the publications through 1992. I'd like to thank the very hard work of Maurizio Tonato and his group for putting together such an excellent program.

Tonato: I need an extra consensus from the panel, Richard. We have a prize from an Italian newspaper that should be given to an Italian researcher in this field and so I propose Fausto Roila as the recipient and I'd like for the panel to support this choice.

Gralla: I'd like to second that.

Eur J Cancer, Vol. 29A, Suppl. 1, pp. S51–S58, 1993
Printed in Great Britain

0964–1947/93 $6.00 + 0.00
Pergamon Press Ltd

Quality of Life Assessment in Individuals with Lung Cancer: Testing the Lung Cancer Symptom Scale (LCSS)

Patricia J. Hollen, Richard J. Gralla, Mark G. Kris and Lisa M. Potanovich

This paper presents the continued development and multi-institutional testing of an instrument focusing on measuring the physical and functional dimensions of quality of life. It emphasises evaluation of symptoms associated with lung cancer and their effect on activity status. The Lung Cancer Symptom Scale (LCSS) is a disease- and site-specific instrument which has both a patient and an observer (health care professional) form. The patient scale required 8 min to administer and the observer scale 2 min. The readability index was second-grade level for the patient scale and ninth-grade level for the observer scale. Content validity revealed a mean of 96% agreement for all major symptoms among 52 experts surveyed (confidence interval = 86–99%, $P = 0.05$). 69 patients with non-small cell and 52 patients with small cell lung cancer confirmed that the symptoms matched their experiences. Interrater reliability showed consistency for all items but one among 21 raters at eight institutions; that one item was consistent for 20 of the 21 raters. Similar results were found on a 9-month interval replication. Using the Kappa statistic to estimate extent of agreement for repeated interrater reliability, almost perfect agreement was obtained (mean coefficients, 0.95–0.98). Using the same rule of agreement as for Kappa (± one category) intrarater agreement was 95–100% for all 21 raters. Past test re–test reliability indicated high patient reproducibility for 52 patients ($r > 0.75$, $P < 0.01$ for all items). We conclude that (1) the LCSS demonstrates good feasibility, reliability, and content validity, (2) high interrater reliability indicates utility in multicentre trials, and (3) continued testing for internal consistency, construct validity and criterion-related validity is warranted.

Eur J Cancer, Vol. 29A, Suppl. 1, pp. S51–S58, 1993.

INTRODUCTION

IN ADDITION TO reporting response and survival rates, the need to determine the effect of cancer and its treatment on quality of life is clear. Disease-specific instruments, detecting differences in patients with malignant disease, are one method of improving the assessment of subjective parameters or "quality of life" in clinical trial research in oncology [1, 2]. Although generic quality of life instruments applicable to any chronic disease population have been developed, the trend is to develop measures specific to the cancer population [3]. Moreover, development of specific instruments focusing on a particular malignancy is becoming more common. Several problems exist with many of the currently available instruments, including a lack of focus on the clinically important area of symptom management, long and cumbersome questionnaires and an absence of adequate psychometric testing to determine the reliability, validity and utility of the instruments.

Initial development and testing of the Lung Cancer Symptom Scale (by R.J. Gralla and M.T. Burke) began in the mid 1980s [4, 5]. It was specifically designed to address the issues of palliation and symptom control in evaluating patients receiving new chemotherapy regimens. The favourable initial psychometric properties of the Lung Cancer Symptom Scale (LCSS) encouraged continued systematic development to

Correspondence to P.J. Hollen.
P.J. Hollen is at the University of Rochester, 601 Elmwood Avenue, Rochester, NY 14642; R.J. Gralla is at the Ochsner Cancer Institute, Ochsner Medical Institutions, 1516 Jefferson Highway, New Orleans, LA 70121; and M.G. Kris and L.M. Potanovich are at the Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, U.S.A.
Received 30 Jan. 1992; accepted 8 May 1992.

establish fully its applicability to multicentre clinical trials [6]. More development of this practical and easy to administer instrument is warranted. There is currently no accepted instrument for quality of life; the need for such an instrument in a disease as widespread as lung cancer is clear. The current psychometric study was conducted to extend the feasibility, reliability and validity estimations of the LCSS.

Review of quality of life instruments

Pursuit of measurement of quality of life factors, such as health and welfare, was suggested in 1960 by the US President's Commission on National Goals [7, 8]. Controversy surrounding the measurement of benefits of chemotherapy beyond response and survival rates prompted renewed interest in developing reliable and valid subjective instruments. The Federal Drug Administration (FDA) ruling in 1985 that new drug approval would necessitate either improvement in survival rates or quality of life, has made the need for instruments with increased specificity more pressing.

