Palliative and Supportive Care (2005), **3**, 87–98. Printed in the USA.
Copyright © 2005 Cambridge University Press 1478-9515/05 $16.00
DOI: 10.1017/S1478951505050169

Assessing symptoms of terminally-ill patients by different raters: A prospective study

SILKE ZLOKLIKOVITS, M.S., ELISABETH ANDRITSCH, B.S., BRIGITTE FRÖHLICH, M.D., JULIJANE VEREBES, M.D., GABRIELE DIETMAIER, M.S., AND HELLMUT SAMONIGG, M.D.

Section of Palliative Care and Division of Clinical Oncology, Department of Internal Medicine, Medical University Graz, Graz, Austria

(RECEIVED December 4, 2004; ACCEPTED January 9, 2005)

ABSTRACT

Objective: For patients who are terminally ill, the efficiency of symptom management is dependent, among other factors, on an accurate assessment by proxy raters. The aim of this prospective study is to describe differences in symptom severity ratings between patients and their nurses, physicians, and family members during their stay in different departments with acute care, and to identify variables associated with the accuracy of the ratings by others.

Methods: Physical, psychological, social, and functional disorders were assessed for 41 in-patients with a standardized 13-item Symptom List for Quality Assurance in Palliative Care drafted by the Working Group on the Core Documentation for Palliative Care Units in Germany. Symptom assessment was completed by different raters (patient, nurse, physician, family member) within the first four days after admission. Socio-demographic, disease-related, and hospitalization data were taken from patient charts.

Results: Reliability of the symptom list was computed with Cronbach's alpha measures for the present sample. Between-group-comparisons on the individual items and on the sum-score level were analyzed separately for the different rater-pairs: patients-nurses (n = 41), patients-physicians (n = 39), patients-family members (n = 12). Multiple regression analyses calculated predictive variables of the staff's deviation scores.

Significance of results: Significant differences for nurses and physicians were found on the sum-score level for psychological and social symptoms, but not for physical and functional symptoms. Family members rated the intensities of the symptoms generally higher than the patients. Suggestions for further analyses are presented and discussed.

KEYWORDS: Symptom assessment, Terminally-ill patients, Self-assessment vs. assessment by others

INTRODUCTION

Patient Assessment

Patients with an advanced and progressive incurable disease suffer in the majority of cases from different symptoms and stressors. Symptom assessment in palliative care provides the basis for achieving an improvement in symptom burden and quality of life of suffering patients. For healthcare providers of palliative patients, it is very important to know as much as possible about the patient's symptom situation, since the accuracy of symptom severity ratings may affect the efficiency of symptom management (Brunelli et al., 1998; Kristjanson et al., 1998; Nekolaichuk et al., 1999*a*; Lobchuk, 2003). For example, an inappropriate interpretation of a patient's symptom expression, such as overestimating pain intensity, may result in an overdose of opioids and toxicity (Nekolaichuk et al., 1999*b*).

Corresponding author: Silke Zloklikovits, M.S., Department of Clinical Oncology and Section of Palliative Care, Medical University Graz, Auenbruggerplatz 15, A-8036 Graz, Austria. E-mail: silke.zloklikovits@klinikum-graz.at

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

Primarily, self-assessments made by patients themselves are used as the "golden standard," but this fact is questionable (Nekolaichuk et al., 1999*a*, 1999*b*). In palliative situations, the question is becoming more and more relevant as to what extent proxy assessments are useful and practicable. If terminally-ill patients are not able to communicate their symptom events due to cognitive impairment, for example, an effective symptom management depends on whose assessments are used as the basis of intervention. It is still mostly uncertain to what extent caregivers such as physicians, nurses, and family caregivers can be appointed proxy raters (Lobchuk et al., 1997; McPherson & Addington-Hall, 2003). Comparisons of patients and proxy symptom assessments can detect the caregiver's ability to perceive and interpret the patient's condition. A discrepancy in the symptom assessment between the terminally-ill patient and the caregiver can be predictive of inadequate symptom management. Research on symptom assessments of different raters will be essential for providing optimal quality of life for palliative patients (Brunelli et al., 1998).

For assessment and management of different disorders, it is necessary to define what specific information is required. In general, symptoms as the patient's subjective experience of a disease are multifaceted events. Symptom occurrence with terms of frequency, duration, and severity, as well as the perceived distress as a separate emotional response to the occurrence of a symptom, are subcomponents of the whole symptom experience. This is a dynamic process of the patient's perception of and response to symptom occurrence and distress (McDaniel & Rhodes, 1995; Rhodes et al., 2000; Lobchuk, 2003).

In our study, we used a unifocal measurement of symptoms and items were analyzed in terms of intensity (characterized as the amount and degree of discomfort with which the symptom occurs).

Rating Assessment

Many studies examining congruence in symptom assessments by different raters have been conducted with different patient populations. It is only since the middle 1990's that patients with an advanced progressive incurable illness such as terminal cancer have been studied. However, comparison studies are scarce due to different measurement tools with diverse symptoms and research questions, different settings with acutely stressed or stabilized patients, and varying populations.

General findings of previous studies on patients in the palliative situation have pointed out that the nature of a symptom influences the degree of accuracy for proxy symptom ratings, with a substantial agreement in the literature regarding more accurate assessments on physical and functional symptoms than on psychological and social symptoms (Brunelli et al., 1998; Sneeuw et al., 1999; Lobchuk & Degner, 2002*b*; Lobchuk, 2003). Physical and functional symptoms have more associated overt behavioral cues with an objective nature, and are therefore perceived more similarly by other persons in comparison to more subjective items like anxiety with more inconsistent assessments (Sprangers & Aaronson, 1992; Nekolaichuk et al., 1999*a*; McPherson & Addington-Hall, 2003). Analyzing total symptom scores only has the disadvantage that differences on the individual symptom level remain undetected. Since information gets lost when averaged across all symptoms, recommendations exist to also analyze individual symptom profiles (Nekolaichuk et al., 1999*b*).

