical well being and ability factor seemed to be a composite of factors 1 and 4, on the initial 42-question validation study. In the two successor studies, using a shorter questionnaire, these merged into factor 1. Questions that were not included in a given study are noted as (...). Questions that were deleted after the first run do not appear. For a complete description of the questionnaire process, see Materials and Methods.

in each of the three runs were the same (Tables 1–5) and reflected the universe of concern identified early in the project and agreed to by the panel that composed the instrument. We take the consistency of factor analysis across the three clinical trials to represent a significant measure of construct validity. Nausea emerged as a separate factor, fourth-ranked in the Edmonton questionnaire. This came as somewhat of a surprise, as we anticipated that nausea would be subsumed by the first factor (physical status).

*Stratification*

Patients answering the Winnipeg "second-generation" questionnaire were stratified into six groups representing broad categories of extent of disease: follow-up, off treatment; on adjuvant therapy without evident clinical disease; on active treatment for evident clinical disease; hospitalized for treatment; hospitalized due to extent of illness; and in a terminal-care unit. The average scores obtained by these groups decreased with extent of disease from a high of 116.6 to a low of 84.6. ($F = 6.373$, $p < 0.00005$). The same stratification was performed for the Edmonton data. Again, average scores of the groups fell with extent of disease ($F = 3.638$, $p = 0.0145$). Figure 1 illustrates the effect of extent of disease on FLIC scores in the two samples and indicates comparable effects in both. (The two hospitalized groups were collapsed in the Winnipeg data.)

**Table 2. Factor Loadings of Questions on Emotional State Factor for Each Data Set: Validation Studies**

| Questions | Winnipeg Run 1, Factor 2 | Winnipeg Run 2, Factor 2 | Edmonton Factor 2 |
|---|---|---|---|
| How much pain today? | 0.199 | 0.151 | 0.197 |
| Pain disrupting activity? | ... | 0.079 | 0.136 |
| Cancer-related pain? | ... | 0.151 | 0.197 |
| Uncomfortable? | ... | −0.006 | 0.307 |
| Maintains leisure activities? | ... | ... | 0.116 |
| Feel well today? | 0.255 | 0.103 | 0.139 |
| Well enough for meals or repairs? | −0.097 | 0.062 | 0.036 |
| Satisfied with work? | 0.224 | 0.346 | 0.266 |
| Able to complete housework? | 0.159 | 0.048 | 0.028 |
| Thinks appears well? | 0.238 | 0.139 | 0.235 |
| Afraid of future? | *0.652* | *0.669* | *0.789* |
| Angry, frightened, or depressed? | *0.663* | *0.679* | ... |
| Afraid? | ... | ... | *0.621* |
| Depressed? | ... | ... | *0.614* |
| Angry? | ... | ... | *0.503* |
| Discouraged? | *0.674* | *0.742* | *0.596* |
| Thinks about illness? | ... | ... | *0.552* |
| Copes well with stress? | ... | ... | *0.543* |
| "Family" hardship from cancer? | −0.232 | 0.019 | 0.163 |
| "Family" disruption from cancer? | −0.160 | 0.186 | 0.204 |
| Personal hardship from cancer? | 0.389 | 0.193 | 0.271 |
| How much nausea? | 0.029 | 0.031 | −0.036 |
| Nausea affecting activity? | ... | ... | 0.053 |
| Spends time with "family"? | 0.061 | 0.164 | 0.101 |
| Spends time with friends? | 0.086 | 0.152 | 0.282 |
| Confident of medical staff? | −0.088 | 0.133 | 0.090 |
| Confident of treatment? | 0.326 | 0.384 | 0.365 |
| Any arguments today? | −0.123 | 0.054 | −0.047 |

NOTE. For a complete description of factor loading, see NOTE to Table 1 and Materials and Methods.

*Concurrent Validation Studies*

Correlation coefficients relating the overall FLIC score to the battery of concurrent validation studies for both the Winnipeg and Edmonton trials are shown in Table 6. Correlation with the McGill/Melzack pain inventory was undertaken only in Edmonton. With the exception of the Katz Activities of Daily Living Index in the Winnipeg population, $p$ values were all $<0.0005$. Correlation of the Katz Activities of Daily Living Index with the FLIC gives Pearson $r$ values of 0.170/0.305 (Winnipeg/Edmonton) and is the lowest correlation. Subscales of the General Health Questionnaire (GHQ), B representing anxiety and insomnia and D representing severe depression, have correlations with the FLIC of 0.441/0.496 and 0.467/0.595, respectively, for the Winnipeg and Edmonton runs. All other scales have correlations with the FLIC in excess of 0.53, the highest being those of the Karnofsky Index 0.693/0.619 (Fig. 2), the GHQ total score 0.724/0.765, and the Beck Depression scale 0.724/0.773.

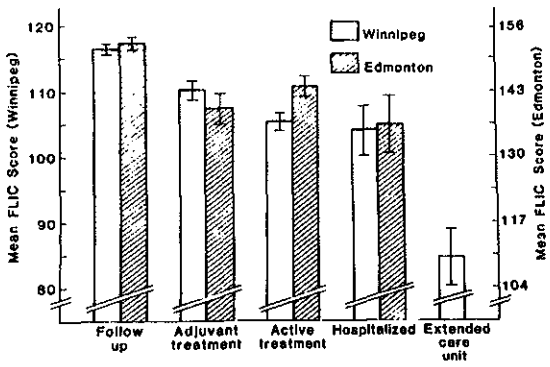



**Fig. 1. Stratification studies: the different scales are a function of the different questionnaire sizes. Scores ranged from 1–7 for each question and were summed to give an overall score.**

Correlation between FLIC factor 1, represent-

**Table 3. Factor Loadings of Questions on Sociability Factor for Each Data Set: Validation Studies**

| Questions | Winnipeg Run 1 | Winnipeg Run 2 | Edmonton |
|---|---|---|---|
| How much pain today? | 0.052 | ... | ... |
| Pain disrupting activity? | ... | 0.089 | −0.093 |
| Cancer-related pain? | ... | 0.102 | −0.175 |
| Uncomfortable? | ... | 0.133 | 0.051 |
| Maintains leisure activities? | ... | ... | 0.091 |
| Feel well today? | 0.200 | 0.217 | 0.291 |
| Well enough for meals or repairs? | −0.083 | 0.007 | 0.405 |
| Satisfied with work? | 0.038 | 0.066 | 0.323 |
| Able to complete housework? | 0.005 | 0.057 | 0.222 |
| Thinks appears well? | 0.197 | 0.068 | 0.294 |
| Afraid of future? | 0.151 | 0.040 | 0.007 |
| Angry, frightened, or depressed? | 0.012 | 0.127 | ... |
| Afraid? | ... | ... | −0.022 |
| Depressed? | ... | ... | 0.201 |
| Angry? | ... | ... | 0.135 |
| Discouraged? | 0.042 | 0.113 | 0.200 |
| Thinks about illness? | ... | ... | −0.051 |
| Copes well with stress? | ... | ... | 0.399 |
| "Family" hardship from cancer? | −0.063 | 0.112 | 0.048 |
| "Family" disruption from cancer? | −0.136 | 0.109 | 0.064 |
| Personal hardship from cancer? | 0.106 | 0.181 | 0.108 |
| How much nausea? | 0.114 | 0.357 | 0.233 |
| Nausea affecting activity? | ... | ... | 0.160 |
| Spends time with "family"? | *0.829* | *0.819* | *0.637* |
| Spends time with friends? | *0.601* | *0.571* | *0.503* |
| Confident of medical staff? | −0.063 | 0.077 | 0.129 |
| Confident of treatment? | 0.080 | 0.085 | 0.110 |
| Any arguments today? | −0.122 | 0.091 | 0.286 |

NOTE. For a complete description of factor loadings, see NOTE to Table 1 and Materials and Methods.

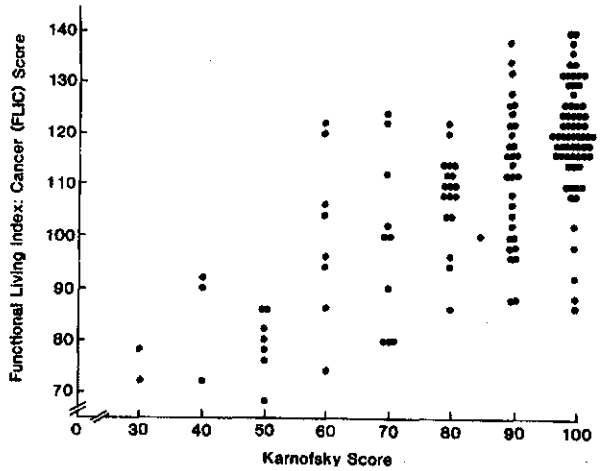



**Fig. 2. Scattergram of Karnofsky versus FLIC. This scattergram graphically represents the correlation between the overall FLIC score and physician Karnofsky scoring (r = 0.693) for the second Winnipeg validation study. While the correlation is good, the plot supports the hypothesis that the Karnofsky score only partially represent overall functional status as measured by the FLIC.**

ing physical well-being, and the Karnofsky scale (0.757/0.554), the GHQ C scale representing social dysfunction (−0.649/−0.571), and the McGill/Melzack Present Pain Index (−0.657) are high. Correlation with measures of psychologic functioning, the GHQ B Scale (anxiety and insomnia), the GHQ D (severe depression) scale, and Spielberger State and Trait Anxiety Scales are all <0.33 (Table 7).

Correlations of factor 2, representing psychologic state, are high with the GHQ D Scale (severe depression) (−0.566/−0.640), the GHQ B Scale (anxiety and insomnia) (−0.557/−0.543), the Spielberger State Anxiety Scale (−0.629/−0.588) (Fig. 3), and the Trait Anxiety Scale (−0.735/−0.609). Factor 2 has much lower correlation with the Karnofsky Scale (0.066/0.157), the GHQ C Scale (−0.279/−0.083), and all of the measures of the McGill/Melzack Pain Scale (Table 7).