Among the many available quality of life instruments, eight are frequently discussed, either because of their approach or because of promising psychometric properties: Karnofsky Performance Status (KPS), Sickness Impact Profile (SIP), Quality of Life Index (QL-Index), Ferrans and Powers' Quality of Life Index (QLI), Priestman and Baum's Linear Analogue Self-Assessment (LASA), Padilla and colleagues' Quality of Life Index (QLI), Functional Living Index—Cancer (FLIC), and Selby and colleagues' Linear Analogue Self-Assessment (LASA) [9–16]. Of these, four are generic [9–12], three are intended for use in the general cancer population [13–15], and one is designed for a specific malignancy [16]. A second version of two of these instruments has been developed, each with increased population specificity [17, 18]. One is not truly a quality of life measure but has been widely used for this purpose in the past [9]. None of the cancer-specific measures in this group adequately addresses symptoms of lung cancer. Measures that have both a patient-rating and an observer-rating form are advocated to enhance validity when selecting quality of life measures since the simultaneous ratings fully capture the cancer experience [19]; none of these measures has both forms. Further examination of psychometric properties is needed in nearly all these instruments [2, 20].

All but one of this group focus on most of the global dimensions of quality of life: physical, psychological, social and spiritual. While such a global assessment can be useful in a variety of settings, it is doubtful that this approach is appropriate for determining a more focused outcome, such as the impact of a new treatment on the symptomatic aspects of cancer. It would appear that palliation of physical symptoms (the physical dimension) is more likely to be influenced by a new treatment than would the other dimensions; thus, assessment of all dimensions might introduce inaccuracy by measuring life areas affected more strongly by factors other than the new treatment. In addition, many of the areas assessed by these instruments may relate to problems existing before the malignancy. This broad scope, rather than a more specific one focused on the critical concern of relief of disease-related symptoms, can lead to disenchantment with the methods of assessing subjective factors in general. This broad scope may also explain why these instruments are not well accepted. Opinions can differ widely. One developer of a quality of life instrument prescribes at least three of the generally accepted dimensions, whereas another believes that at least four are needed [19, 21]. Some have questioned the practicality of trying to measure such broad dimensions as happiness and life satisfaction, since this puts an inordinate burden on the health care system in general, let alone any specific intervention, and may even exceed the charge of the healing professions [22].

Table 1. Items of questionable value in assessing quality of life as part of a new treatment trial

Functional Assessment of Cancer Therapy: FACT-L (Lung) Scale Example items
I regret my smoking.
I feel guilty for not seeking treatment sooner.
I feel distant from my friends.
I feel close to my partner (or main source of support).
I am satisfied with my sex life.
My work (including housework) is fulfilling.
I am doubting my faith in God.
I get comfort or strength from my faith.
I have confidence in my doctor(s).
My doctor is available to answer my questions.

Ref. [25].

Attempts have been made to establish psychometric properties for the above-mentioned eight instruments, resulting in varying degrees of measured reliability and validity. However, using the development criteria for quality of life measures delineated by Donovan and colleagues, none of these instruments adequately meets all five of the following criteria: (a) appropriate conceptualisation of four life areas (physical, psychological, spiritual and social) affected by cancer and its treatment as well as positive aspects, (b) acceptable psychometric properties, (c) responsiveness to disease or treatment

Table 2. Evolving quality of life items

EORTC Core Quality of Life Questionnaire and Lung Cancer-Specific Questionnaire Module Example items
1987 version (45 items)
Did you feel lonely?
Did you worry?
Did you feel that things were going your way?
Overall, how satisfied do you feel?
There was someone to talk to when I wanted.
I felt close to someone.
People seemed to understand my problems.
1988 verson (49 items)
Did you look forward with enjoyment to things?
Could you enjoy a good book or radio or television program?
Could you laugh and see the funny side of things?
Has your condition interfered with your family or social life?
Has your medical treatment interfered with your family or social life?
Has your condition or treatment caused you financial difficulties?
1990 version (43 items)
Did you worry?
Has your physical condition or medical treatment interfered with your family life?
Has your physical condition or medical treatment interfered with your social activities?
Has your physical condition or medical treatment caused you financial difficulties?

Refs [21], [23] and [24].

changes in quality of life, (d) patient-generated data as a base, and (e) patient, health care provider, and researcher acceptability [2].