Comparisons of assessments as a function of the rater person illustrate a consistent response bias of family caregivers who report symptom distress, frequency, and severity more negatively than the patients themselves (Lobchuk et al., 1997; Kristjanson et al., 1998; Lobchuk & Degner, 2002*a*). In a Canadian study on health care providers in an acute palliative care unit, the results showed generally lower symptom severity ratings by the physicians and more closely approximated assessments by the nurses (Nekolaichuk et al., 1999*a*). Studies designed to compare self-assessments with professionals and non-professionals within a palliative sample are rarely found in the literature. Nekolaichuk et al. (1999*b*) found that nurses rated better in physical issues and family members better in psychological issues within the same sample.

Additional research questions in this context provide inconsistent outcomes. Results of studies analyzing the influence of the characteristics of patient and proxies on the accuracy of the ratings are very population associated. Brunelli et al. (1998) found that none of the characteristics of the patient or the proxy (nurses and physicians) showed any statistically significant relationship with the two disagreement indices. An Australian study from Kristjanson et al. (1998) with 78 terminal cancer patients showed that symptom distress was reported more accurately from male and younger family caregivers. The literature concerning the relationship between the symptom severity level and accuracy of proxy reports illustrates the tendency that the higher the symptom levels reported by the patients, the less accurate the report of different healthcare providers. The good agreement on the absence of symptoms found in the studies of Brunelli et al. (1998) and Horton (2002) shows a tendency to underreport

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

problems by professionals. No relationship was found in this context concerning family caregivers by Kristjanson et al. (1998).

Our Study

In our present study, assessments from three different proxy raters were compared with the self-assessments for the same patient sample: nurses, physicians, and family members. A symptom list was chosen, including 13 symptoms to target a large spectrum of aspects (physical, psychological, cognitive, social, functional). Different symptoms were compiled from a German Study Group for a documentation system in palliative medicine (Radbruch et al., 2000*a*, 2000*b*, 2003) which is still in development and only validated in parts. In comparison to the literature on symptomatology and prevalence data (Mercadante et al., 2000; Strömgren et al., 2002; Potter et al., 2003), the symptom list provides a good overview on the different aspects as most measurement tools are dominated from somatic problems (Radbruch et al., 2003). The questionnaire proved to be practicable and short for the German sample of palliative patients. In this pilot investigation, we wanted to test practicability on self and proxy assessments for an Austrian sample and analyze if the items are able to differentiate. Further analyses for the Austrian sample of palliative patients, as well as their different caregivers, are still in the research stage.

The purposes of this prospective study are:

1. to identify differences in symptom severity ratings between patients and their nurses, physicians and family members in an acute hospital setting
2. to determine whether different patient and proxy variables, as well as the level of symptom severity reported by the patients, are associated with the accuracy of the symptom severity ratings of other raters (nurses, physicians, family members).

METHODS

Sample

Over a 6-month period in 2001, a total of 56 terminally-ill patients were recruited consecutively on admission from five different departments—Dermatology, Urology, Ear, Nose and Throat, Neurology and Surgery—of the University Hospital Graz, Austria. The patients had to meet the following criteria for inclusion in the study: (1) patients with an advanced and progressive chronic incurable illness and no existing potentially curable treatment, (2) patients with a Karnofsky Performance Status (KPS) score (Karnofsky & Burchenal, 1949) between 20 and 50 ($20 \leq x \leq 50$), (3) no terminal patients, (4) no patients with an acute life-threatening illness and (5) patients able to communicate and without signs of limited awareness. Patients were identified as eligible by a physician or the charge nurse and a research assistant then explained the aim of the study to the patients with an information sheet. The patients identified one close family member who is primarily involved in his/her care and regularly visits him/her in the hospital. Both patients and family members were also asked by the research assistant to sign informed consent. For proxy assessments, the nurses and the physicians who were primarily responsible for the patient's care on the day of assessment participated in the study. The study was approved by the ethics committee of the hospital.

Measurement

A Symptom List for Quality Assurance in Palliative Care drafted by the Working Group on the Core Documentation for Palliative Care Units in Germany (M. Fuchs, G. Lindena, F. Nauck, K. Neuwöhner, L. Radbruch, D. Schulenberg; German Cancer Society and the German Association of Palliative Medicine; version of the minimal documentation system for palliative care (MIDOS), Radbruch et al., 2002) was used to assess the severity of 13 symptoms referring to the present day as an instrument for self-assessment by the patients as well as assessment by others. The standardized proforma of a symptom list includes physical, psychological, social, and functional aspects which are considered relevant in palliative situations. Several items were already validated from the German Working Group as the data evaluation is still in process (Radbruch et al., 2000*a*, 2000*b*, 2003).

Twelve of the 13 symptoms were scored on a 4-point-likert-scale with "none" (1), "mild" (2), "moderate" (3), and "severe" (4). A higher score corresponds to a higher severity. The advantage of an easily answered response format has been shown especially in a sample of terminally-ill patients and outbalances the disadvantage of the limited discriminative power (Radbruch et al., 2000*a*, 2000*b*). Maximal pain as the 13th item was assessed with a numerical rating scale from 0 (no pain) to 10 (worst imaginable pain). The symptom list takes approximately 15 minutes to complete (based on the experience of the German Working Group); assistance from the research assistant may be necessary in this specific patient sample.