Factor 3, which represents family situational

**Table 4. Factor Loadings of Questions on Family Situational Factor for Each Data Set: Validation Studies**

| Questions | Winnipeg Run 1, Factor 3 | Winnipeg Run 2, Factor 4 | Edmonton Factor 3 |
|---|---|---|---|
| How much pain today? | 0.262 | ... | ... |
| Pain disrupting activity? | ... | 0.304 | 0.219 |
| Cancer-related pain? | ... | 0.085 | 0.236 |
| Uncomfortable? | ... | 0.135 | 0.124 |
| Maintains leisure activities? | ... | ... | 0.306 |
| Feel well today? | 0.225 | 0.118 | 0.067 |
| Well enough for meals or repairs? | −0.123 | 0.141 | 0.432 |
| Satisfied with work? | 0.417 | 0.066 | 0.197 |
| Able to complete housework? | 0.318 | 0.142 | 0.441 |
| Thinks appears well? | 0.226 | 0.206 | 0.094 |
| Afraid of future? | 0.218 | 0.067 | 0.085 |
| Angry, frightened, or depressed? | −0.231 | ... | ... |
| Afraid? | ... | ... | 0.053 |
| Depressed? | ... | ... | 0.029 |
| Angry? | ... | ... | 0.093 |
| Discouraged? | 0.117 | 0.040 | −0.035 |
| Thinks about illness? | ... | ... | 0.057 |
| Copes well with stress? | ... | ... | 0.101 |
| "Family" hardship from cancer? | *−0.634* | *0.791* | *0.825* |
| "Family" disruption from cancer? | *−0.682* | *0.586* | *0.783* |
| Personal hardship from cancer? | *0.589* | 0.210 | 0.481 |
| How much nausea? | 0.061 | −0.009 | 0.132 |
| Nausea affecting activity? | ... | ... | 0.256 |
| Spends time with "family"? | 0.020 | 0.105 | 0.062 |
| Spends time with friends? | 0.157 | 0.068 | 0.128 |
| Confident of medical staff? | −0.054 | 0.122 | 0.030 |
| Confident of treatment? | 0.156 | 0.149 | 0.131 |
| Any arguments today? | −0.031 | 0.132 | 0.337 |

NOTE. For a complete description of factor loadings, see NOTE to Table 1 and Materials and Methods.

interaction, correlates only weakly with the validation tests. Factor 4 in the Winnipeg data and factor 5 in the Edmonton data represent sociability and likewise have relatively weak correlations with the validation tests. The concurrent validation battery was designed for the most part to validate the physical well being and psychologic state factors.

Results obtained on concurrent validation tests representing physical function or medical state did not correlate with those obtained representing psychosocial or social interaction states (Fig. 4).

### *Social Desirability*

A question by question social desirability analysis was undertaken for the Edmonton data using the Jackson Social Desirability Scale.[16-18] Of the 26 items in that questionnaire, four have been withdrawn from the final document partially on grounds of social desirability contamination. The remaining 22 questions are uncontaminated by social desirability issues.

## DISCUSSION

We believe that the FLIC (Fig. 5) represents a validated measure of the overall functional quality of a cancer patient's day-to-day life. We do not view it as an ultimate measure, but rather as a starting point permitting clinical trials to be compared for functional living outcomes in addition to the more traditional measures. The approach to defining "universe of concern" using a series of structured patient interviews followed by use of a panel including patients as well as health professionals may represent a unique approach. We believe that this maneuver in itself provides a substantial measure of face validity for the questionnaire. The observation that the factor analysis was stable through three separate clinical tri-

**Table 5. Factor Loadings of Questions on Nausea Factor for Edmonton Data Set: Validation Studies**

| Questions | Factor Loadings |
|---|---|
| Pain disrupting activity? | 0.269 |
| Cancer-related pain? | 0.370 |
| Uncomfortable? | 0.331 |
| Maintains leisure activities? | 0.050 |
| Feel well today? | 0.331 |
| Well enough for meals or repairs? | 0.226 |
| Satisfied with work? | −0.093 |
| Able to complete housework? | 0.117 |
| Thinks appears well? | 0.213 |
| Afraid of future? | 0.127 |
| Afraid? | 0.074 |
| Depressed? | 0.055 |
| Angry? | −0.085 |
| Discouraged? | −0.062 |
| Thinks about illness? | 0.042 |
| Copes well with stress? | 0.012 |
| "Family" hardship from cancer? | 0.241 |
| "Family" disruption from cancer? | 0.265 |
| Personal hardship from cancer? | 0.187 |
| How much nausea? | *0.777* |
| Nausea affecting activity? | *0.728* |
| Spends time with "family"? | 0.188 |
| Spends time with friends? | 0.252 |
| Confident of medical staff? | 0.058 |
| Confident of treatment? | 0.026 |
| Any arguments today? | −0.143 |

NOTE. For a complete description of factor loadings, see NOTE to Table 1 and Materials and Methods.

**Table 6. Correlations of Functional Living Index: FLIC with Concurrent Validation Tests**

| Validation Test | FLIC | |
|---|---|---|
| | Winnipeg | Edmonton |
| Katz activities of daily living | −0.170 | −0.305 |
| General Health Questionnaire | | |
| A scale: somatic symptoms | −0.629 | −0.668 |
| B scale: anxiety and insomnia | −0.441 | −0.496 |
| C scale: social dysfunction | −0.683 | −0.581 |
| D scale: severe depression | −0.467 | −0.595 |
| Total | −0.724 | −0.765 |
| Beck depression | −0.724 | −0.773 |
| Karnofsky | 0.693 | 0.619 |
| Spielberger | | |
| State anxiety | −0.557 | −0.578 |
| Trait anxiety | −0.551 | −0.599 |
| Melzack | | |
| Present pain index | −0.586 | ... |
| Pain rating index | −0.548 | ... |

NOTE. Pearson (r) correlation coefficients for the overall FLIC score against the concurrent validation battery. The only weak correlation is with the Katz ADL index (see text).

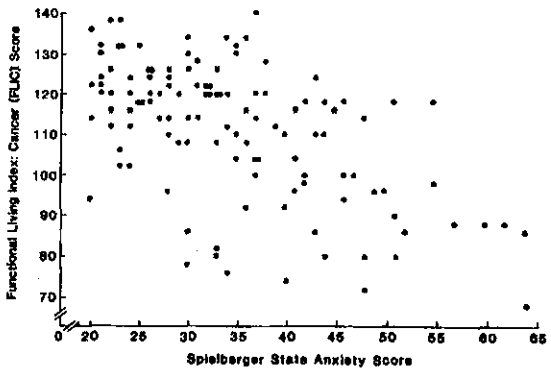



**Fig. 3 Scattergram of Spielberger versus FLIC. The correlation of the FLIC score with measures of psychosocial function such as the Spielberger is good (Pearson r = −0.735, Winnipeg run 2). The inverse correlation occurs since higher FLIC scores represent better function, while increasing anxiety results in a lower Spielberger score.**

als involving two populations likewise signifies a strong measure of construct validity, in addition to lending credence to the belief that our four component factors do represent a meaningful approximation to functional quality of life.

The physical function and ability factor of the FLIC correlates well with those concurrent validation studies measuring physical attributes or their consequences, but does not correlate strongly with the psychosocial measures. Analogously, FLIC's emotional function factor correlates strongly with measures of depression and anxiety, but weakly with physical ability measures. These data provide clear evidence that the

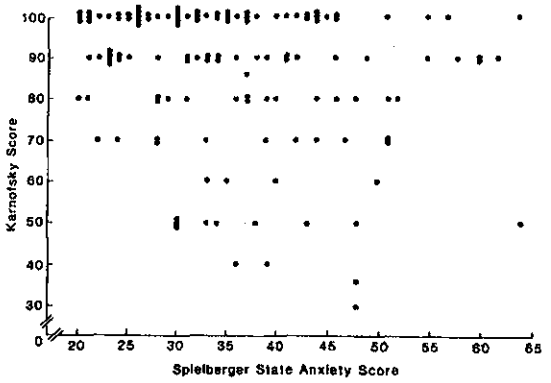
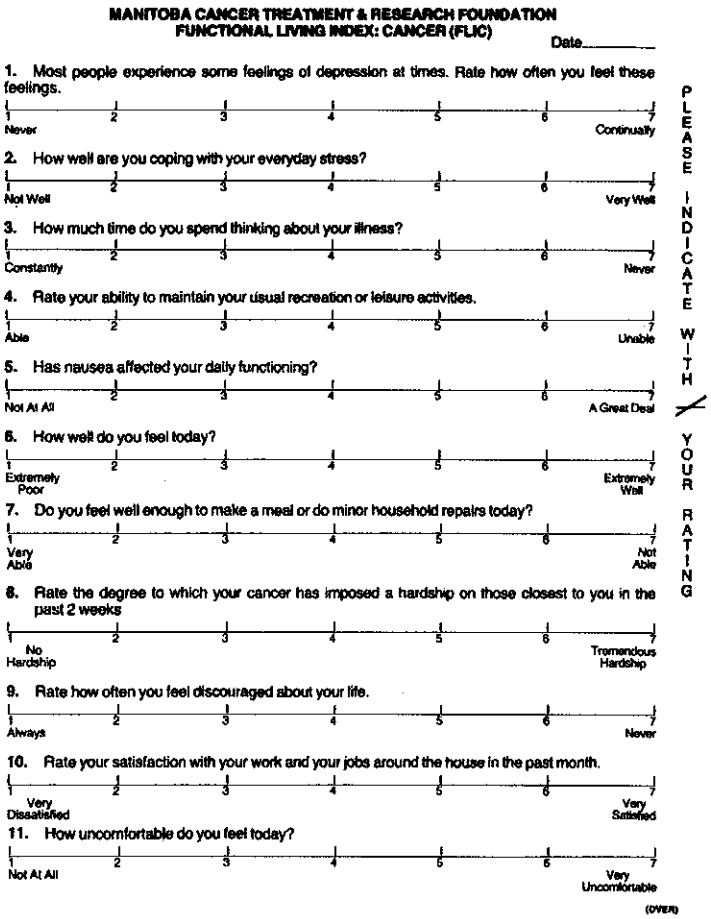