In addition to these eight instruments, two instruments specific to lung cancer were also reviewed. A multidimensional quality of life instrument has been developed as a collaborative effort by the Study Group on Quality of Life and the European Organisation for Research and Treatment of Cancer (EORTC) [21, 23, 24]. 13 of the 43 items of this modular instrument are related to disease- and treatment-related symptoms for lung cancer. A second site-specific instrument is the Functional Assessment of Cancer Therapy (FACT) scale [25]. This 47-item instrument has a nine-item module specific to site, and has been developed for four types of cancer (breast, lung, colorectal, head and neck) and one general version. While both of these disease- and site-specific instruments appear promising, both are in the early stages of psychometric development. More importantly, the question arises whether such global instruments are justified for evaluation of new agents or new treatment trials. As shown in Table 1, some of the items within the FACT-L may not be applicable for drug or treatment efficacy trials. This problem may be becoming clearer to instrument developers. As with the EORTC measure, items are evolving, with a reduction in the number of items in the psychological, social and spiritual dimensions (Table 2). Additionally, the EORTC instrument demonstrates the development of summation items more focused on an individual's physical condition or medical treatment within these dimensions (Table 2).

The importance of each quality of life dimension has not been determined; the process is complicated because individuals' value of life experiences differ, affecting their perceptions of quality of life [12, 26]. In contrast to the other instruments, the LCSS focuses primarily on the physical dimension, particularly symptoms and symptomatic distress specific to lung cancer, and subsequent quality of life. It also includes summative items to assess perception of activity status and overall quality of life. It has been designed to measure whether a treatment itself gives symptom control for individuals with lung cancer and how this affects activity status and quality of life. A summary of the previously completed psychometric testing, encouraging the further testing, is shown in Table 3.

MATERIALS AND METHODS

Design and settings

This psychometric study had three phases: feasibility, content validity and reliability. A brief description of each of the phases is given below.

Content validity testing included 25 regional cancer centres in the U.S.A. and Canada. Intra- and interrater reliability were measured at eight sites in the U.S.A. and Canada.

Feasibility. Few quality of life subjective instruments have been evaluated for ease of comprehension as part of feasibility testing [2]. No absolute standard of reading level has been established; however, according to a 1971 U.S. literacy study estimate, 37% of adults read below the eighth-grade reading level [27]. Based on 1982 data of a national survey by the US Department of Education, analysed in 1986, 13% or 18.7 million adults were functionally illiterate ($P<0.05$) [28]. This may be why one instrument developer employed sixth-grade reading level as a base for his frequently used psychological symptom scale [29]. For the LCSS, the reading level of both the patient and observer forms were tested.

Table 3. Initial psychometric properties for the Lung Cancer Symptom Scale (LCSS)

Properties	Methods	Samples	Results
Feasibility	Completion time (patient scale) (observer scale) Skill level (observer scale)	$n=20$ LC patients 2 raters	Mean time 8 min 2 min Minimal skill required
Content validity	Patient acceptance Panel of experts	$n=52$ $n=4$ Experts	High acceptance Agreement/ consensus
Reliability	Test–retest with 1-h interval prior to R_x Interrater 1-h interval prior to R_x	$n=52$ LC patients 2 raters with $n=52$ LC patients	Pearson $r>0.75$, $P<0.01$ for all items Pearson $r>0.75$, $P<0.01$ for all items except cough (0.65) and weakness* (0.54)
Construct validity	Patient scale with observer scale	$n=52$ LC patients	Pearson r, ($P<0.01$) Cough (0.74) Dyspnoea (0.66) Haemoptysis (0.69) Pain (0.71) Weight loss (0.61) *weakness (0.49)
Criterion-related validity	Patient activity item with KPS prior to R_x	$n=40$ LC patients	Pearson $r=0.59$, $P<0.01$

Refs [4] and [5].
*Item dropped due to low reliability.

Content validity. Patient- and health care professional-generated items in addition to literature reports are the foundation of many quality of life instruments [20]. There is also a growing opinion that the individual's perspective is the only reasonable one in estimating quality of life; objective data reported by an observer is important though supplementary [26]. Thus, content validity of an instrument by both a panel of patients with small cell and non-small cell lung cancer and by experts experienced with lung cancer gives the best base for establishing the universe of content for a lung cancer scale.

Reliability. A sequential training process was employed that includes training raters to use the instruments, evaluating performance at the end of the training, and ensuring rater performance over time [30]. Interrater reliability was evaluated within and among sites to determine consistency of the scales for multicentre trials. Intrarater reliability was estimated at a 9-month interval to evaluate the extent to which a rater applies the same scoring criteria on two occasions.

Subjects

Content validity. Three separate panels were enlisted: (1) individuals with advanced lung cancer, (2) physicians specialising in medical oncology with an emphasis on lung cancer, and (3) oncology nurses. The 69 individuals with advanced non-small cell lung cancer consisted of 67% men and 33% women whose median age was 59 (range: 31–79 years). Prior therapy for these patients included chemotherapy (80%), radiotherapy (38%), and surgery (20%). The median time from diagnosis was 8 months (range: 0.25–56 months). KPS ratings were in the following ranges: 20–50% (28%); 60–70% (30%); and 80–90% (42%). A panel of 52 individuals with small cell lung cancer consisted of 38% men and 62% women whose median age was 61 (range: 34–82). Prior treatment for this group included chemotherapy (94%), radiotherapy (29%), surgery (11.5%), and no treatment (6%). Their median time from diagnosis was 5 months (range: 0–120 months). The KPS for this group were classified as: 50% (8%), 60–70% (42%), 80–90% (50%).