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

DATA COLLECTION

Assessment of symptoms was completed within the first 4 days after admission to the ward. Two research assistants asked different raters (the patient, nurse, physician and a family member if available) to complete the questionnaire separately and approximately at the same time in the afternoon. They were instructed to evaluate the symptoms in terms of that day's maximal (pain) or mean intensity (other 12 symptoms). Assessments were done between 2:00 and 4:00 in the afternoon in order to evaluate the results and to ensure that the proxies had adequate time to observe the patient. It was estimated that it would take a maximum of two hours for all the raters to complete the questionnaires. Socio-demographic data were obtained from the patients, and disease-related and hospitalization data were taken from patient charts. Information concerning the staff (gender, age, years of experience) and the family members (gender, age, relationship to the patient, living with the patient) was collected by the research assistants.

Statistical Analysis

Descriptive statistics, such as frequencies, percentages, means, SD and median, were used to describe sample characteristics and prevalence data of the symptom assessments. To test the reliability of the symptom list, analyses of the internal consistency with the Cronbach's Alpha were conducted and presented for each rater group (values >.80 were considered as good, values between .60 and .80 as acceptable and values <.60 as moderate).

Symptom intensities were converted to sum-scores for the physical and functional symptoms (= SUM PHY-FUNC) consisting of 7 items: weakness, lack of appetite, nausea, breathlessness, constipation, reduction in daily activities, and reduction in mobility (range 7–28); for the psychological, cognitive, and social symptoms (= SUM PSY-SOC) with 5 items: difficulty with concentration and cognition, anxiety, depression, communication problems, and worrying about future care (range 5–20); and for all 12 symptoms (= SUM TOTAL) with the range 12–48. The pain item was analysed as an individual score (range 0–10).

In order to compare differences between the ratings, nonparametric statistics were used due to ordinal data (Ludbrook, 2002). The Mann-Whitney-U-Test (between-group-analyses; Mann & Whitney, 1947) for two independent samples was calculated to compare patient and nurse rating, patient and physician ratings as well as patient and family member ratings for the individual items and the sum-scores.

Unsigned deviation scores (DS) were computed to describe the different assessments between the rater pairs (patients-nurses and patients-physicians) as follows: $DS = \Sigma_i |x_{ai} - y_{bi}|$. The sum of the absolute values of differences between the two defined raters (a and b) were calculated separately for the items of the sum-score PHY-FUNC (DS PHY-FUNC; range 0–21) and of the sum-score PSY-SOC (DS PSY-SOC; range 0–15). A higher value means a higher deviation from the patient's ratings. A third deviation score (DS PAIN) was calculated for the difference between patients and nurses/physicians maximum pain rating with a range of −10 to +10. A negative value means overestimation of the patient's maximal pain, with higher scores representing higher deviation and positive values underestimation. Zero indicates no difference between the two assessments.

To test the association between the deviation scores and variables concerning the patient and the proxy raters (age and gender of the patient, age, gender and experience of the staff raters, Karnofsky-Index, cancer diagnosis, first or re-admission to the ward, symptom occurrence), stepwise multiple regression analyses were performed. The significance level was constituted for $p < .05$.

RESULTS

Sample Statistics

From a total of 56 eligible patients, 15 patients with sample description data similar to those of the rest of the sample were excluded. Thirteen (23.2%) refused to take part in the study and 2 (3.6%) patients were planned to be transferred to the palliative care unit. The reasons for not agreeing to participate were poor well-being (5 cases), communication problems due to somatic reasons (e.g. tracheostoma) (7 cases) and no specified reason (one patient). Analyses of the initial assessment were conducted with data from 41 patients. Concerning the hospital staff, 41 nurses and 39 physicians completed the symptom questionnaire (the responsible physicians for two patients were not available later on the initial assessment day). Twelve family members agreed to evaluate the patient's symptom severity in the hospital setting.

Table 1 shows the socio-demographic, disease-related, and hospitalisation data of the 41 patients. The 20 (48.8%) female and 21 (51.2%) male patients had a mean age of 65,49 ± 12,33 years (median = 67; range 22–86), 60.9% of the patients were married or lived in common law; 32 (78.0%) did not live alone. The majority (82.9%) of the patients had a diagnosis of cancer, whereas the special sample of-

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

Table 1. *Socio-demographic, disease-related, and hospitalisation data of the patients (n = 41)*

Socio-demographic data	*N*	Percent
Sex		
Male	21	51.2%
Female	20	48.8%
Age		
Mean ± SD: 65.49 ± 12.33		
Median: 67		
Range: 22–86		
Marital status		
Single	4	9.8%
Married	22	53.7%
Common law	3	7.2%
Divorced	1	2.4%
Widowed	11	26.8%
Living status		
Solitarily	9	22.0%
Non solitarily	32	78.0%

Disease-related data	*N*	Percent
Diagnosis		
Cancer (primary tumor)	34	82.9%
mouth and pharynx	4	9.8%
gastrointestinal tract	9	21.9%
respiratory tract	4	9.8%
skin/connective tissue	8	19.5%
male genital tract	7	17.1%
urinary tract	2	4.9%
Noncancer	7	17.1%
neurological diseases	2	4.9%
heart diseases	2	4.9%
cutaneous diseases	3	7.3%
Karnofsky Performance Status (KPS) on admission		
higher KPS (40 and 50)	33	80.5%
lower KPS (20 and 30)	8	19.5%
Mean ± SD: 43.41 ± 9.11		
Median: 50		
Range: 20–50		

Hospitalisation data	*N*	Percent
First admission or re-admission to this ward		
First admission	11	26.8%
Readmission	30	73.2%
Reasons for readmission (multiple answering)		
Pain	22	53.7%
Nutrition	14	34.1%
Physical symptoms	29	70.7%
Psychological problems	6	14.6%
Social problems	4	9.8%
Nursing problems	12	29.3%
Diagnostic measures	3	7.3%
Planned medical treatment	10	24.4%
Other reasons	4	9.8%
Duration of stay (number of days)		
Mean ± SD: 15.32 ± 10.08		
Median: 12		
Range: 3–42		
End of stay		
Discharge home	28	68.3%
Discharge in a nursing home or hospice	2	4.9%
Transferred to another department or hospital	8	19.5%
Deceased	3	7.3%

tumor types depended on the particular participating specialized departments. Thirty (73.2%) patients had stayed on this ward at least once before the current admission. The characteristics of nurses, physicians, and family members who answered the initial symptom assessment are detailed in Table 2. Nurses were mostly female (92.7%), but only 16 (41.0%) of the physicians were women. As expected,