**Fig. 4. Scattergram of Karnofsky versus Spielberger. This plot demonstrates the independence of the traditional Karnofsky score, from a representative measure of psychosocial function (Pearson r = −0.194, Winnipeg run 2).**


MANITOBA CANCER TREATMENT & RESEARCH FOUNDATION
FUNCTIONAL LIVING INDEX: CANCER (FLIC)
Date_____

PLEASE INDICATE WITH ✓ YOUR RATING

1. Most people experience some feelings of depression at times. Rate how often you feel these feelings.
1 (Never) 2 3 4 5 6 7 (Continually)

2. How well are you coping with your everyday stress?
1 (Not Well) 2 3 4 5 6 7 (Very Well)

3. How much time do you spend thinking about your illness?
1 (Constantly) 2 3 4 5 6 7 (Never)

4. Rate your ability to maintain your usual recreation or leisure activities.
1 (Able) 2 3 4 5 6 7 (Unable)

5. Has nausea affected your daily functioning?
1 (Not At All) 2 3 4 5 6 7 (A Great Deal)

6. How well do you feel today?
1 (Extremely Poor) 2 3 4 5 6 7 (Extremely Well)

7. Do you feel well enough to make a meal or do minor household repairs today?
1 (Very Able) 2 3 4 5 6 7 (Not Able)

8. Rate the degree to which your cancer has imposed a hardship on those closest to you in the past 2 weeks
1 (No Hardship) 2 3 4 5 6 7 (Tremendous Hardship)

9. Rate how often you feel discouraged about your life.
1 (Always) 2 3 4 5 6 7 (Never)

10. Rate your satisfaction with your work and your jobs around the house in the past month.
1 (Very Dissatisfied) 2 3 4 5 6 7 (Very Satisfied)

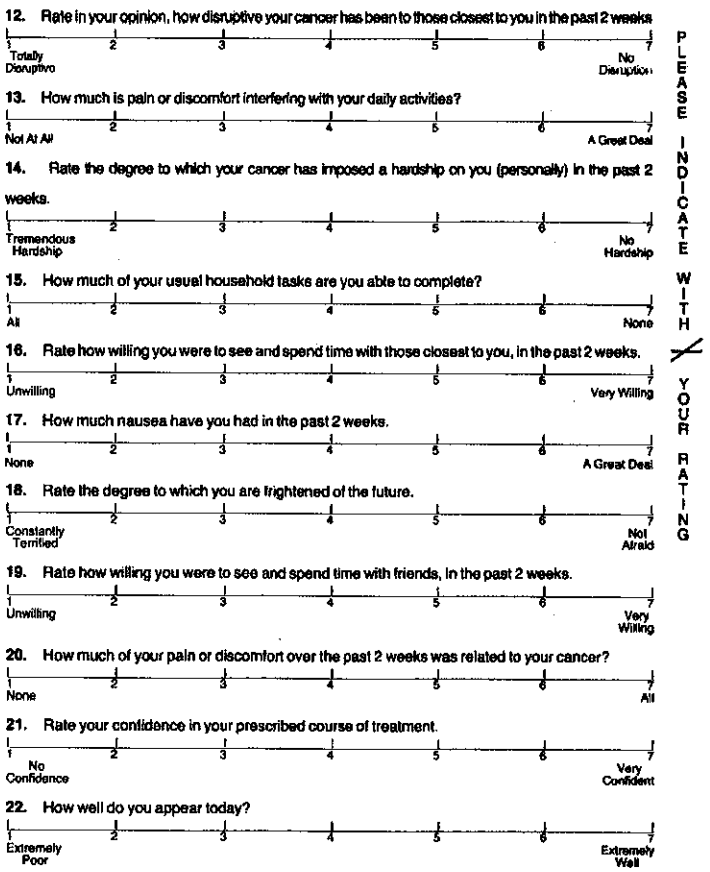
11. How uncomfortable do you feel today?
1 (Not At All) 2 3 4 5 6 7 (Very Uncomfortable)

(over)

**Fig. 5. Sample of the Manitoba Cancer Treatment and Research Foundation Functional Living Index: Cancer questionnaire.**

**Table 7. Correlations of Factor Scores with Concurrent Validation Tests: Comparison of Winnipeg (Run 2) and Edmonton Data**

| Validation Test | Physical Well-Being and and Ability Factor of FLIC | | Emotional State Factor of FLIC | |
|---|---|---|---|---|
| | Winnipeg (Run 3) | Edmonton | Winnipeg (Run 3) | Edmonton |
| Katz activities of daily living | −0.228 | −0.207 | 0.042 | 0.20 |
| General Health Questionnaire | | | | |
| A scale: somatic symptoms | −0.541 | −0.606 | −0.315 | −0.346 |
| B scale: anxiety and insomnia | −0.204 | −0.224 | −0.557 | −0.543 |
| C scale: social dysfunction | −0.649 | −0.571 | −0.279 | −0.083 |
| D scale: severe depression | −0.263 | −0.328 | −0.556 | −0.427 |
| Total | −0.549 | −0.572 | −0.594 | −0.469 |
| Beck depression | −0.567 | −0.640 | −0.517 | −0.528 |
| Karnofsky | 0.757 | 0.554 | 0.066 | 0.157 |
| Spielberger | | | | |
| State anxiety | −0.308 | −0.326 | −0.629 | −0.588 |
| Trait anxiety | −0.296 | −0.338 | −0.735 | −0.609 |
| Melzack | | | | |
| Present pain index | −0.657 | ... | −0.067 | ... |
| Pain rating index | −0.555 | ... | −0.128 | ... |




FLIC measures a composite of distinct factors contributing to overall functional living. If these distinctions in correlation data had not emerged it could have been argued that the FLIC was little more than an elaborate way of asking "How do you feel?"

In addition to providing substantial evidence for concurrent validity, the trials in Winnipeg and Edmonton in which the FLIC was compared with a battery of established measures of physical and psychologic function lead to some significant observations. Possibly most important, physician-based Karnofsky scores, which like their Eastern Cooperative Oncology Group (Zubrod) counterparts form the basis of functional status assessment in the clinical trial setting, do not correlate at all well with measures of psychologic well being, sociability, or even somatic discomfort, and form only one component of a functional life style–based response assessment. As Yates et al have shown,[20] the Karnofsky Scale has reasonable correlation with the overall medical state of illness, but we believe that broader issues of patient function must be addressed in assessing and designing clinical trials. It is entirely possible for medical parameters in a particular patient to indicate a most successful treatment while in fact the patient has been reduced to total societal debility by virtue of psychologic, somatic, or social interactional factors that may be the result of either the disease or its treatment. These issues appear not to enter into the estimation of functional status using a Karnofsky Index. If that is the case then a broader-based index may provide answers to the apparent compliance difficulties, and response-function discrepancies that are increasingly seen within clinical trials. We consider the social desirability data particularly relevant. Subjects frequently attempt to provide answers that will please their physicians, hence masking their true perceptions. Frequently this is not considered in the design of such docu-

ments. The Jackson measure is designed to determine the sensitivity of specific questionnaire items to this social desirability contamination. The observation that our questionnaire is uncontaminated by such factors adds to the validity of the index as established by the other test measures.

Why did the Katz Activities of Daily Living Index not correlate very well with our index? Case-by-case analysis of our data suggests that this apparent noncorrelation may be a function of patient selection. The Activities of Daily Living Index is designed primarily as a measure of patient self-help function for those institutionalized or being considered for immediate institutional placement. As such, it represents a population considerably more disabled than many patients for whom our questionnaire is designed, which may explain the observation that most patients in our study scored at the ceiling of the Katz Index. This restriction of scoring range accounts for the low correlation between our index and the Katz. There does in fact seem to be a reasonable correlation between the Katz and the FLIC when patients scoring at the lower levels are considered.

The stratification data shows that this index is valid across the general range of extent of illness seen in cancer patients. It seems to differentiate the overall levels of dysfunction in patients who are disease free and off treatment, in those undergoing adjuvant therapy, in those in the midst of treatment for active disease, and in those in the palliative-care setting.

We have not encountered difficulties administering the test. The 20- and 26-item questionnaires were answered easily by patients in $<10$ minutes. All patients were able to answer all of the questions. The entire concurrent validation package took approximately half an hour to complete and it serves as a measure of the acceptability of our index that while considerable difficulty was encountered with the McGill/Melzack Pain Inventory, and some difficulty was encountered with both the Beck and Spielberger questionnaires, particularly when the issue of suicide was raised, no patient found difficulty with the FLIC.