A panel of 24 physicians and 28 nurses were surveyed as experts. The mean number of years in oncology for the physicians was 13 (range: 6 to >20 years) with 100% being board certified in internal medicine and 96% in medical oncology. The median number of publications for physicians was 20 (range: 2 to >100) with 88% having three or more lung cancer publications. The mean number of years as an oncology nurse was 8 (range: 3–14 years). 68% were members of the Oncology Nursing Society with 39% of these nurses certified in this specialty. The number of publications in oncology for the nurses averaged four (range: 0–60) with a mean of one in lung cancer.

Reliability. Two or three observers from each of eight sites, for a total of 21, served as raters for the reliability phase of the study. All but two of the observers were registered nurses with at least 1 year of oncology experience; two others with the same experience were physicians working in data management.

Instruments

Expert and patient panel survey instruments. A two-page, short-answer questionnaire was used to assess major symptoms of lung cancer and demographics for each of the panels.

LCSS. This subjective rating scale focuses on the physical and functional dimensions of quality of life, primarily measuring major lung cancer symptoms and their effect on activity status. It consists of two instruments, one for patients and one for health professionals as observers [4–6].

The patient scale consists of nine items: six measuring major symptoms for lung malignancies, and three summation items related to total symptomatic distress, activity status and overall quality of life, all using visual analogue scales (Fig. 1). The interval level visual analogue scale uses 100-mm lines to measure the intensity of patient responses for appetite, fatigue, cough, dyspnoea, haemoptysis, pain, symptomactic distress from lung cancer, activity level and quality of life. The time frame is the past day based on the rationale that the scale items represent states or conditions that may quickly fluctuate, not enduring traits. Each item, on separate card, is given an individual score equal to the length of the line marked by the patient with zero corresponding to the lowest rating and 100 representing the highest rating. The instrument was shown to be feasible, reliable (using Nunnally's standard of a coefficient of $r \geqslant 0.70$ for new instruments [31]), and some support for validity, as well as used in prior research (Table 3).

The observer scale is a 5-point ordinal level scale similar in content to the patient scale (Fig. 2). It measures the intensity of six major lung cancer symptoms (loss of appetite, fatigue, cough, dyspnoea, haemoptysis and pain). Scores are recorded after a patient interview in which the observer is directed to assess lung cancer symptoms using the time frame of the past day. Response categories range from 100 = none and 0 = severe with an individual score obtained for each of the items. The observer scale has been shown in prior research to be feasible, reliable, and some support for validity has previously been established (Table 3).

Karnofsky scale (KPS). The observer also determined the performance status using the KPS, a well accepted instrument within oncology. The KPS measures physical functioning based on level of activity, symptoms of disease, and amount of assistance needed, and it categorises a patient's level of activity from 0% = deceased, to 100% = normal activities with no signs or symptoms of malignancy, divided by deciles [9].

Procedures

Feasibility. To determine the reading level required for the LCSS, the Flesch–Kincaid Readability Index was used. This index examines the average number of syllables per word and the average sentence length [32]. Estimating the readability indices for the LCSS was accomplished by using RightWriter™, a software programme for grammatical evaluation that includes several readability indexes (RightSoft, Inc., Sarasota, FL).

Content validity. The survey instrument for experts was administered to the panel of physicians after a presentation of the nature of the psychometric project. A mailed survey approach, using the same instrument, was used to obtain the responses of the panel of expert nurses.

Each patient for the panel was interviewed by a nurse at the time of an outpatient appointment. To standardise the procedure for this phase, the measure was completed: (a) on a day of treatment, (b) before the patient received results from any clinical tests, and (c) prior to chemotherapy, generally 1 h preceding treatment.

Fig. 1. Lung Cancer Symptom Scale (LCSS): patient scale

Directions: Please place a mark along each line where it would best describe the symptoms of your lung cancer DURING THE PAST DAY.

1. How is your appetite?
   As good as it could be |_____| As bad as it could be
2. How much fatigue do you have?
   None |_____| As much as it could be
3. How much coughing do you have?
   None |_____| As much as it could be
4. How much shortness of breath do you have?
   None |_____| As much as it could be
5. How much blood do you see in your sputum?
   None |_____| As much as it could be
6. How much pain do you have?
   None |_____| As much as it could be
7. How bad are your symptoms from lung cancer?
   I have none |_____| As bad as they could be
8. How much has your illness affected your ability to carry out normal activities?
   Not at all |_____| So much that I can do nothing for myself
9. How would you rate the quality of your life today?
   Very high |_____| Very low

Note: The Lung Cancer Symptom Scale is under ongoing development. Permission for use will be granted by the developers when the psychometric testing is complete in exchange for pertinent data related to enhancing these properties.