Table 2. *Characteristics of the nurses (n = 41), physicians (n = 39), and family members (n = 12)*

	Nurses		Physicians		Family member	
Characteristics	*N*	Percent	*N*	Percent	*N*	Percent
Sex						
Male	3	7.3%	23	59.0%	5	41.7%
Female	38	92.7%	16	41.0%	7	58.3%
Age	Mean ± SD: 38.95 ± 13.53 Median: 37 Range: 20–60		Mean ± SD: 37.10 ± 4.28 Median: 37 Range: 29–54		Mean ± SD: 57.58 ± 9.78 Median: 57.50 Range: 45–73	
Professional experience in years	Mean ± SD: 18.02 ± 13.20 Median: 17 Range:1–40		Mean ± SD: 6.77 ± 5.19 Median: 5 Range:1–27			
Relationship to the patient						
Partner					8	66.7%
Child					3	25.0%
Parent					1	8.3%
Living status						
With the patient					12	100.0%
Not with the patient					0	0%

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

due to different lengths of education, nurses had the same median age (37 years) but a broader distribution (SD = 13.53 vs. 4.28) than the physicians, whereas physicians had fewer years of professional experience compared to the nurses (median = 5 vs. 17). Family members were mostly partners (66.7%) with a median age of 57.5 years.

Intragroup Comparisons

Results of the present Graz sample reliability analyses for the sum-scores SUM PHY-FUNC (7 items), SUM PSY-SOC (5 items) and the SUM TOTAL (12 items) show good internal consistency for the proxy raters with Cronbach's Alpha coefficients of .76, .86, and .87 for the nurses, .76, .71, and .78 for the physicians and .78, .79, and .84 for the family members. Internal consistency when rated by the patients were moderate for the sum-score SUM PSY-SOC with .57 and acceptable for the sum-score SUM PHY-FUNC (.66) and the sum-score SUM TOTAL (.68).

Intergroup Comparisons

Between-group-comparisons concerning the symptom severity ratings were made separately between each of the three rater-pairs because of different sample sizes. Table 3–5 summarise results such as initial prevalence data and mean values on individual symptoms and sum-scores as reported by the three different rater-pairs patients-nurses ($n = 41$), patients-physicians ($n = 39$), and patients-family members ($n = 12$) as well as p-values of the between group differences (Mann-Whitney-U-Test). In addition, the percentages of cases with concordant ratings (con) as well as the percentages of cases overreporting (>) and underreporting (<) the patient intensity score were shown.

For the 41 patient-nurse dyads (Table 3), the physical-functional symptoms reduction in mobility, reduction in daily activities and weakness as well as pain are reported as most prevalent. Depressiveness as the most prevalent psycho-social symptom occurred in 63.4% when self-assessed and

Table 3. *Prevalence data, mean values, and percentages of concordance by the patients compared to the nurses (n = 41)*

	Patients				Nurses							
	Prevalence				Prevalence							
Symptoms	*N*	%	Mean	SD	*N*	%	Mean	SD	*p**	Con	>	<
Maximal pain (0–10)	29	70.7	3.24	2.96	35	85.4	3.12	2.56	.974	36.6[1]	31.7	31.7
Physical-functional:												
Weakness	31	75.6	2.76	1.22	41	100.0	3.05	0.83	.382	39.0[2]	36.6	24.4
Lack of appetite	20	48.8	2.17	1.34	30	73.2	2.29	1.08	.387	34.1[2]	41.5	24.4
Nausea	15	36.6	1.76	1.11	19	46.3	1.71	0.92	.762	51.2[2]	26.8	22.0
Breathlessness	12	29.3	1.49	0.84	8	19.5	1.29	0.64	.277	68.3[2]	7.3	24.4
Constipation	14	34.1	1.85	1.28	22	53.7	2.02	1.11	.306	43.9[2]	29.3	26.8
Reduction in daily activities	29	70.7	2.61	1.26	33	80.5	2.85	1.20	.371	48.8[2]	34.1	17.1
Reduction of mobility	32	78.0	2.80	1.21	34	82.9	2.78	1.51	.911	39.0[2]	29.3	31.7
Psycho-social:												
Difficulty concentrating	10	24.4	1.37	0.69	19	46.3	1.68	0.85	.050	53.7[2]	36.6	9.7
Depressiveness	26	63.4	2.02	1.04	31	75.6	2.34	1.06	.152	26.8[2]	46.4	26.8
Anxiety	12	29.3	1.51[a]	0.92	25	61.0	1.95[a]	0.86	.009[a]	29.3[2]	48.8	21.9
Communication problems	1	2.4	1.02[a]	0.16	17	41.5	1.56[a]	0.81	.000[a]	58.5[2]	41.5	0.0
Worrying about future care	5	12.2	1.29[a]	0.81	16	39.0	1.73[a]	1.07	.010[a]	61.0[2]	31.7	7.3
SUM phy-func (7)			15.44	4.80			16.00	4.52	.518			
SUM psy-soc (5)			7.22[a]	2.38			9.27[a]	3.77	.009[a]			
SUM total (12)			22.66	5.90			25.27	7.54	.159			

N = number of cases
SD = standard deviation
* = Mann-Whitney-U-Test, p-value; [a] = significantly different, $p < .05$
con = percentage of concordance, >/< = percentage of overestimating/underestimating
[1]The percentage of concordance were calculated with all cases of pairs who rated the same score or within one score higher or lower ($-1 < x < +1$)
[2]The percentage of concordance were calculated with only cases of pairs who rated the same score

