The Edmonton Symptom Assessment System: a 15-year retrospective review of validation studies
(1991-2006)
Cheryl Nekolaichuk, Sharon Watanabe and Crystal Beaumont
Palliat Med 2008; 22; 111
DOI: 10.1177/0269216307087659

The online version of this article can be found at:
http://pmj.sagepub.com/cgi/content/abstract/22/2/111

Cheryl Nekolaichuk, Sharon Watanabe and Crystal Beaumont Division of Palliative Care Medicine, Department of Oncology, University of Alberta, Edmonton, Alberta, Canada

Introduction: The purpose of this review was to identify and critique validation studies focusing on the Edmonton Symptom Assessment System (ESAS), a commonly used symptom assessment tool for advanced cancer and palliative patients. Methods: Using a comprehensive literature search, the authors identified and screened 87 publications. Thirteen articles were selected for in-depth review, based on the following inclusion criteria: psychometric studies with a primary focus on the ESAS, 1991–2006 publication dates and peer-reviewed English language publications. Results: Most studies involved cancer patients (n = 11). The ESAS format varied across studies, in terms of scale format, item number, item selection and language. Studies focused on gathering reliability estimates (n = 8), content validity evidence (n = 1), concurrent validity evidence (n = 5), predictive validity evidence (n = 1), and sensitivity and/or specificity (n = 3). None of these studies involved patients’ perspectives as a source of validity evidence. Discussion: The use of varying instrument formats and limited psychometric evidence support the need for further ESAS validation studies, including the involvement of patients. Palliative Medicine (2008); 22: 111–122

Key words: Edmonton Symptom Assessment System; ESAS; instrument development; literature review; psychometrics; validation studies

Introduction

Palliative cancer patients experience a complex configuration of symptoms associated with advancing disease. Approximately 60–80% of patients will experience pain before death.1 Other debilitating physical symptoms, including anorexia, nausea, asthenia, dyspnoea and delirium, occur with similar or higher frequencies.2–4 Psychological distress, such as depression or anxiety, is often associated with these debilitating symptoms.5–10 Up to 30% of patients will experience an adjustment disorder,9 while 10–25% will develop a major depressive episode.7

To address these complex symptom experiences, Bruera and colleagues developed the Edmonton Symptom Assessment System (ESAS), a brief and clinically useful bedside tool for self-reporting symptom intensity by advanced cancer patients.11 The ESAS was designed to enable repeated quantitative measurements of symptom intensity with minimal patient burden. It includes nine common symptoms of advanced cancer (pain, tiredness, nausea, depression, anxiety, drowsiness, appetite, wellbeing, shortness of breath), with the option of adding a 10th, patient-specific symptom. In the original version of the ESAS, patients rated the intensity of eight symptoms, using visual analog scales, by placing a mark along a line ranging from 0 (no symptom) to 100 mm (worst possible symptom).11 In a subsequent version, Bruera, et al. added a ninth symptom (shortness of breath), as well as the option of rating a 10th symptom.11 More recently, the visual analog scales have been replaced with 11-point numerical rating scales, with higher scores representing worse symptom intensity.12

Since its inception in 1991, the ESAS has been adopted by palliative care programs nationally, across Canada, and internationally, for clinical, administrative and research purposes.13–15 Despite this overwhelming endorsement, the systematic validation of this assessment tool has lagged behind its widespread use. In addition, recent concerns have been raised about the use of the ESAS, particularly in clinical practice. A survey of palliative care nurses’ perceptions of the feasibility and usefulness of the ESAS identified concerns about the potential for patients to misunderstand the tool when self-reporting symptom intensity.16 Garayali, et al. also identified concerns about potential errors in patient self-assessments of symptom intensity using the ESAS.17 Studies to further evaluate the psychometric properties of the ESAS, with the ultimate aim of improving accuracy (reliability) and truthfulness (validity), are warranted.

The purpose of this review was to identify and critique validation studies focusing on the ESAS, published over a
15-year period (1991–2006). Based on this review, directions for further validation research were also identified.