Initially this questionnaire was designed to compare populations. Our primary intention was not to scale it, but to use it as a overall global measure of patient function. We believe it is valid in that context. However, the clarity with which factor 1 representing physical ability and factor 2 representing psychosocial function seem to represent distinct and important dimensions suggests that at least for these two items the index is in fact scalable and may possibly be usable not only for population studies, but also as a guide to intervention in particular patients. Predictive validity studies are underway and in the event that responses to this index herald subsequent changes in patients' overall status leading to appropriate interventions, then its value will be augmented. At present, we are recommending the use of this index to provide adjunctive information in the interpretation of comparative clinical trials. In this setting a variety of outcomes are possible. These may both provide explanations for apparent discrepancies between patient function and clinical response, and pose some new and potentially difficult, but beneficial questions for clinical-trials investigators. For example, if based on traditional criteria two clinical trials provide similar outcomes, a significant discrepancy between quality of life outcomes would readily influence subsequent treatment selection. More perplexing however, is the potential dilemma of divergent results, that is, that an outcome that is viewed inferior by classical criteria is associated with a significantly improved overall quality of life response. Which treatment arm does the investigator then recommend, the one with the greater response and/or survival, or the one with a quality of life outcome that appears superior? We cannot of course propose a universal answer, but the availability of a valid measure of quality of life provides alternatives: (*1*) The investigator will be aware of the quality of life distinctions and may be in a position to alter therapy in the "traditionally" superior arm to improve the quality of life outcome. (*2*) If the former is not possible, at least we are in the position to offer patients not only response, disease-free survival, and overall survival and medical toxicity information, but also, in a reasonably concrete form, quality of life information. In such circumstances the patient will be better able to participate in making an informed decision. Particularly in diseases such as non-small cell lung cancer where the entire question of chemotherapy efficacy is unresolved and where the number of confounding factors is acknowledged to be large, it is important to provide

information to patients regarding quality of life aspects of proposed treatments. This index represents an attempt to provide such information.

## ACKNOWLEDGMENTS

We thank Loretta Hallgren, W. W. Cross Cancer Institution, Edmonton; Professor H. A. Minden, York University, Toronto; and Sharon Reisdorf.

## REFERENCES

1. Simon RM: Design and conduct of clinical trials in cancer, in DeVita V, Hellman S, Rosenberg SA (eds): Cancer: Principles and Practice of Oncology. Philadelphia, Lippincott, 1982, p 198–225

2. Holland JCC: The humanistic side of cancer care: Changing issues and values. Proceedings of the American Cancer Society, 1981, p 1–13

3. Meyerowitz BE: Psychosocial correlates of breast cancer and its treatments. Psycho Bull 87:108–131, 1980

4. Gorzynski G, Massie MJ, Holland J: Diagnosis and management of psychiatric syndromes in cancer patients, in Derogatis L (ed): Psychopharmacology in Clinical Practice. New York, Addison-Wesley Press, 1981

5. Hersh SP: Psychologic aspects of patients with cancer, in DeVita V, Hellman S, Rosenberg SA (eds): Cancer: Principles and Practice of Oncology. Philadelphia, Lippincott, 1982, p 267–277

6. Schain WS: Sexual problems of patients with cancer, in DeVita V, Hellman S, Rosenberg SA (eds): Cancer: Principles and Practice of Oncology. Philadelphia, Lippincott, 1982, p 278–290

7. Spitzer WO, Dobson AJ, Hall J, et al: Measuring the quality of life of cancer patients. A concise QL-index for use by physicians. J Chronic Dis 34:585–598, 1981

8. Padilla G, Presant CA, Grant M, et al: Assessment of quality of life in cancer patients. Proc Am Soc Clin Oncol 22:397, 1981 (Abstr)

9. Selby PJ, Boyd NF, Campbell J, et al: A measurement instrument of quality of life (QL) in patients with breast and lung cancer. Proc Am Soc Clin Oncol 1:45, 1982 (Abstr)

10. Katz S, Akpom CA: A measure of primary sociobiological functions. Int J Health Serv 6:493–508, 1976

11. Karnofsky DA, Burchenal JH: The clinical evaluation of chemotherapeutic agents in cancer, in Macleod CM (ed): Evaluation of Chemotherapeutic Agents. New York, Columbia University Press, 1949, p 191–205

12. Melzack R: The McGill Pain Questionnaire: Major properties and scoring methods. Pain 1:277–299, 1975

13. Speilberger CD, Gorsuch RL, Lushene RE: STAI Manual for the State-Trait Anxiety Inventory. Palo Alto, Calif, Consulting Psychologists Press, 1970

14. Beck AT, Ward CH, Mendolson M, et al: An inventory for measuring depression. Arch Gen Psychiatry 4:561–571, 1961

15. Goldberg DP, Hellier VF: A scaled version of the General Health Questionnaire. Psychol Med 9:139–145, 1979

16. Jackson DN: Structured personality assessment (SPA), in Wolman B (ed): Handbook of Psychology. Englewood Cliffs, NJ, Prentice Hall, 1974

17. Jackson DN: Dynamics of structured personality tests. Psychol Rev 78:229–248, 1971

18. Jackson DN: Sequential system for personality scope development, in Spielberger CA (ed): Current Topics in Clinical and Community Psychology, Vol. II. New York, Academic Press, 1970

19. Nie NH, Hull CH: Statistical Package for the Social Sciences. New York, McGraw Hill, 1975

20. Yates JW, Chalmer B, McKegney FP: Evaluation of patients with advanced cancer using the Karnofsky Performance Status. Cancer 45:2220–2224, 1980

VOLUME 31 · NUMBER 29 · OCTOBER 10 2013

JOURNAL OF CLINICAL ONCOLOGY

SPECIAL ARTICLE

# End Points and Trial Design in Geriatric Oncology Research: A Joint European Organisation for Research and Treatment of Cancer–Alliance for Clinical Trials in Oncology–International Society of Geriatric Oncology Position Article

*Hans Wildiers, Murielle Mauer, Athanasios Pallis, Arti Hurria, Supriya G. Mohile, Andrea Luciani, Giuseppe Curigliano, Martine Extermann, Stuart M. Lichtman, Karla Ballman, Harvey Jay Cohen, Hyman Muss, and Ulrich Wedding*

Hans Wildiers, Murielle Mauer, Athanasios Pallis, Andrea Luciani, Giuseppe Curigliano, Martine Extermann, and Ulrich Wedding, European Organisation for Research and Treatment of Cancer, Brussels; Hans Wildiers, University Hospitals Leuven and Katholieke Universiteit Leuven, Leuven, Belgium; Hans Wildiers, Arti Hurria, Harvey Jay Cohen, and Ulrich Wedding, International Society of Geriatric Oncology, Geneva, Switzerland; Arti Hurria, City of Hope, Duarte, CA; Arti Hurria, Karla Ballman, Harvey Jay Cohen, and Hyman Muss, Alliance, Chicago, IL; Supriya G. Mohile, University of Rochester, Rochester; Stuart M. Lichtman, Memorial Sloan-Kettering Cancer Center and Weill Cornell Medical College, New York, NY; Andrea Luciani, S. Paolo Hospital; Giuseppe Curigliano, European Institute of Oncology, Milano, Italy; Martine Extermann, University of South Florida, Tampa, FL; Karla Ballman, Mayo Clinic, Rochester MN; Harvey Jay Cohen, Duke University, Durham; Hyman Muss, University of North Carolina, Chapel Hill, NC; and Ulrich Wedding, Jena University Hospital, Jena, Germany.

Published online ahead of print at www.jco.org on September 9, 2013.

Supported by a donation from the Fondation Contre le Cancer in Belgium through the European Organisation for Research and Treatment of Cancer Charitable Trust. H.W. is senior clinical investigator of the Fund for Scientific Research (FWO-Vlaanderen).

This article is endorsed by European Organisation for Research and Treatment of Cancer, Alliance for Clinical Trials in Oncology, and International Society of Geriatric Oncology.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Hans Wildiers, MD, PhD, University Hospitals Leuven, Herestraat 49, Leuven, Belgium, B-3000; e-mail: hans.wildiers@uzleuven.be.



0732-183X/13/3129w-3711w/$20.00

DOI: 10.1200/JCO.2013.49.6125

## ABSTRACT

Selecting the most appropriate end points for clinical trials is important to assess the value of new treatment strategies. Well-established end points for clinical research exist in oncology but may not be as relevant to the older cancer population because of competing risks of death and potentially increased impact of therapy on global functioning and quality of life. This article discusses specific clinical end points and their advantages and disadvantages for older individuals.

Randomized or single-arm phase II trials can provide insight into the range of efficacy and toxicity in older populations but ideally need to be confirmed in phase III trials, which are unfortunately often hindered by the severe heterogeneity of the older cancer population, difficulties with selection bias depending on inclusion criteria, physician perception, and barriers in willingness to participate. All clinical trials in oncology should be without an upper age limit to allow entry of eligible older adults. In settings where so-called standard therapy is not feasible, specific trials for older patients with cancer might be required, integrating meaningful measures of outcome. Not all questions can be answered in randomized clinical trials, and large observational cohort studies or registries within the community setting should be established (preferably in parallel to randomized trials) so that treatment patterns across different settings can be compared with impact on outcome. Obligatory integration of a comparable form of geriatric assessment is recommended in future studies, and regulatory organizations such as the European Medicines Agency and US Food and Drug Administration should require adequate collection of data on efficacy and toxicity of new drugs in fit and frail elderly subpopulations.

*J Clin Oncol 31:3711-3718. © 2013 by American Society of Clinical Oncology*

## INTRODUCTION

The choice of appropriate end points is important to assess the benefit of therapy. In oncology, there are well-established clinical end points for clinical research in randomized clinical trials (RCTs); in the curative/adjuvant setting, disease-free survival (DFS) and overall survival (OS) are the most recognized and well accepted. For metastatic solid tumors, progression-free survival (PFS), time to tumor progression (TTP), time to treatment failure (TTF), response rate (RR), and OS are the most commonly used end points.