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

Table 4. *Prevalence data, mean values, and percentages of concordance by the patients compared to the physicians (n = 39)*

	Patients				Physicians							
	Prevalence				Prevalence							
Symptoms	*N*	%	Mean	SD	*N*	%	Mean	SD	*p**	Con	>	<
Maximal pain (0–10)	28	71.8	3.28	2.98	32	82.1	2.62	2.33	.492	53.8[1]	15.4	30.8
Physical-functional:												
Weakness	29	74.4	2.72	1.23	37	94.9	3.08	0.93	.305	46.2[2]	33.3	20.5
Lack of appetite	20	51.3	2.23	1.35	27	69.2	2.31	1.03	.547	20.5[2]	38.5	41.0
Nausea	14	35.9	1.77	1.13	17	43.6	1.82	1.07	.666	48.8[2]	25.6	25.6
Breathlessness	12	30.8	1.51	0.85	9	23.1	1.28	0.60	.369	59.0[2]	15.4	25.6
Constipation	13	33.3	1.85	1.29	20	51.3	1.90	1.09	.416	43.6[2]	30.8	25.6
Reduction in daily activities	28	71.8	2.62	1.25	33	84.6	2.90	1.12	.319	46.2[2]	38.5	15.3
Reduction of mobility	30	76.9	2.79	1.22	33	84.6	2.90	1.12	.781	53.8[2]	25.5	20.5
Psycho-social:												
Difficulty concentrating	9	23.1	1.33[a]	0.66	22	56.3	1.82[a]	0.85	.006[a]	43.6[2]	46.2	10.2
Depressiveness	25	64.1	2.05[a]	1.05	33	84.6	2.44[a]	0.91	.035[a]	41.0[2]	43.6	15.4
Anxiety	10	25.6	1.49[a]	0.94	28	71.8	2.08[a]	0.87	.009[a]	17.9[2]	61.5	20.6
Communication problems	1	2.6	1.03[a]	0.16	16	41.0	1.59[a]	0.82	.000[a]	56.4[2]	41.0	2.6
Worrying about future care	5	12.8	1.31[a]	0.83	20	51.3	2.05[a]	1.15	.010[a]	46.2[2]	46.2	7.6
SUM phy-func (7)			15.49	4.85			16.18	4.55	.385			
SUM psy-soc (5)			7.21[a]	2.42			9.97[a]	3.15	.000[a]			
SUM total (12)			22.69[a]	5.94			26.15[a]	6.31	.018[a]			

N = number of cases
SD = standard deviation
* = Mann-Whitney-U-Test, *p*-value; [a] = significantly different, $p < .05$
con = percentage of concordance, >/< = percentage of overestimating/underestimating
[1]The percentage of concordance were calculated with all cases of pairs who rated the same score or within one score higher or lower ($-1 < x < +1$)
[2]The percentage of concordance were calculated with only cases of pairs who rated the same score

75.6% when assessed by nurses. Breathlessness appeared to be the only symptom which was scored higher in prevalence by the patients (29.3%) than by the nursing staff (19.3%). Between-group-analyses for patients and nurses show significantly higher intensity scores for the psycho-social sum-score (SUM PSY-SOC; mean = 7.22 vs. 9.27, $p = .009$) by nurses, but no significant results for maximal pain, the physical-functional and the total sum-score. On the individual symptom level, anxiety, communication problems and worrying about future care were significantly overestimated by the nurses. Percentages of concordance ranged between 34.1% and 68.3% for SUM PHY-FUNC and 26.8% and 61.0% for the SUM PSY-SOC, respectively, at 36.6% for the maximal pain rating. Accurate ratings were highest for breathlessness (68.3%), and lowest for the items depressiveness (26.8%), and anxiety (29.3%). For all symptoms except breathlessness (7.3% vs. 24.4%), reduction in mobility (29.3% vs. 31.7%) and maximal pain (31.7% vs. 31.7%), more nurses overestimated than underestimated a symptom.

Analyses of the 39 patient-physician pairs demonstrate findings similar to those of the nursing staff (see Table 4). Physicians evaluated the same symptoms as most prevalent than patients themselves and nurses: pain, weakness, reduction in mobility, reduction in daily activities, and depressiveness. Except breathlessness, they assessed all individual symptoms as more prevalent than the patients. A statistically significant difference in the sum-scores was found for the psycho-social sum-score (SUM PSY-SOC; mean = 7.21 vs. 9.97, $p = .000$) and the total sum-score (SUM TOTAL; mean = 22.69 vs. 26.15, $p = .018$) as well as for all five psycho-social single items where physicians overestimated the symptom intensities. Physicians' intensity ratings concerning pain as well as physical and functional items demonstrate no significant differences compared to the self-assessments by the patients. Depending upon the symptoms, the intensity ratings between patients and physicians were concordant between 43.6% and 59.0% for the physical-functional items except lack of appetite in 20.5% of cases. The psycho-social symptoms were

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

Table 5. *Prevalence data, mean values, and percentages of concordance by the patients compared to the family members (n = 12)*