Methods

Overview

For the purpose of this review, we adopted the following definitions for reliability and validity:

(a) **reliability** – consistency (reproducibility) of test scores;\(^{18}\)
(b) **validity** – an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness of inferences* and *actions* based on test scores or other modes of assessment.\(^ {19}\)

To minimize bias, we applied the principles for conducting a systematic review\(^ {20}\) by using clear objectives, a comprehensive search strategy, explicit inclusion/exclusion criteria and blinded independent reviews of publications by three individuals. Traditional systematic reviews focus on *intervention* studies, which include an assessment of the quality of included studies. To our knowledge, there are no standardized assessments for evaluating the quality of validation studies in instrument development. Thus, for the purposes of this review, we used the measurement principles of reliability outlined by Crocker and Algina,\(^ {18}\) and a framework for gathering validity evidence proposed by Messick,\(^ {19}\) as an overarching template to critique articles and identify future research directions (Table 1).

Search strategy

The articles evaluated in this report were identified through a comprehensive search of six electronic databases: MEDLINE (1966–December 2006), CINAHL (1937–December 2006), PubMed (1950–December 2006), HealthStar (1966–December 2006), Science Direct and EMBASE (1988–December 2006). The following search terms were used: ‘ESAS’, ‘Edmonton Symptom Assessment Scale’, ‘Edmonton Symptom Assessment System’, ‘Edmonton Symptom Assessment Tool’, and ‘Edmonton Symptom Assessment’. These terms were typed into an advanced keyword search. Reference lists of relevant articles were also reviewed, where appropriate.

Inclusion and exclusion criteria

The following inclusion criteria were used for article selection: research articles with a primary focus on gathering reliability and/or validity evidence for the ESAS; published in peer reviewed and English language periodicals; year of publication from January 1991 to December 2006. Exclusion criteria included review articles, articles published in a language other than English, research abstracts (eg, poster and oral conference presentation abstracts) and research studies, which included the ESAS but did not have a key focus of ESAS validation.

Data extraction and collection

All articles identified through the search strategy were retrieved. Each article was summarized in a database, using the following categories: publication date, author(s), population/sample, study design, types of tools used, study purpose and relevant outcomes. The three authors initially screened all articles independently, by reviewing the title and abstract of each article, as well as the database summary. Where appropriate, the texts of the articles were also reviewed. Each author independently created a list of potential articles for in-depth review, based on the inclusion criteria. The three authors met on two separate occasions to review the screening process. If all three authors were in agreement regarding the inclusion of a particular article, then no further discussion was required. In cases where there were differences in opinions regarding whether or not to include a particular article, further discussion ensued, until 100% consensus was achieved. The three authors independently conducted

<table>
<thead>
<tr>
<th>Source</th>
<th>Approach for gathering evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content relevance and representativeness</td>
<td>Judgmental and logical analyses (eg, use of expert panels, Delphi technique)</td>
</tr>
<tr>
<td>Structure and variance components both internally and externally</td>
<td>Correlational studies (eg, factor analysis, correlations with other variables)</td>
</tr>
<tr>
<td>Processes underlying item responses and task performance</td>
<td>Cognitive probing (eg, think aloud protocol analysis)</td>
</tr>
<tr>
<td>Differences in test processes and structures over time or across groups and settings</td>
<td>Longitudinal studies; group and/or setting comparisons</td>
</tr>
<tr>
<td>Appropriateness of alterations in instrument scores in response to interventions</td>
<td>Intervention studies</td>
</tr>
<tr>
<td>Value implications and social consequences of interpreting and using test scores</td>
<td>Consequential validity methods</td>
</tr>
<tr>
<td>Year</td>
<td>Author</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>1991</td>
<td>Bruera, et al.</td>
</tr>
<tr>
<td>1993</td>
<td>Bruera, et al.</td>
</tr>
<tr>
<td>1998</td>
<td>Philip, et al.</td>
</tr>
<tr>
<td>1999a</td>
<td>Nekolaichuk, et al.</td>
</tr>
<tr>
<td>1999b</td>
<td>Nekolaichuk, et al.</td>
</tr>
<tr>
<td>2000</td>
<td>Chang, et al.</td>
</tr>
<tr>
<td>2002</td>
<td>Stromgren, et al.</td>
</tr>
<tr>
<td>2003</td>
<td>Pautex, et al.</td>
</tr>
<tr>
<td>2006a</td>
<td>Davison, et al.</td>
</tr>
<tr>
<td>2006b</td>
<td>Davison, et al.</td>
</tr>
<tr>
<td>2006</td>
<td>Garyali, et al.</td>
</tr>
<tr>
<td>2006</td>
<td>Moro, et al.</td>
</tr>
<tr>
<td>2006</td>
<td>Vignaroli, et al.</td>
</tr>
</tbody>
</table>