Fig. 2. Lung Cancer Symptom Scale (LCSS): observer scale

Directions: Direct the interview to assess lung cancer symptoms using the timeframe of DURING THE PAST DAY.

1. Loss of appetite: (Score: _____)
   - 100 *None.*
   - 75 *Mild*; occasional loss of appetite but does not interfere with food intake.
   - 50 *Moderate*; occasional loss of appetite which occasionally interferes with food intake.
   - 25 *Marked*; frequent loss of appetite which generally interferes with food intake.
   - 0 *Severe*; appetite so poor that medical intervention for feeding (intravenously or feeding tube) is needed.
2. Fatigue: (Score: _____)
   - 100 *None.*
   - 75 *Mild*; occasionally troubled by modest fatigue.
   - 50 *Moderate*; usually troubled by modest fatigue.
   - 25 *Marked*; occasionally troubled by major fatigue.
   - 0 *Severe*; usually troubled by major fatigue.
3. Cough: (Score: _____)
   - 100 *None.*
   - 75 *Mild*; present and increased over a year ago, but not bothersome; no medications needed.
   - 50 *Moderate*; bothersome; leads to SOB on occasion.
   - 25 *Marked*; bothersome; disturbs sleep and other normal functioning.
   - 0 *Severe*; nearly constant; disrupts any normal activities.
4. Dyspnoea: (Score: _____)
   - 100 *None.*
   - 75 *Mild*; noticed only with major activity (e.g. climbing more than one flight of stairs); SOB does not limit usual activities.
   - 50 *Moderate*; present when walking at normal pace; interferes with ability to carry out some usual activities.
   - 25 *Marked*; present with minimal activity; supplemental O_2 used only occasionally.
   - 0 *Severe*; supplemental O_2 required most or all of the time.
5. Haemoptysis: (Score: _____)
   - 100 *None.*
   - 75 *Mild*; blood in sputum, less frequently than daily.
   - 50 *Moderate*; blood in sputum at least daily but generally just "flecks" as part of the sputum.
   - 25 *Marked*; sputum is often purely bloody (not just flecks) on a daily basis.
   - 0 *Severe*; same as marked but blood loss by haemoptysis measurably lowering haemoglobin.
6. Pain: (Score: _____)
   - 100 *None.*
   - 75 *Mild*; present but either no medications required or only non-narcotic, non-codeine type oral agents; pain control satisfactory or reasonable.
   - 50 *Moderate*; codeine or codeine containing oral medications needed; pain control satisfactory or reasonable.
   - 25 *Marked*; narcotic oral agents are required; pain control satisfactory, or reasonable.
   - 0 *Severe*; narcotic oral medications required but pain control not satisfactory, or parenteral narcotics are required.

Reliability. Standardised training procedures were used to obtain inter-rater reliability for the LCSS observer scale. A half-day training session was held for the observers. Rater training followed a protocol in which the content had been carefully identified and ordered. An instructional manual delineated procedures for administering test materials and scoring. At the end of the training session, rater performance was evaluated using two videotaped interviews with lung cancer patients to ensure standardisation in observation. 2 cases, 1 male and 1 female patient, were chosen to provide examples of differing performance status (KPS of 70% and of 90%).

Raters assessed the two cases using the LCSS observer form and the KPS at the end of the training session and repeated this procedure midway through patient accrual (9 months later) by mailed response. Using the mode scores of raters to provide feedback, they were told of their relative placement on specific items for the two cases if there was a statistically significant difference in scoring from the other raters.

RESULTS

A total of 24 physicians, 28 nurses, and 173 lung cancer patients were included in the feasibility, reliability and content validity phases. Content validity included 25 regional cancer centres in the U.S.A. and Canada, with eight of these sites participating in the reliability phase.

Feasibility

The LCSS was assessed initially for administration time. The patient form requires a mean of 8 min for administration and the observer form, 2 min (Table 3). The Flesch–Kincaid Readability Index determined a second-grade reading level for the patient form and a ninth-grade level for the observer form.

Content validity

The LCSS had been originally assessed for universality or representation of items for content validity by four experts in lung cancer at one cancer centre (Table 3). This phase was expanded to include a larger panel of experts (physicians and nurses) from multiple institutions and a panel of patients to validate the major symptoms with their experiences. Content validity revealed a mean of 96% agreement for all symptoms among the 52 experts surveyed (95% confidence interval = 86–99%). 69 non-small cell and 52 small cell lung cancer patients confirmed that the symptoms matched their experiences (Table 4).