	Patients				Family members							
	Prevalence				Prevalence							
Symptoms	*N*	%	Mean	SD	*N*	%	Mean	SD	*p**	Con	>	<
Maximal pain (0–10)	9	75.0	3.25	3.05	9	75.0	3.83	3.24	.705	67.7[1]	16.7	16.7
Physical-functional:												
Weakness	11	91.7	3.17	1.03	11	91.7	3.42	1.01	.082	58.3[2]	33.3	8.4
Lack of appetite	7	58.3	2.33	1.30	8	67.7	2.75	1.36	.233	58.3[2]	33.3	8.4
Nausea	7	58.3	2.08	1.08	8	67.7	2.25	1.21	.116	50.0[2]	33.3	16.7
Breathlessness	3	25.0	1.50	0.91	5	41.7	1.92	1.65	.257	75.0[2]	25.0	0.0
Constipation	5	41.7	2.17[a]	1.47	10	83.3	3.17[a]	1.19	.003[a]	50.0[2]	41.7	8.3
Reduction in daily activities	10	83.3	2.67	1.07	10	83.3	3.00	1.21	.338	50.0[2]	33.3	16.7
Reduction of mobility	11	91.7	3.00	0.95	10	83.3	3.08	1.24	.424	50.0[2]	25.0	25.0
Psycho-social:												
Difficulty concentrating	2	16.7	1.17[a]	0.39	6	50.0	2.08[a]	1.24	.041[a]	58.3[2]	41.7	0.0
Depressiveness	8	67.7	2.25[a]	1.22	10	83.3	2.75[a]	1.05	.037[a]	25.0[2]	50.0	25.0
Anxiety	4	33.3	1.58[a]	0.97	8	67.7	2.50[a]	1.31	.009[a]	50.0[2]	50.0	0.0
Communication problems	0	0.0	1.00[a]	0.00	3	25.0	1.33[a]	0.65	.009[a]	75.0[2]	25.0	0.0
Worrying about future care	1	8.3	1.25	0.87	2	16.7	1.50	1.17	.613	91.7[2]	8.3	0.0
SUM phy-func (7)			16.92[a]	4.36			19.58[a]	5.53	.011[a]			
SUM psy-soc (5)			7.25[a]	2.53			10.17[a]	4.09	.024[a]			
SUM total (12)			24.17[a]	6.04			29.75[a]	8.53	.007[a]			

N = number of cases
SD = standard deviation
* = Mann-Whitney-U-Test, p-value; [a] = significantly different, $p < .05$
con = percentage of concordance, >/< = percentage of overestimating/underestimating
[1]The percentage of concordance were calculated with all cases of pairs who rated the same score or within one score higher or lower ($-1 < x < +1$)
[2]The percentage of concordance were calculated with only cases of pairs who rated the same score

assessed accurately between 41.0% and 56.4%, with the exception of anxiety with only 17.9% concordance.

The initial assessments of patients and family members (n=12) demonstrate that relatives assessed nearly all individual symptoms as highly prevalent, especially pain and nine other symptoms with at least 50% prevalence. Concerning all sum-scores, family members rated statistically significantly higher intensities than the patients themselves. For the physical-functional sum-score (SUM PHY-FUNC; mean = 16.92 vs. 19.58, $p = .011$), constipation on the individual level was significantly overestimated by the family ($p = .003$). Family members scored the individual items difficulty concentrating ($p = .041$), depressiveness ($p = .037$), anxiety ($p = .009$), and communication problems ($p = .009$) from the psycho-social sum-score (SUM PSY-SOC; mean = 7.25 vs. 10.17, $p = .024$) significantly higher. For the overall sum-score (SUM TOTAL), significant differences between patients and family members were shown (mean = 24.17 vs. 29.75, $p = .007$). Percentages of concordance show good agreement with more relatives overestimating than underestimating in most cases. Only for depressiveness, less than 50% of family members assessed the symptom intensity accurately.

Regression Analyses

The stepwise multiple regression analyses, with the three deviation scores (DS PHYS-FUNC, DS PSY-SOC, DS PAIN) as dependent variables and selected demographic and disease-related variables as predictors, revealed only a few significant results for both staff rater-pairs (Table 6). For nurses, predictor variables were patient's age and gender, nurse's age and experience in years (gender was not included due to 92.7% percentage of women), diagnosis (cancer vs. non-cancer), Karnofsky Performance Status (KPS, lower vs. higher), first or re-admission to this ward and three symptom occurrence variables (number of prevalent symptoms of the physical-functional (range 0–7) and the psycho-

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

Table 6. *Significant results of the stepwise multiple regression analyses*

	Mean	SD	Median	Range	R^2	Beta	p
Nurses							
DS PHY-FUNC	5.39	2.73	5.00	1/13			
—First or re-admission to this ward					.140	.374	.016
DS PSY-SOC	3.95	2.89	3.00	0/13			
—Karnofsky Performance Status					.306	−.478	.000
—Presence of maximal pain					.432	.363	.006
DS PAIN	0.12	2.66	0.00	−5/+8			
—Presence of maximal pain					.226	.475	.002
Physicians							
DS PHY-FUNC	5.97	2.94	5.00	2/13			
—Gender of patients					.129	.360	.027
DS PSY-SOC	4.41	2.56	4.00	0/10			
—Presence of maximal pain					.132	.364	.025
DS PAIN	0.67	2.60	0.00	−5/+9			
—Presence of maximal pain					.251	.471	.002
—Diagnosis					.341	.301	.036

social sum-score (range 0–5) as well as presence of maximal pain intensity rated by the patients themselves divided by the median intensity value in lower vs. higher) were included. The variable "first or re-admission to the ward" was associated significantly (beta = .374, p = .016) with the deviation score for the physical-functional symptoms (DS PHY-FUNC), accounting for 14% of the variance of the dependent variable. The absolute deviation of symptom intensity ratings was lower for patients who stayed for the first time on this ward (mean = 3,73, SD = 2,37) as for patients who had a re-admission (mean = 6,00, SD = 2.63). For the psycho-social deviation score (DS PSY-SOC), the "KPS" was found significantly related with a beta score of −.553 (p = .000), contributing 30.6% of the variance with higher deviation for patients with a lower KPS (mean = 7.43, SD = 4.04) than for patients with a better functional status (mean = 3.24, SD = 2.02). "Presence of maximal pain" increased the explanation of the variance to 43,2% (beta = .363, p = .006) with a higher mean DS PSY-SOC (mean = 5.30, SD = 3.21) for patients who assessed higher pain scores than for patients with lower ratings (mean = 2.67, SD = 1.83). The "presence of maximal pain" assessed by the patients themselves was also significantly related to the pain deviation score (DS PAIN) (beta = .418, p = .007, R^2 = 0.175), indicating underestimation for higher pain scores (mean = 1.40, SD = 2.98) and overestimation for lower ones (mean = −1.10, SD = 1.58).