HRQL, health-related quality of life; HADS, Hospital Anxiety and Depression Scale; PCU, palliative care unit; STAS, Support Team Assessment Schedule.
in-depth reviews of 13 articles,11,17,21–31 which were then summarized in table format (Tables 2 and 3).

Results

A total of 85 articles were identified through the comprehensive online database search. Two additional articles were retrieved through checking reference lists of relevant papers, resulting in a total of 87 articles (see Appendix). Using a comprehensive screening process, 74 articles were excluded for the following reasons: gathering ESAS reliability and validity evidence was not a primary focus of study (n = 63), review article (n = 6), minimal or no reference to the ESAS (n = 4), and poster presentation abstract (n = 1). This resulted in a total of 13 articles, which were selected for in-depth review (Figure 1).11,17,21–31

A detailed summary of the articles selected for in-depth review appears in Tables 2 and 3. As shown in Table 2, the majority of studies (n = 11) involved cancer inpatients and/or outpatients, in palliative care settings (n = 9), a geriatric hospital (n = 1) and a medical oncology unit (n = 1). Two studies were conducted in non-cancer settings, involving patients receiving renal dialysis. The highest frequency of studies was in Canada (n = 6), followed by the

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>ESAS format</th>
<th>Other measures</th>
<th>Psychometric evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>Bruera, et al.</td>
<td>VAS 9</td>
<td>English</td>
<td>yes&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>1993</td>
<td>Bruera, et al.</td>
<td>VAS 9</td>
<td>English</td>
<td>no</td>
</tr>
<tr>
<td>1998</td>
<td>Philip, et al.</td>
<td>VAS 9</td>
<td>English</td>
<td>no</td>
</tr>
<tr>
<td>1998a</td>
<td>Nekolaichuk, et al.</td>
<td>VAS 9</td>
<td>English</td>
<td>no</td>
</tr>
<tr>
<td>1998b</td>
<td>Nekolaichuk, et al.</td>
<td>VAS 9</td>
<td>English</td>
<td>no</td>
</tr>
<tr>
<td>2000</td>
<td>Chang, et al.</td>
<td>VAS 9</td>
<td>English</td>
<td>no</td>
</tr>
<tr>
<td>2002</td>
<td>Stromgren, et al.</td>
<td>VAS 9</td>
<td>not specified</td>
<td>yes&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>2003</td>
<td>Pautex, et al.</td>
<td>NRS/VAS 9</td>
<td>French</td>
<td>yes&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>2006a</td>
<td>Davison, et al.</td>
<td>NRS/VAS 10</td>
<td>English</td>
<td>yes&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>2006b</td>
<td>Davison, et al.</td>
<td>NRS/VAS 10</td>
<td>English</td>
<td>yes&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>2006</td>
<td>Moro, et al.</td>
<td>NRS 10</td>
<td>Italian</td>
<td>no</td>
</tr>
<tr>
<td>2006</td>
<td>Vignaroli, et al.</td>
<td>NRS 10</td>
<td>English</td>
<td>yes&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

**ESAS Modifications:**

<sup>a</sup>Eight-item version was used for descriptive study, but nine-item version (including shortness of breath) with additional ‘empty VAS’ item for rating other symptoms was also described in the report.