A caveat is that the definitions of these so-called standard outcomes have varied in different trials in the past, challenging the ability to compare across studies and provide evidence-based care. There are international efforts to streamline this, such as the DATECAN (Definition for the Assessment of Time-to-Event End Points in Cancer Trials) project.[1]

However, these standard end points may not be the most appropriate to balance the benefits with the risks of therapy in older patients with cancer, because older patients often die as a result of other diseases, and relapse will not always affect survival, whereas cancer-directed therapy can sometimes cause severe acute or chronic toxicities and decreased quality of life (QoL). For young patients with familial/social obligations (eg, toward young children), prolongation of life might be the most important end point; however, older adult patients with incurable disease may prefer QoL above quantity of life, especially if treatment also has an impact on their functional capacity and ability to carry out

daily tasks, their cognitive function, their social situation/capability to stay at home, or their caregiving abilities.[2] Therefore, there is a need for delineation of relevant clinical end points for older individuals, which can then be uniformly incorporated into future clinical trials.[3,4]

The best-established form of clinical trial design is the RCT. When designing RCTs for older patients with cancer, selection of what should be the standard arm may vary because this can be different for fit, vulnerable, and frail patients. As a result, it will often not be possible to have the same standard arm for all older patients, so other trial designs should be considered, especially for vulnerable and frail patients.

This article describes several potential outcome measures/end points and their advantages and disadvantages for elderly-specific clinical trials and discusses potential trial designs that could be used to greatly expand evidence-based treatment outcomes for the older population with cancer.

## OUTCOME MEASURES/END POINTS FOR CLINICAL TRIALS IN OLDER INDIVIDUALS

### OS

OS is considered the gold standard in clinical trials, especially when evaluating the superiority of new treatments; other end points such as PFS and DFS are commonly used to report on clinical benefit, but this has been subject to criticism (Table 1).[5] Surrogacy of these end points for OS has been demonstrated in some specific settings and is under investigation in others. Compared with younger patients, elderly patients with cancer often present with significant comorbidities and therefore die as a result of other, non–cancer-related diseases more frequently.[6,7] Elderly patients are more likely to experience severe toxicities from cancer-directed therapies, including treatment-related mortality.[8,9] Non–disease-related deaths and treatment discontinuation/reduced dosage because of toxicity might dilute treatment benefit, and larger sample sizes would be needed to demonstrate treatment effects. It should be emphasized that this diluted benefit is an accurate estimate of the true clinical benefit in the older population, and larger sample sizes are the price society has to pay if it wants to ensure that older patients are not subjected to toxic therapies that provide no tangible clinical benefit. The mentioned concerns have resulted in age limits and stringent inclusion criteria, leading to the exclusion of large numbers of older patients from clinical trials.[3,10,11] Although excluding older patients with comorbidities could help a trial determine whether a benefit from treatment exists (especially if the benefit is small), this approach limits generalizability of the treatment for the vast majority of cancers, where most of the patients are older. On average, the trial population in chemotherapy trials is 5 to 10 years younger than the general population with the disease. Because there are no regulatory requirements for establishing the efficacy or toxicity of new therapies in older adults, the limited data in this population ultimately lead to the risk of expensive treatments being used in the older, less studied population, resulting in higher toxicity and smaller benefit than in younger patients with cancer.

### Disease-Specific Survival

Whereas primary end points such as OS or PFS would still be suitable to provide a realistic estimation of treatment benefit in the targeted population in the presence of competing risks, measuring cancer-specific end points such as disease-specific survival (DSS) and performing competing risks analyses could generate crucial data. Nout et al[12] nicely demonstrated that including or excluding non–breast cancer–related deaths and contralateral breast cancer significantly affected outcome reporting in early breast cancer. DSS better indicates how many patients die as a result of disease and how many die as a result of other causes. A precondition to using DSS as the primary end point is that the cause of death can be reliably ascertained, and other causes of death are not related to the treatment. In that case, DSS as the primary end point might help in requiring a smaller sample size.[13] However, a reduction in the risk of one type of event (eg, death resulting from cancer) can lead to an increase in the number of observed events for competing types, just because patients remain at risk for those events for a longer period. At any rate, information on cause of death should always be reported to distinguish cancer deaths from treatment-related deaths and deaths resulting from other causes. We recommend reporting DSS always in addition to OS.

### Coprimary End Points

Coprimary end points should also be considered because this allows capturing more than efficacy alone. Multiple single end points can be chosen as coprimary end points of equal importance, and a statistical design can be built to test each separately. However, coprimary end points also have disadvantages; statistical design is difficult because the correlation between the different end points is rarely known. Moreover, if the trial objective is to have positive results for at least one or all coprimary end points, the type I or II error, respectively, must be adjusted for multiple testing, which necessitates in increase of sample size.[14]

### Composite End Points

Composite end points are another way of integrating other aspects into the end point, such as QoL, treatment effects on disease-related symptoms, functional capacity, and ability to carry out daily tasks. As the International Conference on Harmonisation stated,[15] composite end points avoid the need for arbitrary choice and deal with multiplicity in an efficient manner when several outcome measures are of equal importance to the patient. A composite end point in an RCT consists of multiple single end points that are combined so that an event is indicated if any of the end points occurs. Composite end points have sometimes been used in oncology (eg, skeletal-related events in clinical trials with bisphosphonates or denosumab[16]) but have been more widely used and studied in other medical disciplines, mainly in cardiology.[17,18] Major advantages of a composite end point are the simplicity of the statistical design, which is based on a single end point (ie, the composite one), and the resultant increase in statistical efficiency. However, there are also risks, and caution must be applied. The major possible issues include: lack of a strong rationale given for the composite (ie, mixture of end points with different clinical importance; eg, death and hospital admission), difficulty in interpretation of the results in case of positive results on the composite but observed divergent effects on the components, and inadequate or incorrect reporting of the results (eg, declaring positive effects on the most important component when statistical significance is only reached for the composite, and when the more important component, such as death, accounts only for a minority of the events). Less frequent but important to consider is the situation in which negative results can be observed for the composite, while

**Table 1.** Relevant End Points in Clinical Trials in the Older Cancer Population

| End Point | Definition | Current Situation | Pro | Con |
|---|---|---|---|---|
| OS: time or proportion | Time from diagnosis of treatment situation/study entry until death or rate of patients alive at specified time point | Considered gold standard in clinical trials, especially when evaluating superiority of new treatments | Remains hardest end point, also in elderly | Oncologic relevance in elderly can be hampered by increased number of non–cancer-related deaths (all life ends with death) |
| | | | Easy and distinct to measure, high impact for patients | Does not include QoL aspects |
| DSS: time or proportion | Time from diagnosis of treatment situation/study entry until death resulting from index disease or rate of patients without death related to index disease at specified time point | Important to collect in addition to OS because it gives better insight into contribution of non–cancer-related deaths | Cancer treatment primarily aims at decreasing cancer death | Some cancer treatments might also influence non–cancer-related deaths (eg, treatment-related mortality)<br>May lead to overestimation of true benefit for patients in presence of competing risks (eg, treatment benefit in localized prostate cancer)<br>Reason for death will be of no/minor meaning for patients<br>Reason for death can remain unclear |
| Coprimary end points | Combination of ≥ two equal primary end points | Rarely used in oncology | Allows capturing more than efficacy alone | Difficult statistical design because correlation between different end points is rarely known |
| | | | | Might increase sample size |
| Composite end points | Combination of different end points in one defined end point | Rarely used in oncology (one example: skeletal-related events) but should be encouraged more | Can take into account multiple dimensions in definition of treatment benefit, including efficacy and toxicity | Requires individual components of composite that are clinically meaningful and of similar relative importance |
| | | | Simple and efficient statistical design<br>Allows separate reporting of different end points | Difficult interpretation if there are divergent results for each component separately |
| TFFS and TTF: time or proportion | TFFS is time elapsing between random assignment and early treatment discontinuation because of any reason (including disease progression, treatment toxicity, early death), disease progression, death (resulting from any cause), or any other event of interest; TTF is similar, but death resulting from other cause is not considered an event | Often used in addition to OS | Integrates efficacy and toxicity | Difficult to distinguish between efficacy and toxicity (eg, toxic but effective)<br>Treatments might be stopped for other reasons (eg, chemotherapy holiday) |
| QoL-related end points: level at specified time point or time until deterioration compared with baseline | Evaluation of QoL through validated instruments at baseline and during course of disease/treatment/study | Often used as secondary end point in clinical trials but should be promoted as primary end point or part of composite end point | QoL may be more important than duration of life for many older individuals | Difficult to measure and identify clinically relevant cutoffs that determine whether therapy is worthwhile |
| Maintenance of functional capacity/dependence: level at specified time point or time until deterioration compared with baseline | Evaluation of evolution of functioning and (in)dependence through validated instruments during course of disease/treatment/study | Rarely measured in oncology trials but crucial to include | Main contributor to QoL in elderly patients with cancer | No general consensus on optimal measurement or clinically relevant cutoffs determining whether therapy is worthwhile |

Abbreviations: DSS, disease-specific survival; OS, overall survival; QoL, quality of life; TFFS, treatment failure–free survival; TTF, time to treatment failure.

statistical significance can be reached for the most important component. The pros and cons of composite end points have been summarized by Kleist.[19] Use of this approach is usually justified under the following assumptions:

- The individual components of the composite are clinically meaningful and of similar relative importance to clinical care.
- The expected effects on each component are similar based on clinical/biologic plausibility (which is, in the end, the rationale for using a composite end point).
- For the study to be ultimately positive, the clinically more important components of a composite end point should at least not be affected negatively.