Analyses of physicians' deviation scores included the same patient predictor variables as for nurses as well as gender, age, and professional experience in years of the physicians. "Gender of the patients" was found significantly predictive for the physical-functional deviation score (DS PHY-FUNC) with a higher deviation for male patients (mean = 6.95 vs. 4.95; SD = 3.22 vs. 2.27; beta = .360; p = .027) explaining 12.9% of the variance. "Presence of maximal pain" explained 13.2% of the variance for the dependent variable DS PSY-SOC (beta = .364; p = .025) with higher pain scores predicting higher deviation (mean = 5.26 vs. 3.60; SD = 2.62 vs. 2.28). The variables "presence of maximal pain" (beta = .471, p = .002) and "diagnosis" (beta = .301, p = .036) were found to predict significantly the pain deviation score (DS PAIN), explaining 34.1% of the variance. Higher pain scores are associated with higher deviation towards underestimation (mean = 1.95 vs. −0.55, SD = 2.88 vs. 1.57). For cancer patients, physicians' ratings were closer (mean = 0.33, SD = 2.23) than for patients with a non-cancer diagnosis (mean = 2.50, SD = 3.85) whose maximal pain was mostly underestimated.

DISCUSSION

It was difficult to compare the research literature results on proxy rater agreement due to diverse measurement tools with different symptoms, instructions, dimensions, and scaling. Furthermore, no consensus exists on which evaluation tools are most appropriate to use, since no selection of questionnaires can cover all of the patient's needs (Strömgren et al., 2002; Barresi et al., 2003; Radbruch et al., 2003).

The significant findings of the present study can be divided into four major areas: the degree of different assessments concerning pain as a differ-

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

entiated aspect, the physical-functional and the psycho-social aspects as well as the predictive variables for the degree of accurate ratings.

1. For the maximal pain item, we found statistically non-significantly higher prevalence scores by the staff on the one hand and lower mean ratings like Nekolaichuk et al. (1999*a*) and Sprangers & Aaronson (review, 1992) on the other. Nurses gave better maximal pain ratings than physicians, possibly because they spend more time with the patient or routinely ask the patient about his/her symptom burden for chart documentation. The fact that family members scored the maximal pain insignificantly higher than the patients has also been described in other studies (Kristjanson et al., 1998; Lobchuk & Degner, 2002*a*).

2. Results concerning the better congruence for physical-functional symptoms compared to psychological and social items are similar to those confirmed in the literature (Sprangers & Aaronson, 1992; Brunelli et al., 1998; Kristjanson et al., 1998; Lobchuk & Degner, 2002*a*; McPherson & Addington-Hall, 2003; Lobchuk, 2003). The more objective physical-functional items were assessed without a statistically significant difference by nurses and physicians on the individual and the sum-score level. Descriptively, each of these symptoms (except breathlessness) showed a marginal overestimation in prevalence and mean intensity ratings. The discrepancy in breathlessness confirmed by Nekolaichuk et al. (1999*a*) may possibly arise from the fact that patients' ratings are also affected by the anxiety of no longer being able to breathe and of dying. Health care professionals know that dyspnea has a non-imminent life-threatening character and therefore they evaluate the intensity slightly lower. Family members rated the intensity of all physical-functional symptoms higher than patients with statistically significant differences for the sum-score as well as the individual symptom constipation.

3. Psycho-social symptoms, which are more subjective in nature, were generally overestimated by the proxy raters (staff as well as relatives). Other studies also describe differences in comparisons with higher scoring of these symptoms (Sprangers & Aaronson, 1992; Nekolaichuk et al., 1999*b*; Lobchuk & Degner, 2002*a*; McPherson & Addington-Hall, 2003). Evaluations of Nekolaichuk et al. (1999*a*, 1999*b*) also showed inconsistent ratings for psychological aspects (such as anxiety and depression) by health care professionals but more consistent overestimation by family caregivers. Current data describing percentages of over- and underestimation partly confirm inconsistent intensity ratings by nurses and physicians for psychosocial symptoms. For anxiety and depression, 21.9% and 26.8% of nurses and 20.6% and 15.4% of physicians scored the intensity lower than patients, whereas between-group-analyses showed an overestimation by staff raters. Special patient assessments like anxiety were not supported by other raters and it could be postulated that symptoms may mean different things to different individuals (Lobchuk, 2003) which could be a topic for future research. Particularly for these kinds of symptoms, it seems problematic to accept patients' assessments as a "golden standard" since different mechanisms like repression, social desirability, intention to protect others—primarily their relatives against distress—may affect the self-ratings (Faller et al., 1995; Addington-Hall & Kalra, 2001). The present sample includes hospital staff that does not specialize in treating palliative patients within a bio-psycho-social treatment concept. They may presume that high intensity levels are natural when suffering from a terminal illness.

4. The results regarding variables associated with the deviation scores for socio-demographic variables of patients and proxies partly correspond with other literature results which were often inconsistent. The functional status of the patients seams to influence the perception of others particularly regarding psychosocial aspects. Concerning nurses' assessments, a lower KPS score of the patients is associated with a higher deviation score for psychological, social, and cognitive symptoms. Horton (2002) and McPherson & Addington-Hall (2003) also found less agreement since patients' disabilities increase. Further results show a higher deviation for male and re-admitted patients, patients with a non-cancer diagnosis, as well as patients with higher pain scores. It is feasible that patients who experience pain are evaluated by other raters with higher intensity levels for other symptoms because of their pain. This is an interesting subject for further analyses.