<sup>b</sup>Modified anchor (worst possible to severe), additional symptom (pruritis).

<sup>c</sup>Replaced tiredness with fatigue, additional symptom for sleep, different order (well-being at end), modified anchor (worst possible to worst imaginable), symptom ratings over past 24 hours as opposed to time of assessment.

<sup>d</sup>Replaced tiredness with weakness, added pain relief question at the end.

<sup>e</sup>Not administered to patients.

<sup>f</sup>Replaced tiredness with fatigue, different order (well-being at end), primary focus on depression and anxiety, described scale as 10 items but only reported nine in table.

BPI, Brief Pain Inventory; EORTC QLQ-C30, European Organization for Research and Treatment of Cancer – Quality of Life – Core 30; ESAS, Edmonton Symptom Assessment System; FACT, Functional Assessment Cancer Therapy; HADS, Hospital Anxiety and Depression Scale; HD, haemodialysis; HRQL, health-related quality of life; KPS, Karnofsky Performance Status; KDQOL-SF, Kidney Dialysis Quality of Life Short Form; MMSE, Mini Mental Status Examination; MQOL, McGill Quality of Life Questionnaire; MSAS, Memorial Symptom Assessment Scale; NRS, Numerical Rating Scale; PCU, palliative care unit; POS, Palliative Care Outcome Scale; RSCL, Rotterdam Symptom Check List; SDS, Symptom Distress Scale; STAS, Support Team Assessment Schedule; VAS, visual analog scale.
USA (n = 3), Italy (n = 1), Switzerland (n = 1), Australia (n = 1) and Denmark (n = 1). Most studies (n = 11) were prospective. As shown in Figure 2, the majority of studies were published in the past nine years (n = 11), with the highest frequency of publications in 2006 (n = 5).

The content and format for the ESAS varied considerably across studies, as modifications were made in terms of scale format, number of items, item selection and language. As shown in Table 3, the scale format included the original visual analog scale (n = 7), a numerical rating scale (n = 3) or a combined visual analog–numerical rating scale (n = 3). The number of items varied from eight to 10, with significant modifications being made to the scale anchors, types of symptoms assessed and symptom order. For example, anchors were changed from ‘worst possible’ to ‘severe’ or ‘worst imaginable’. In two studies, ‘tiredness’ was changed to ‘fatigue’ or ‘weakness’. Other symptoms or questions were added, including pruritis, sleep and pain relief. In two studies, the order of items was changed, in which patients rated ‘well-being’ last. In one study, symptoms were assessed over a 24-hour period rather than at the time of the assessment. The majority of studies used an English version of the ESAS (n = 11); while two studies used a translated version, one in Italian and one in French.

As also shown in Table 3, the types of psychometric evidence gathered varied across studies and included different types of reliability estimates (n = 8), content validity evidence (n = 1), concurrent validity evidence (n = 5), predictive validity evidence (n = 1), and sensitivity and/or specificity (n = 3). Reliability estimates included test–retest, internal consistency, inter-rater and generalizability coefficients. In terms of concurrent validity evidence,
the types of scales used for comparison included quality of life, anxiety, depression, distress, performance status and other symptom assessment measures. None of these studies involved the patient’s perspective as a source of validity evidence.

Discussion

The majority of ESAS validation studies have been conducted in advanced cancer populations, primarily within inpatient settings. Only one study involved cancer patients earlier in the cancer trajectory and two studies used a non-cancer population involving renal dialysis patients. As illustrated by Davison, et al., there may be other symptoms that are more relevant for non-cancer populations, such as pruritis in the case of renal dialysis patients. Further studies, validating the use of the ESAS in cancer patients earlier in the cancer trajectory, in different settings other than inpatients, and in non-cancer populations are warranted.