All components of a composite end point should also be analyzed separately and reported as such. The separate reporting of end points is also essential to facilitate cross-study comparisons (although there are also intrinsic limitations to this) or to generate assumptions for designing future trials. It is important to mention that for the US Food and Drug Administration, a regulatory end point should clearly distinguish the efficacy of the drug from toxicity, patient or physician withdrawal, or patient intolerance.[20]

An interesting example of a composite end point in older individuals is therapeutic success.[21] This end point combines efficacy, toxicity, and patient compliance with treatment and has been defined as a patient receiving at least three cycles of chemotherapy, at the planned dose (without dose reduction) and schedule (no treatment delay beyond 2 weeks), and having a response (either complete or partial) without experiencing grade 3 or 4 toxicity according to the Common Toxicity Criteria criteria.[22] Variations of this design are possible, such as defining therapeutic success as being progression free at a fixed time point without having grade 3 or 4 nonhematologic or grade 4 hematologic toxicity. This seems to be an attractive end point in settings where significant differences in toxicity between two treatments are expected and requires further exploration. Looking simultaneously at toxicity and efficacy can be a disadvantage as well as an advantage; therapies might be temporarily toxic, requiring dose reduction, but might be efficacious. Dose, toxicity, and response are related (eg, in patients with non–small-cell lung cancer, those with a higher rate of hematologic toxicity survive longer[23]).

Another example is the use of overall treatment utility (OTU) as an end point in the FOCUS (Fluorouracil, Oxaliplatin, and CPT11 [irinotecan]—Use and Sequencing) trial of older patients with metastatic colorectal cancer,[24] in which good OTU indicated no clinical or radiologic evidence of disease progression and no major negative treatment effects in terms of toxicity or patient acceptability. Intermediate OTU signified either clinical deterioration but no negative treatment effect or a significant negative treatment effect but no clinical deterioration. Poor OTU indicated both clinical deterioration and a major negative treatment effect or death.

### Treatment Failure–Free Survival and TTF

Treatment failure–free survival (TFFS) and TTF are well-known examples of composite end points and could also be interesting end points to consider for clinical trials in the elderly. TFFS is defined as the time that elapses between random assignment and early treatment discontinuation because of any reason (including treatment toxicity and patient refusal of further treatment), disease progression, death resulting from any cause, or any other event of interest. TTF is similar, but only disease-specific and treatment-related deaths are considered events. Treatment-related toxicity is a major issue in elderly patients with cancer, especially those with advanced disease stages where the goal of treatment is palliation rather than cure. TFFS and TTF provide an opportunity to take into account the role of toxicity and not concentrate only on efficacy. This is important because older patients are less willing than younger patients to continue treatments with severe toxicities,[2,25] especially if these have functional consequences that limit independence. One limitation, however, is that in some situations, treatment breaks are introduced not because of toxicity or progression but to provide a period without chemotherapy (ie, chemotherapy holiday), although this can be handled by not considering these breaks as treatment failures. Another limitation is that early treatment discontinuations are still considered failures in situations where significant toxicity occurs, but patients have good disease outcomes (perhaps with improvement of toxicities) thereafter.

### QoL-Related End Points

The main goal of cancer treatment, certainly in the palliative setting, should be to reduce discomfort related to or caused by cancer progression and its related consequences (eg, loss of functionality, inability to stay at home, deterioration of QoL). Health-related QoL (HRQoL) is a major concern for patients with cancer, and it can be affected by symptoms caused by cancer as well as by treatment-induced toxicity.[26] For many older patients, the goal of cancer-directed treatment is not just how much additional time they can gain but how valuable that time is. Elderly patients are less willing to compromise their HRQoL for the potential for increased survival.[27] Thus, HRQoL may be an appropriate outcome for elderly-specific trials, but it remains to be defined how to measure or quantify HRQoL optimally, how to quantify the different domains of HRQoL in one score, and which cutoffs are relevant as end points for clinical trials, although a 10-point decrease (on score of 100) is frequently used as relevant change.[28] The EORTC (European Organisation for Research and Treatment of Cancer) QoL Group recently developed an elderly-specific QoL module,[29] which adds specific QoL-related aspects in older individuals to the general EORTC Quality of Life Questionnaire C30. HRQoL should be captured in all trials of palliative chemotherapy in older patients regardless of the primary end point of the trial. The Q-TWIST (quality-adjusted time without symptoms of disease or toxicity of treatment) approach measuring quality-adjusted survival is another QoL-related end point, which partitions the survival time of the patient into three consecutive health states (ie, time with toxicity resulting from treatment, time without symptoms of disease or toxicity, and time from progression/relapse to death) and assigns utility weights to each state.[30] The Q-TWIST value is the sum of the weighted health state durations and is used for treatment comparisons. This approach quantitatively adjusts periods in which treatment toxicities or symptoms of disease progression are present to reflect the potentially reduced value for the patient. In principle, this is a valuable approach for older patients with cancer, but the great difficulty lies in determining or quantifying the weight factor for QoL during the different periods.

### Preservation of Functional Capacity/Independence

In a similar way, maintenance of function and independence should be one of the major principles of cancer management in the elderly. A negative impact on a patient's functional capacity will have a negative impact on survival as well.[31] The prolongation of active life expectancy seems much more important than the prolongation of life expectancy as such. The GERICO (French Geriatric Oncology Group) trial[32] nicely showed that functionality measured by instrumental activities of daily living does not decrease significantly (by $\geq$ two points) in older patients with breast cancer receiving adjuvant chemotherapy. Using single or multiple domains of geriatric assessment as outcome events would also be of great value to clinicians.

### Surgical Trial End Points

Several trials in the surgical field, including elderly-specific trials such as the PACE (Pre-operative Assessment of Cancer in the Elderly)

study,[33] have used (primary) end points such as 30-day morbidity, 30-day serious morbidity (grade 3 to 4), and 30-day mortality, which are relevant but should be accompanied by information on longer-term outcome end points, as we have discussed here.

## TRIAL DESIGN IN OLDER PATIENTS WITH CANCER

### *Trials for Older Patients Versus Trials Without Upper Age Limit*

Table 2 lists issues in clinical trial design in older patients with cancer. Clinical trials need to be representative of the whole population in whom the treatment will be used later, which is not the case at present. Several studies have shown that there is substantial underrepresentation of older patients in clinical trials.[10,34,35] The differential effects of aging on organ function and the variety of comorbidities that characterize the older population result in significant heterogeneity.[36] This variance could result in considerable differences in the efficacy and safety of cancer treatments. For studies using therapy regimens expected to be used in all age categories, patients should be enrolled across the entire age spectrum, and a minimum cohort of elderly patients should be required. If treatment regimens are expected to be tolerated by only fit older patients or younger patients, severe selection bias will be present, and conclusions from these kinds of trials will not be generalizable to the whole population, especially the frail elderly. It is important to capture the fitness status of the older patients enrolled onto a clinical trial to provide information about the generalizability of the results. Documentation of the nonincluded population is also important. One option for ensuring sufficient accrual of older patients could be to require registration trials to remain open after they have met their target accrual until a minimum cohort of elderly patients is enrolled. It should be noted that older fit patients are likely included in clinical trials and so should likely receive the standard treatments. However, it is clear that several standard treatments administered to younger patients are not suitable for unfit or frail elderly adults (and sometimes even fit elderly adults) because of expected higher or unacceptable risk of toxicity or other competitive risks determining the long-term prognosis. For example, allogenic bone marrow transplantation; high-dose cytarabine, anthracycline, or cisplatin; major surgery; and concurrent chemoradiotherapy are treatments generally reserved for younger or sometimes fit older patients. In this setting, elderly-specific trials are certainly needed, because there is no clear standard therapy in this group of patients, who are not likely to tolerate the standard therapy administered to fit patients. In frail older patients, separate clinical trials could be designed because these patients could be better served by trials comparing modified approaches (eg, adapted chemotherapy/biologic agents) with pure palliative/supportive care. For vulnerable patients, a possible trial design could include standard therapies versus less aggressive therapies or no therapy, depending on the setting.

**Table 2.** Issues in Clinical Trial Design for Older Patients With Cancer

| Issue |
|---|
| RCTs remain gold standard when possible |
| Clinical trials should preferably integrate whole age range, including fit and frail older individuals |
| Elderly-specific clinical trials in older patients with cancer are required if standard therapy is different from that for younger patients |
| Trials of treatment strategy comparing different strategies (eg, therapy *v* best supportive care) should be encouraged |
| Randomized phase II or even single-arm phase II trials in specific subsets of older patients can provide insight into range of efficacy and toxicity in older populations but ideally should be confirmed in large phase III trials, which might be hard to perform for various reasons (eg, insufficient interest from sponsors/investors, difficulty in finding sufficient numbers of patients) |
| Not all questions can be answered with randomized trials, and large observational cohort studies or registries in community can provide further insight for frail population with less selection bias (preferably in parallel with or linked to RCTs) |
| Comparable/uniform geriatric assessment should be integrated into future trials in geriatric oncology |
| Regulatory authorities should require evaluation of efficacy and safety of new drugs in older and frail patients as well as in younger patients |

Abbreviation: RCT, randomized clinical trial.

Randomized phase III trials remain the gold standard for clinical research, in older as well as younger people. However, designing these trials that address heterogeneity in all elderly populations might be challenging for many reasons (insufficient interest from sponsors/investors, difficulty in finding sufficient numbers of patients, and so on). Often, phase III data exist only for younger populations. Randomized phase II trials can provide insight into the range of efficacy and toxicity in older populations. If the treatment is too toxic, this would be established in a phase II trial. If a phase II trial in an older (nonfit) population shows that the toxicity is acceptable and confirms efficacy in the same range as previous phase III trials in younger people, there might not be a need to repeat the phase III trial again in an older (nonfit) population. However, if the phase II results are indeterminate concerning toxicity and/or efficacy, then confirmation in a phase III trial is likely. Randomized phase II trials in specific subsets of older patients can thus potentially provide relevant information. In these cases, physical status (frailty and vulnerability) could be used as a stratification factor to explore the benefit of treatment in different older populations. Often, no real standards exist for this population (because standard therapy for that disease/indication is expected to be too intense for that person), and all treatments/study arms could actually be seen as experimental arms. Although it might be difficult to select a control arm in a randomized phase II trial, one possibility would be to make the control arm the physician's decision. Because of the methodologic difficulties of defining appropriate control arms for the reasons mentioned in this article, randomized phase II trials might sometimes turn out to be infeasible. A pragmatic option for frail patients could be to perform only single-arm phase II studies with toxicity as an end point, allowing indirect comparison of toxicity (and efficacy) with fit young/old populations from previous studies. This kind of study could provide relevant information if the appropriate end points (HRQoL, functionality, and so on) are included but would be scientifically much less robust than randomized phase II or III studies. Nevertheless, this type of study is sometimes the only feasible option, and regimens studied this way, such as the R-miniCHOP (rituximab plus low-dose cyclophosphamide, doxorubicin, vincristine, and prednisone) regimen in patients age > 80 years with diffuse large B-cell lymphoma,[37] have been adopted in clinical care because higher-level data are lacking.