Reliability

Interrater reliability was expanded from the initial two raters at one cancer centre (Table 3). Interrater reliability showed consistency for all items but one, using a null hypothesis that there would be no significant difference among the 21 raters at the eight cancer centres and no difference between the two or three raters at each site ($P > 0.05$, Friedman Test); the dyspnoea item was consistent among 20 of the 21 raters. Similar results were found on a 9-month interval replication. Though it was not necessary to repeat interrater reliability, measurement at midway through accrual (9-month interval) provides additional support of the stability of raters when using the LCSS.

Table 4. Comparison of major lung cancer symptoms

Symptoms	Non-small cell patients (%) ($n=69$)	Small cell patients (%) ($n=52$)
Fatigue	84	79
Decreased activity	81	63
Cough	71	62
Dyspnoea	59	56
Decreased appetite	57	60
Weight loss	54	52
Pain	48	54
Haemoptysis	25	14

With the Kappa statistic, as a weighted formula to measure the degree of agreement among raters for the LCSS Observer scale, the coefficients for repeated inter-rater reliability were 0.95–1.00 for time 1 and 0.95–0.97 for time 2 with a mean agreement of 0.95–0.98 for all six items. A rule for weighting (± one rating category) was developed for the weights for the Kappa statistic; six lung cancer researchers agreed as experts that using a weighted rule of agreement was appropriate. The rationale for weighting was that expecting perfect agreement across all raters may be unreasonable, i.e. many accept that the KPS may be off by 10 (± one rating category).

For evaluation of intrarater reliability of the LCSS Observer form among the 21 raters over a 9-month interval, items were analysed by per cent of agreement for both a strict comparison and a relaxed comparison (± one rating category). Perfect agreement for the six items ranged from 57 to 100% for the case of the man and 48–95% for the case of the woman; for the relaxed comparison, agreement ranged from 95 to 100% for both cases. The dyspnoea item had the most inconsistency regardless of type of agreement (perfect agreement, 57–71%; relaxed, 95–100%). There was too much missing data in the case of the man to estimate the appetite item, in that this item was not fully assessed in the videotaped interview; however, raters were consistent over time for the case of the woman (perfect agreement, 81%; relaxed agreement, 100%). Three experts had originally reviewed the taped interviews, but an additional researcher, not related to the study, concluded that more questions related to these two items would have enhanced agreement among raters.

As shown in Table 3, original patient test–retest reliability results (administered prior to treatment; with a 1-h interval between tests and with the multiple items given in a different order with each test) indicated high reproducibility for 52 patients ($r > 0.75$, $P < 0.01$ for all items). Original interrater reliability results were high for all items except cough (0.65) and weakness (0.54). Although initial construct validity will not be discussed in depth, the weakness item was again low ($r = 0.49$, $P < 0.01$) when patient scores were correlated with observer scores. This item was later changed to "fatigue", as discussed below.

Using the Friedman test, there was no significant difference in the observers classifying the two patients whose interviews had been videotaped, by the Karnofsky performance scale: time 1 ($F = 1.47$, $P = 0.20$); time 2 ($F = 1.06$, $P = 0.45$).

DISCUSSION

Feasibility

The LCSS takes little time to administer and can easily be understood by patients and observers. Its feasibility has been demonstrated by patient, health care provider and researcher acceptability in previous research (Table 3). The initial psychometric properties of the LCSS were promising; these

results encouraged a systematic and complete psychometric evaluation.

Content validity

The LCSS is designed specifically for use with lung cancer patients. While certain items may apply to different malignancies or diseases, the set of major symptoms were validated by expert health care professionals and patients for both small cell and non-small cell lung cancer. The LCSS has the potential to detect differences in symptoms and symptomatic distress along the illness continuum; it can be useful in assessing this important area in clinical trials in lung cancer.

Reliability

Interrater reliability with 21 observers at eight cancer centres yielded satisfactory results and was comparable to the initial research.

According to Landis and Koch, useful benchmarks for the Kappa statistic are: 0.00–0.20 = slight agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement, and 0.81–1.00 = almost perfect agreement [33]. Thus using the kappa statistic with a weighted rule of agreement, among 21 raters from eight cancer centres, the LCSS observer scale had almost perfect interrater reliability and similar results for a 9-month repetition.

Using per cent of agreement for intrarater reliability, there was 95–100% agreement for all items, allowing one rating category as less than perfect agreement.