The international interest in conducting validation studies, as well as the increasing frequency of studies published over the past nine years, serve as an indirect marker for the uptake of the ESAS across countries and cultures. Based on this review, only two studies used a translated version of the ESAS, and only one of these studies focused on translation as a validation concern. Little is known about the cultural differences in terms of patients’ understanding of the symptoms and anchors that are used for the ESAS. Although other translated versions of the ESAS are available, minimal validation research has been conducted in this area. Further research is warranted to better understand patients’ understanding and interpretation of the ESAS across diverse cultures and settings in different countries, as well as to gather validity evidence for versions translated into non-English languages. Given the international interest and uptake of this tool, this research is critical to ensure that the instrument is being used consistently across settings, particularly when comparing findings cross-culturally.

The diversity of ESAS modifications (ie, scale formats, number of items, item selection and order) across studies further reinforces the need for the development of a standardized tool and administration process. Given this current diversity, it is difficult to make cross-study comparisons. When modifications are made to the ESAS, it is important that these changes are explicitly reported. In addition, any modification may potentially influence the validity and interpretation of patient responses. Ideally, validity evidence for these modified formats needs to be gathered, to ensure that the changes are not significantly influencing the interpretation of responses. For example, to our knowledge, there are no published studies comparing the original ESAS version, using visual analog scales, to the numerical rating format. Paice and Cohen did conduct a validation study comparing visual analog and numerical rating scale formats, but this was limited to pain intensity. Similarly, there are no validation studies comparing other modified forms with the original version of the ESAS.

A further ESAS modification has been in the reporting and interpretation of responses. Although most studies used continuous scores (eg, 0–100 for the visual analog; 0–10 for the numerical rating scale) to describe symptom intensity, some studies collapsed the continuous responses into categorical variables. Vignaroli, et al. provide preliminary sensitivity and specificity data using cut-off points for the ESAS symptoms of depression and anxiety. Similarly, some studies used a total symptom distress score for the ESAS as a measure of overall symptom burden, while other studies focused on each symptom independently. Chang, et al. used both individual items and a total ESAS symptom distress score for gathering concurrent validity evidence. Further validation studies are needed to assess potential relationships amongst individual scales and the total symptom distress score.

In terms of reliability, a number of different approaches have been used to obtain reliability estimates. For test–retest reliability, the time frame between assessments varied from one hour to one day to one week. Generally, test–retest reliability estimates are conducted to measure consistency over time, with the intent of obtaining consistently high estimates. Given the dynamic nature of symptom experience in advanced cancer, often due to fluctuating functional status and treatment interventions, which may influence symptom intensity, it is difficult to obtain consistently high estimates for the ESAS over time. Thus, test–retest reliability estimates may not necessarily be appropriate for this population. The use of interrater reliability estimates may be more appropriate, particularly if more than one person (ie, patient, caregiver) may be involved in symptom intensity assessment or if patients are not able to conduct their own symptom assessments. The type of reliability estimate that would be most appropriate is driven by the way in which the instrument will ultimately be used. More research regarding the types of reliability estimates that would be most appropriate for this population, such as reliability estimates of third-party assessments, is warranted.

In terms of validity evidence, the majority of studies focused on gathering concurrent validity evidence, using a wide range of comparison instruments, measuring diverse domains such as quality of life, distress, pain, anxiety, depression, functional status and symptom distress. In some studies, an overall symptom distress score for the ESAS was used to compare with other measures such as quality of life, while other studies compared specific ESAS scales, such as depression or anxiety, with a com-
parison measure for depression and anxiety, the Hospital Anxiety and Depression Scale.24 A significant challenge is to be able to select measures that are appropriately capturing the underlying construct for comparison.

Most of the reliability and validity evidence to date has focused on test–retest reliability and concurrent validity. Minimal validity evidence, focusing on the internal structure of the ESAS, has been gathered. For example, are the nine scales functioning independently of each other or are there some inter-item relationships that may be influencing patient responses? Do some of the items cluster together, potentially forming subscale scores? How well does the total symptom distress score capture symptom burden? Are there some scales, such as well-being, which may be serving as a surrogate marker for an overall score of symptom distress? In addition, there are no known validation studies specifically focusing on the appropriate responsiveness of the ESAS to treatment interventions, an important area for future validation research.