Aging is a highly individualized process that results in several changes in organ function, affecting the pharmacokinetics of anticancer drugs.[38] These organ system changes may result in altered drug metabolism, with a major impact on treatment tolerability. For that reason, pharmacokinetic studies and phase I studies should be designed specifically for older patients. New drugs could, for instance, be studied in amended phase I studies in populations with higher levels of comorbidity or functional limitations in parallel with standard phase I trials or after the drugs have shown promising results in the general population. An approach in the same line is to design phase I/II–type trials with progressively increasing inclusion criteria. The regimen of interest is first administered to patients in good condition, then in cohorts with increasing levels of functional limitations or comorbidities. This would provide evidence-based thresholds for dose reductions or regimen changes. Risk indicators that could be used for this approach include the CRASH (Chemotherapy Risk Assessment Scale for High-Age Patients) score,[39] the CARG (Cancer and Aging Research Group) score,[40] or criteria such as those used in lymphoma studies.[41,42]

Although incorporating geriatric assessment into oncology trials is usually feasible,[43] the major obstacle to using this as a stratification or even randomization factor is the exact/optimal definition of frailty or vulnerability. Balducci and Extermann[44] formulated an operational definition of frail, fit, and vulnerable patients in 2000 that is commonly used in the oncology world but has significant shortcomings; unfortunately, 10 years later, it is still not clear which are the best criteria and tests to be used to make this stratification.

### Trials of Treatment Regimens Versus Trials of Treatment Strategies Versus Observational Cohort Studies

Randomized trials of treatment regimens comparing treatment A versus treatment B can provide important information. The CALGB (Cancer and Leukemia Group B) 49907 adjuvant breast cancer trial, for instance, showed that classical adjuvant chemotherapy (AC [doxorubicin and cyclophosphamide] or CMF [cyclophosphamide, methotrexate, and fluorouracil]) was clearly superior to so-called soft chemotherapy with capecitabine.[45] New drugs also need to be tested specifically in the older population because specific adverse effects might occur that potentially change the toxicity/benefit ratio. The older population represents a huge potential market for the pharmaceutical industry, but the enhanced risk of toxicity as well as non–treatment-related adverse events that sometimes occur in older patients might lessen the enthusiasm of the industry to support such trials and might hamper drug development and registration.

Trials of treatment strategy comparing no treatment with treatment (eg, prostate cancer surgery or no surgery; breast cancer adjuvant chemotherapy or not) are some of the most important kind of trials that need to be performed. However, several challenges exist. Persuading a patient to participate in a trial of therapy versus no therapy is generally much more difficult than participation in a trial of treatment A versus B, and selection bias and crossover will occur. In the former situation, the impact of random assignment (eg, chemotherapy or not) on older patients is much bigger than in the latter situation (eg, chemotherapy A *v* B). There are possible trial designs that might make this more palatable to patients, such as a cluster randomization design or postrandomization (double) consent design (also called the Zelen design), but these designs are less rigorous because they rely on unverifiable assumptions (eg, patient referral patterns). For both of these approaches, patient consent is sought for the study after the patient already knows which treatment (if any) he or she would receive, removing the anxiety that impending random assignment may produce. Another aspect is that funding is much more difficult to obtain for treatment strategy studies, because there is generally no benefit for industry (on the contrary, the omission of treatment might be disadvantageous for industry). Several attempts at trials of treatment strategy have failed in the past because of these and other reasons, as was nicely demonstrated in the ACTION (Adjuvant Chemotherapy in Older Women) trial for early breast cancer.[46] It should be noted that problems of accrual to trials that compare different treatment modalities or the omission of treatment in one arm are the same for younger, fit populations. Although treatment strategy trials are difficult, it is important that work continue on developing and using alternative designs for these types of trials in the nonfit older population. There is no perfect solution for this, but one pragmatic strategy is to invest much more in large observational cohort studies in the nonfit older population[47] or even in registry studies in the community. If possible, they can be linked to randomized trials, allowing the capturing of the nonincluded population as well as the assessment of different treatments and strategies with regard to outcome. This integration of an RCT into a registry trial increases the quality of an RCT, because the patient selection is better described, and it is better known to which patient populations the results of the RCT can be generalized.

### Incorporation of Geriatric Assessment Into Clinical Trials

Geriatric assessment has not been used often in previous clinical trials, but it should become more frequently required in the future. Without geriatric assessment information, it is impossible to evaluate which older individuals were included in a trial (eg, fit patients only or fit as well as frail patients), limiting extrapolation of the study data to the general older population. This should be mandatory in registration trials and elderly-specific trials and should be encouraged in all trials including older people. However, many different forms of geriatric assessment exist, which complicates comparisons across trials. It is important to agree on a (more or less) uniform or at least comparable evaluation of the older population. EORTC has made an attempt by providing a minimal data set for geriatric assessment to be included in clinical trials,[48] and CALGB has also demonstrated the feasibility of a mainly self-administered tool in its trials,[49] but there are other options,[41,42] and it is important to continue international discussion on this topic.

### Eligibility Criteria

The generally long list of inclusion and exclusion criteria during the last decade has led to selection bias and exclusion of older patients. Exclusion criteria are not based on a high level of evidence. In clinical trials, especially those focusing on older patients with cancer, an attempt should be made to have as few inclusion and exclusion criteria as possible. A National Institutes of Health team concluded that decreasing function and comorbidity restrictions can dramatically increase elderly accrual to clinical trials.[34]

### *European Medicines Agency and US Food and Drug Administration Geriatric Investigation Plan*

In the medical care of pediatric patients, the European Medicines Agency (EMA) has established a pediatric investigation plan to ensure that drugs are examined appropriately in the pediatric population. There is a need for a global strategy within the EMA/US Food and Drug Administration (FDA) to do the same in the older population. Compulsory use of uniform geriatric assessment and frailty tools in drug registration trials could be helpful in establishing a better view of the fitness of older patients included in clinical trials. The EMA/FDA could require adequate representation of older adults in registration trials if applicable (with information from geriatric assessment) or require postmarketing safety studies in the general older population. The EMA recently established a geriatric expert group for this purpose.[50] Longitudinal as well as baseline evaluation of geriatric parameters (eg, functionality, social situation, QoL) is crucial to better understanding the impact of new therapies on older individuals and to improving care for this important population.

## DISCUSSION

Choosing end points for clinical trials in older patients with cancer requires careful reflection on the ultimate goals of therapies. OS is a crucial end point, but DSS should also be recorded in trials where older patients with cancer are included, because deaths resulting from other causes (eg, other diseases, treatment toxicity) occur much more frequently in the older population. Composite end points allow the integration of multiple dimensions in addition to efficacy (eg, QoL, evolution of functionality) into the definition of treatment benefit and have clear advantages in RCTs involving older patients with cancer, such as simplicity of statistical design and statistical efficiency. Composite end points are not feasible in all settings, but they are justified if the individual components of the composite are clinically meaningful and of similar relative importance to clinical care. QoL and preservation of functional capacity and independence are important for the older population and should be included more often as end points in clinical trials in this population.

Although clinical trials in principle should include the entire age range of the population, the heterogeneity of this population generally does not allow the capture of the whole older population, leading to selection bias and difficulty in drawing firm conclusions for the frailer elderly who are often not included. Specific trials for subgroups of older patients with cancer are needed, with additional pharmacokinetic studies if required, and with appropriate control arms depending on the setting. Randomized or single-arm phase II trials can provide insight into the range of efficacy and toxicity in older populations, but ideally they should be confirmed in large phase III trials that are unfortunately often hindered by insufficient interest from sponsors/investors or difficulty in finding sufficient numbers of patients. Large observational cohort studies in the nonfit older population should be considered, preferably linked to randomized trials, to capture the nonincluded population. Incorporation of a preferably uniform geriatric assessment in elderly-specific or registration trials is crucial to better understanding the effect of treatments in different elderly populations. Regulatory authorities including the EMA/FDA should require geriatric assessment information and adequate representation of older adults, including patients of different health statuses such as vulnerable and frail patients, in trials. Better clinical trial design is crucial to understanding the impact of new therapies on older individuals and to improving care for this important population.