The exception in reliability from the original data is an item describing weakness. This item was less reliable than other items for both patients and observers. The item was changed to "fatigue" and validated by the panel of experts. Eighty-four per cent of the non-small cell lung cancer patient panel and 79% of the small cell panel also validated the fatigue symptom. A similar result was previously reported in a study of fatigue in which 96% of patients with either breast or lung cancer experienced some level of fatigue [34].

A limitation in the testing procedure may have affected perfect agreement among the observers in that more discriminative questions could have been asked to improve the quality of the taped interviews for assessing certain items, specifically dyspnoea and appetite.

In addition to consistency with the LCSS, the observers also demonstrated ability to consistently assess patients along another severity continuum, the Karnofsky performance scale.

Utility

The LCSS is unique in that it provides both a patient and an observer form to measure the same symptoms. Although lung cancer symptomatic distress is a subjective state, data from an observational measure (by the health care professional) provides context and confirmation for patients' reports. This is particularly helpful when treatment efficacy is the research question.

The LCSS also has three items for patients to summarise their perception of symptomatic distress from lung cancer, effect on activity status and overall quality of life. The use of summation items allows capture of experiences not specified in the other LCSS items, while still allowing an assessment of the entire quality of life experience.

An additional strength of the LCSS is that its items are limited to those of lung cancer rather than including items specific to chemotherapy side-effects; such effects can be documented by already established and widely used instruments, and by measurable objective data [35]. This allows more versatility in the instrument's use since the chemotherapeutic agent or treatment modality may not remain constant and may, therefore, introduce many unique side-effects under different research plans.

The LCSS differs from most quality of life instruments in that it has attempted to meet in depth all of the recently stated criteria for quality of life measures, with one exception [1, 2]. This one exception, conceptualisation of all life areas, is less relevant for symptom control evaluation, the stated goal of the instrument. The concept of all life areas is, however, addressed in a summative question, allowing an assessment in less depth. The psychological, social and spiritual dimensions are of interest for some clinical research questions but are not applicable for all clinical trials. For example, the spiritual dimension is not likely to be affected in a new therapeutic agent study. Some instruments evaluate satisfaction with medical and nursing care. Again, if the primary question involves the comparison of the efficacy of two treatments, this evaluation of staff qualities could in fact confound the assessment of a new agent. Depending on the research questions of interest, the importance of certain quality of life dimensions varies: for example, research investigating drug efficacy may view the physical and functional dimensions as most important.

Practical disease- and site-specific quality of life instruments, such as the LCSS, need to be tested for clinical trials in patients with cancer, especially for the research question of drug efficacy. The priority of quality of life dimensions for clinical trials needs to be altered from a global or comprehensive structure of three or four dimensions for this research aim. Focusing primarily on two dimensions, disease-related symptoms, with accompanying distress, and activity status effect, with subsequent quality of life measured in a summative format, can provide the least patient and staff burden during administration of the measures, especially in serial measurement. The choice of a patient-based questionnaire with an observer counterpart form, such as in the LCSS, is logical for capturing the patient's perspective of the cancer experience with the observer's evaluation providing context to the experience.

The objective of this paper is to outline the rationale for and the continuing development of a disease- and site-specific quality of life instrument, the LCSS, measuring subjective factors in patients with lung cancer. The scale was developed to create an instrument that fulfils a need in evaluating the treatment of patients with lung cancer and to provide an instrument that is easy to use for both patients and staff. We have shown that the LCSS is reliable and valid, possessing appropriate psychometric properties. Based on the criteria presented, the LCSS has the potential to be the strongest of instruments currently available to measure health-related quality of life for the lung cancer experience.

We conclude that further evaluation has shown that the LCSS continues to demonstrate favourable results in feasibility, reliability and content validity. High inter-rater reliability indicates a good potential for utility in multicentre trials. Completion of ongoing testing for internal consistency, construct validity and criterion-related validity is warranted.