None of the studies reviewed involved the patient’s perspective as a source of validity evidence. One validation strategy for obtaining patient perspectives is the ‘think aloud’ method or protocol analysis, a well-documented approach for gathering validity evidence focusing on patients’ perceptions and underlying cognitive processes.19,34 Using this method, respondents are asked to think aloud during the completion of a cognitive task (eg, completion of the ESAS) or retrospectively describe the thought processes they used. These verbal reports are often transcribed and then analyzed using qualitative approaches, such as discourse analysis.19 The use of think aloud methods would enhance the understanding of the cognitive processes that patients use to complete the ESAS and identify potential sources for misinterpretation and error. This approach would also be appropriate for understanding potential cultural differences in symptom assessment, using the ESAS. We are currently in the process of conducting a ‘think aloud’ validation study, probing the cognitive processes of advanced cancer patients referred to a pain and symptom consultation team in an oncology setting, as they complete the ESAS. Further research is warranted in this area.

Ultimately, the design of validation studies needs to be driven by the way in which the instrument responses will be used. Based on the findings from this review, a number of subsequent psychometric studies could be developed (Table 4). The development of a series of validation studies for the ESAS will strengthen its use in clinical practice and research, further enhancing its credibility as a well-recognized and commonly used standard assessment tool for pain and symptom assessment.

Acknowledgements
The authors would like to acknowledge the support of Viki Muller, Matthew Woynorowski and Linda Harris for assistance with conducting the literature search and compiling the preliminary database of articles. The authors would also like to acknowledge the financial support of the Canadian Institutes for Health Research (CIHR) New Emerging Team Grant for Difficult Pain Problems.

References
3 Bruera, E, Fox, R, Chadwick, S, Brenneis, C, MacDonald, N. Changing pattern in the treatment of

Table 4 Future directions for further validation research for the ESAS

<table>
<thead>
<tr>
<th>Type of evidence</th>
<th>Sources of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>Obtaining reliability estimates based on how the instrument will be used in clinical practice</td>
</tr>
<tr>
<td>Validity</td>
<td>Examining potential differences across different groups (populations) and settings:</td>
</tr>
<tr>
<td></td>
<td>earlier in the cancer trajectory</td>
</tr>
<tr>
<td></td>
<td>outpatient cancer settings</td>
</tr>
<tr>
<td></td>
<td>palliative settings and cultures in different countries</td>
</tr>
<tr>
<td></td>
<td>non-cancer populations and settings</td>
</tr>
<tr>
<td></td>
<td>gathering validity evidence for modified ESAS formats:</td>
</tr>
<tr>
<td></td>
<td>Comparing numerical and visual analog scales of the ESAS by having patients complete both formats of the tool</td>
</tr>
<tr>
<td></td>
<td>Comparing non-English versions with the English version of the ESAS</td>
</tr>
<tr>
<td></td>
<td>Comparing continuous responses with categorical variables (eg, mild, moderate, severe) to identify appropriate cut-off points for these categories</td>
</tr>
<tr>
<td></td>
<td>Examining the internal structure of the ESAS, as well as the external relationships of ESAS responses to other measures</td>
</tr>
<tr>
<td></td>
<td>Probing patients’ underlying cognitive processes for item responses and performance, through methods such as think aloud protocols</td>
</tr>
<tr>
<td></td>
<td>Examining if ESAS responses appropriately change as a result of specific interventions</td>
</tr>
</tbody>
</table>


Appendix


55 Nekolaichuk CL, Maguire TO, Suarez-Almazor M, Rogers WT, Bruera E. Assessing the reliability of


73 Vignaroli E, Pace EA, Willey J, Palmer JL, Zhang T, Bruera E. The Edmonton Symptom Assessment...