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

*Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.*
**Employment or Leadership Position:** None **Consultant or Advisory Role:** Arti Hurria, GTx (C), Seattle Genetics (C) **Stock Ownership:** None **Honoraria:** None **Research Funding:** Arti Hurria, Celgene, GlaxoSmithKline **Expert Testimony:** None **Patents:** None **Other Remuneration:** None

## AUTHOR CONTRIBUTIONS

**Conception and design:** Hans Wildiers, Murielle Mauer, Ulrich Wedding
**Collection and assembly of data:** All authors
**Data analysis and interpretation:** All authors
**Manuscript writing:** All authors
**Final approval of manuscript:** All authors

## REFERENCES


1. Bonnetain F, Doussau A, Mathoulin-Pelissier S, et al: International experts' panel for the development of guidelines for the definition of time to event endpoints in clinical trials (DATECAN project): Results for pancreatic cancer. J Clin Oncol 30:251s, 2012 (suppl; abstr 4053)
2. Fried TR, Bradley EH, Towle VR, et al: Understanding the treatment preferences of seriously ill patients. N Engl J Med 346:1061-1066, 2002
3. Pallis AG, Fortpied C, Wedding U, et al: EORTC elderly task force position paper: Approach to the older cancer patient. Eur J Cancer 46:1502-1513, 2010
4. Lichtman S: Clinical trial design in older adults with cancer: The need for new paradigms. J Geriatr Oncol 3:368-375, 2012
5. Booth CM, Eisenhauer EA: Progression-free survival: Meaningful or simply measurable? J Clin Oncol 30:1030-1033, 2012
6. Schairer C, Mink PJ, Carroll L, et al: Probabilities of death from breast cancer and other causes among female breast cancer patients. J Natl Cancer Inst 96:1311-1321, 2004
7. Piccirillo JF, Feinstein AR: Clinical symptoms and comorbidity: Significance for the prognostic classification of cancer. Cancer 77:834-842, 1996
8. Pallis AG, Karampeazis A, Vamvakas L, et al: Efficacy and treatment tolerance in older patients with NSCLC: A meta-analysis of five phase III randomized trials conducted by the Hellenic Oncology Research Group. Ann Oncol 22:2448-2455, 2011
9. Laskey RA, Poniewierski MS, Lopez MA, et al: Predictors of severe and febrile neutropenia during primary chemotherapy for ovarian cancer. Gynecol Oncol 125:625-630, 2012
10. Hutchins LF, Unger JM, Crowley JJ, et al: Underrepresentation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med 341: 2061-2067, 1999
11. Javid SH, Unger JM, Gralow JR, et al: A prospective analysis of the influence of older age on physician and patient decision-making when considering enrollment in breast cancer clinical trials (SWOG S0316). Oncologist 17:1180-1190, 2012
12. Nout RA, Fiets WE, Struikmans H, et al: The in- or exclusion of non-breast cancer related death and contralateral breast cancer significantly affects estimated outcome probability in early breast cancer. Breast Cancer Res Treat 109:567-572, 2008

13. Manola JB, Gray RJ: When bad things happen to good studies. J Clin Oncol 29:3497-3499, 2011

14. Offen W, Chuang-Stein C, Dmitrienko A, et al: Multiple co-primary endpoints: Medical and statistical solutions—A report from the Multiple Endpoints Expert Team of the Pharmaceutical Research and Manufacturers of America. Drug Inf J 41:31-46, 2007

15. ICH Harmonised Tripartite Guideline: Statistical principles for clinical trials—International Conference on Harmonisation E9 Expert Working Group. Stat Med 18:1905-1942, 1999

16. Lipton A, Fizazi K, Stopeck AT, et al: Superiority of denosumab to zoledronic acid for prevention of skeletal-related events: A combined analysis of 3 pivotal, randomised, phase 3 trials. Eur J Cancer 48:3082-3092, 2012

17. Freemantle N, Calvert M, Wood J, et al: Composite outcomes in randomized trials: Greater precision but with greater uncertainty? JAMA 289: 2554-2559, 2003

18. Cordoba G, Schwartz L, Woloshin S, et al: Definition, reporting, and interpretation of composite outcomes in clinical trials: Systematic review. BMJ 341:c3920, 2010

19. Kleist P: Composite endpoints: Proceed with caution. Appl Clin Trials 15:50-57, 2006

20. US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER): Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf

21. Ardizzoni A, Favaretto A, Boni L, et al: Platinum-etoposide chemotherapy in elderly patients with small-cell lung cancer: Results of a randomized multicenter phase II study assessing attenuated-dose or full-dose with lenograstim prophylaxis—A Forza Operativa Nazionale Italiana Carcinoma Polmonare and Gruppo Studio Tumori Polmonari Veneto (FONICAP-GSTPV) study. J Clin Oncol 23:569-575, 2005

22. National Cancer Institute: Cancer Therapy Evaluation Program: Protocol development. http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm

23. Pallis AG, Agelaki S, Kakolyris S, et al: Chemotherapy-induced neutropenia as a prognostic factor in patients with advanced non-small cell lung cancer treated with front-line docetaxel-gemcitabine chemotherapy. Lung Cancer 62:356-363, 2008

24. Seymour MT, Thompson LC, Wasan HS, et al: Chemotherapy options in elderly and frail patients with metastatic colorectal cancer (MRC FOCUS2): An open-label, randomised factorial trial. Lancet 377:1749-1759, 2011

25. Extermann M, Albrand G, Chen H, et al: Are older French patients as willing as older American patients to undertake chemotherapy? J Clin Oncol 21:3214-3219, 2003

26. Bottomley A, Vanvoorden V, Flechtner H, et al: The challenges and achievements involved in implementing quality of life research in cancer clinical trials. Eur J Cancer 39:275-285, 2003

27. Yellen SB, Cella DF, Leslie WT: Age and clinical decision making in oncology patients. J Natl Cancer Inst 86:1766-1770, 1994

28. Osoba D, Rodrigues G, Myles J, et al: Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 16:139-144, 1998

29. Johnson C, Fitzsimmons D, Gilbert J, et al: Development of the European Organisation for Research and Treatment of Cancer quality of life questionnaire module for older people with cancer: The EORTC QLQ-ELD15. Eur J Cancer 46:2242-2252, 2010

30. Gelber Rd, Cole FB, Gelber S, et al: Comparing treatments using quality-adjusted survival: The Q-Twist method. Am Stat 149:161-169, 1995

31. Reuben DB, Rubenstein LV, Hirsch SH, et al: Value of functional status as a predictor of mortality: Results of a prospective study. Am J Med 93:663-669, 1992

32. Brain EG, Mertens C, Girre V, et al: Impact of liposomal doxorubicin-based adjuvant chemotherapy on autonomy in women over 70 with hormone-receptor-negative breast carcinoma: A French Geriatric Oncology Group (GERICO) phase II multicentre trial. Crit Rev Oncol Hematol 80:160-170, 2011

33. Audisio RA, Pope D, Ramesh HS, et al: Shall we operate? Preoperative assessment in elderly cancer patients (PACE) can help—A SIOG surgical task force prospective study. Crit Rev Oncol Hematol 65:156-163, 2008

34. Lewis JH, Kilgore ML, Goldman DP, et al: Participation of patients 65 years of age or older in cancer clinical trials. J Clin Oncol 21:1383-1389, 2003

35. Scher KS, Hurria A: Under-representation of older adults in cancer registration trials: Known problem, little progress. J Clin Oncol 30:2036-2038, 2012

36. Aapro MS, Köhne CH, Cohen HJ, et al: Never too old? Age should not be a barrier to enrollment in cancer clinical trials. Oncologist 10:198-204, 2005

37. Peyrade F, Jardin F, Thieblemont C, et al: Attenuated immunochemotherapy regimen (R-miniCHOP) in elderly patients older than 80 years with diffuse large B-cell lymphoma: A multicentre, single-arm, phase 2 trial. Lancet Oncol 12:460-468, 2011

38. Wildiers H, Highley MS, de Bruijn EA, et al: Pharmacology of anticancer drugs in the elderly population. Clin Pharmacokinet 42:1213-1242, 2003

39. Hurria A, Togawa K, Mohile SG, et al: Predicting chemotherapy toxicity in older adults with cancer: A prospective multicenter study. J Clin Oncol 29:3457-3465, 2011

40. Extermann M, Boler I, Reich RR, et al: Predicting the risk of chemotherapy toxicity in older patients: The Chemotherapy Risk Assessment Scale for High-Age Patients (CRASH) score. Cancer 118:3377-3386, 2012

41. Tucci A, Ferrari S, Bottelli C, et al: A comprehensive geriatric assessment is more effective than clinical judgment to identify elderly diffuse large cell lymphoma patients who benefit from aggressive therapy. Cancer 115:4547-4553, 2009

42. Soubeyran P, Khaled H, MacKenzie M, et al: Diffuse large B-cell and peripheral T-cell non-Hodgkin's lymphoma in the frail elderly: A phase II EORTC trial with a progressive and cautious treatment emphasizing geriatric assessment. J Geriatr Oncol 2:36-44, 2011

43. Puts MT, Hardt J, Monette J, et al: Use of geriatric assessment for older adults in the oncology setting: A systematic review. J Natl Cancer Inst 104:1134-1164, 2012

44. Balducci L, Extermann M: Management of the frail person with advanced cancer. Crit Rev Oncol Hematol 33:143-148, 2000

45. Muss HB, Berry DA, Cirrincione CT, et al: Adjuvant chemotherapy in older women with early-stage breast cancer. N Engl J Med 360:2055-2065, 2009

46. Leonard R, Ballinger R, Cameron D, et al: Adjuvant chemotherapy in older women (ACTION) study: What did we learn from the pilot phase. Br J Cancer 105:1260-1266, 2011

47. Mohile SG, Wildiers H: A call for observational cohort studies in geriatric oncology. J Geriatr Oncol 3:291-293, 2012

48. Pallis AG, Ring A, Fortpied C, et al: EORTC workshop on clinical trial methodology in older individuals with a diagnosis of solid tumors. Ann Oncol 22:1922-1926, 2011

49. Hurria A, Cirrincione CT, Muss HB, et al: Implementing a geriatric assessment in cooperative group clinical cancer trials: CALGB 360401. J Clin Oncol 29:1290-1296, 2011

50. Cerreta F, Eichler HG, Rasi G: Drug policy for an aging population: The European Medicines Agency's Geriatric Medicines Strategy. N Engl J Med 367:1972-1974, 2012