1. Schipper H, Levitt M. Measuring quality of life: risks and benefits. *Cancer Treatment Rep* 1985, **69**, 1115–1123.
2. Donovan K, Sanson-Fisher RW, Redman S. Measuring quality of life in cancer patients. *J Clin Oncol* 1989, **7**, 959–968.
3. Aaronson NK, Meyerowitz BE, Bard M, *et al.* Quality of life research in oncology: past achievements and future priorities. *Cancer* 1991, **67**, 839–843.
4. Burke MT, Gralla R, Kris M, Howard J, Monras P. Subjective evaluation in non-small cell lung cancer: comparison of Karnofsky performance status with a patient generated visual analogue scale measuring activity. *Proc IVth World Conference on Lung Cancer*, Toronto, Canada, 1985, **4**, 42.
5. Monras P, Gralla RJ, Burke MT, *et al.* Development of specific instruments for subjective evaluation of patients with lung cancer: comparison of observer assessment with patient generated visual analogue scales (VAS). *Proc Am Soc Clin Oncol* 1985, **4**, 251 (abstract C-978).
6. Hollen PJ, Gralla RJ, Kris MG. Lung Cancer Symptom Scale (LCSS): development of a physical domain quality of life instrument. *Proc Am Soc Clin Oncol* 1991, **10**, 347 (abstract 1234).
7. President's Commission on National Goals. *Goals for Americans*. Columbia University, The American Assembly, 1960.
8. Flanagan JC. Measurement of quality of life: current state of the art. *Arch Phys Med Rehabil* 1982, **63**, 56–59.
9. 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, 191–205.
10. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981, **19**, 787–805.
11. 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 Chron Dis* 1981, **34**, 585–597.
12. Ferrans CE, Powers MJ. Quality of life index: development and psychometric properties. *ANS* 1985, **8**, 15–24.
13. Priestman TJ, Baum M. Evaluation of quality of life in patients receiving treatment for advanced breast cancer. *Lancet* 1976, **1**, 899–901.
14. Padilla GV, Presant C, Grant MM, Metter G, Lipsett J, Heide F. Quality of Life Index for patients with cancer. *Res Nurs Health* 1983, **6**, 117–126.
15. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the Functional Living Index-Cancer: development and validation. *J Clin Oncol* 1984, **2**, 472–483.
16. Selby PJ, Chapman JAW, Etazadi-Amoli J, Dalley D, Boyd NF. The development of a method for assessing the quality of life of cancer patients. *Br J Cancer* 1984, **50**, 13–22.
17. Padilla GV, Grant MM. Quality of life as a cancer nursing outcome variable. *ANS* 1985, **8**, 45–60.
18. Ferrans CE. Development of a quality of life index for patients with cancer. *Oncol Nurs Forum* 1990, **17** (Suppl.), 15–19.
19. Cella DF, Tulsky DS. Measuring quality of life today: methodological aspects. *Oncology* 1990, **4**, 29–38.
20. Grant M, Padilla GV, Ferrell BR, Rhiner M. Assessment of quality of life with a single instrument. *Semin Oncol Nurs* 1990, **6**, 260–270.
21 Aaronson NK, Bullinger M, Ahmedzai S. A modular approach to quality-of-life assessment in cancer clinical trials. *Recent Results Cancer Res* 1988, **111**, 231–249.
22. Aaronson NK. Quality of life: What is it? How should it be measured? *Oncology* 1988, **2**, 69–74.
23. Aaronson NK, Bakker W, Stewart AL, *et al.* Multidimensional approach to the measurement of quality of life in lung cancer clinical trials. In: Aaronson NK, Beckmann J, eds. *The Quality of Life of Cancer Patients*. New York, Raven Press, 1987, 63–82.
24. EORTC Core Quality of Life Questionnaire: Lung Cancer—Specific Questionnaire. *Oncology* 1990, **4**, (Appendix 1), 215.
25. Cella DF, Lee-Riordan D, Silberman M, *et al.* Quality of life in advanced cancer: three new disease-specific measures. *Proc Am Soc Clin Oncol* 1989, **8**, 315 (abstract 1225).
26. Ferrans CE. Quality of life: conceptual issues. *Semin Oncol Nurs* 1990, **6**, 248–254.
27. Monteith MK. How well does the average American read? Some facts, figures and opinions. *J Reading* 1980, **23**, 460–464.
28 Costa M. *Adult Literacy/Illiteracy in the United States: A Handbook for Reference and Research*. Santa Barbara, CA: ABC-CLIO, 1988.
29 Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. *Psychol Med* 1983, **13**, 595–605.
30. Castorr AH, Thompson KO, Ryan JW, Phillips CY, Prescott PA, Soeken KL. The process of rater training for observational instruments: implications for interrater reliability. *Res Nurs Health* 1990, **13**, 311–318.
31. Nunnally JC. *Psychometric Theory*, 2nd ed. New York, McGraw-Hill, 1978.
32. Miller LR. Predictive powers of multiple-choice and cloze-derived readability formulas. *Reading Improvement* 1975, **5**, 52–58.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977, **33**, 159–174.
34. Wickham R, Blesch KS, Paice J, *et al.* Fatigue and its correlates in breast and lung cancer patients. *Oncol Nurs Forum* 1990, **17** (Suppl.), 146 (Abstract 35A).
35. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981, **47**, 207–214.

Acknowledgements—The authors gratefully acknowledge support from the following eight cancer centres: Memorial Sloan-Kettering Cancer Center, Ochsner Cancer Institute, University of Maryland, University of Southern California, Duke University, University of New Mexico, Penn State University, and Princess Margaret Hospital in Toronto, Canada. This work was supported in part by a grant from Ciba-Geigy Corporation.