It is of particular importance to consider the present results within the limitations of the study. Small sample sizes are not unusual in studies with terminally-ill patients since the severity of the disease can cause unexpected events. Due to the particularly small number of family members and the small cell numbers of some predictor variables of the regression model in our study, it is necessary to interpret the results carefully. However, the preliminary findings confirm that further research questions could be more explicitly examined. The relatively low results with the alpha measures (Cronbach's alpha) for the patients sample were due to the pilot nature of the study.

One may also assume that the study was limited since the health care professionals knew about the aim of the study. However, we would expect no

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

significant differences since persons under observation tend to try harder (Hawthorne effect) (Brunelli et al., 1998). Parts of the concordant assessments as well as the overestimations may be due to the opportunity to assess the symptom situation (especially psychological symptoms) and from expectations that the terminally-ill must be distressed when facing death. Never asking about psycho-social aspects may also result in underestimating symptom intensities.

Caution must be taken when accepting the patients' ratings as the "golden standard" (Lobchuk et al., 1997; Nekolaichuk et al., 1999*b*). Adaptation processes, social desirability, and protection behavior could explain the low prevalence score of anxiety (25.6%) by the patients as well as other factors discussed elsewhere. Obvious low prevalence data also appeared for communication problems and worrying about future care, which calls for further discussion about additional validations of the symptom list despite acceptable reliability scores.

CONCLUSION

Further research based on the present results can be recommended. Advanced evaluations should be conducted with complete sample sizes for all different raters. We would suggest collecting data a second time because of the possible variations in time of the concordant ratings. Another interesting aspect would be to measure whether the staff's qualifications in palliative medicine make a difference. In general, future studies should deal with factors which could be associated with the discrepancies of symptom assessment—e.g. subjective concepts of illness and beliefs, different coping mechanisms, one's own expectations and experience, one's own distress and burden, as well as different instructions and perspectives (Addington-Hall & Kalra, 2001; McPherson & Addington-Hall, 2003).

In summarizing the outcomes, results should always be considered as a function of the rater as well as the different symptoms. We are able to accept general overestimation for family members for all of the different symptom intensities. Assessments made by the staff are generally inconsistent and the accuracy is dependent on the type of symptom. Based on the results of our preliminary research and a review of other studies (Kristjanson et al., 1998; Nekolaichuk et al., 1999*b*), we would recommend an analysis of the differences on the individual symptom level and on the sum-scores. Further analyses would be necessary to detect bias-causing factors related to symptom assessments of different raters.

Accurate assessment of symptomatology and the consequently appropriate treatment/care make an essential contribution to the patient's quality of life and comfort; they also enhance the trust that patients have in medical and nursing staff. On account of this, the importance of research in the area of self-perception and perception of others appears to be evident.

ACKNOWLEDGMENTS

We express our gratitude to the nurses and physicians of the participating units (University Hospital, LKH Graz) for their cooperation in patient recruitment and support in data assessment, Gerhild Krenn for her assistance in data collection and especially all patients and their family members for their willingness to participate in our study. We are especially grateful for the support and the helpful comments on earlier versions of the paper from Prof. Dr. Lea Baider, Hadassah University Hospital, Jerusalem.

REFERENCES

Addington-Hall, J. & Kalra, L. (2001). Measuring quality of life: Who should measure quality of life? *British Medical Journal*, *322*, 1417–1420.

Barresi, M.J., Shadbolt, B., Byrne, D., et al. (2003). The development of the Canberra symptom scorecard: A tool to monitor the physical symptoms of patients with advanced tumours. *BMC Cancer*, *3*, 32.

Brunelli, C., Constantini, M., Di Giulio, P., et al. (1998). Quality-of-Life Evaluation: When Do Terminal Cancer Patients and Health-Care Providers Agree? *Journal of Pain and Symptom Management*, *15*(3), 151–158.

Faller, H., Lang, H., & Schilling, H. (1995). Emotional distress and hope in lung cancer patients, as perceived by patients, relatives, physicians, nurses and interviewers. *Psycho-Oncology*, *4*, 21–31.

Horton, R. (2002). Differences in assessment of symptoms and quality of life between patients with advanced cancer and their specialist palliative care nurses in a home care setting. *Palliative Medicine*, *16*, 488–494.

Karnofsky, D.A. & Burchenal, J.H. (1949). The clinical evaluation of chemotherapeutic agents in cancer. In *Evaluation of chemotherapeutic agents*, MacLeod, C.M. (ed.), pp. 191–205. Symposium, Microbiology Section, New York Academy of Medicine, New York: Columbia University Press.

Kristjanson, L.J., Nikoletti, S., Porock, D., et al. (1998). Congruence between patients' and family caregivers' perceptions of symptom distress in patients with terminal cancer. *Journal of Palliative Care*, *14*(3), 24–32.

Lobchuk, M.M. (2003). The Memorial Symptom Assessment Scale: Modified for use in understanding family caregivers' perceptions of cancer patients' symptom experience. *Journal of Pain and Symptom Management*, *26*, 644–654.

Lobchuk, M.M. & Degner, L.F. (2002*a*). Symptom experiences: Perceptual accuracy between advanced-stage cancer patients and family caregivers in the home care setting. *Journal of Clinical Oncology*, *20*(16), 3495–3507.

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press

Lobchuk, M.M. & Degner, L.F. (2002*b*). Patients with cancer and next-of-kin response comparability on physical and psychological symptom well-being. Trends and Measurement Issues. *Cancer Nursing*, *25*(5), 358–374.

Lobchuk, M.M., Kristjanson, L., Degner, L., et al. (1997). Perceptions of symptom distress in lung cancer patients: I. Congruence between patients and primary family caregivers. *Journal of Pain and Symptom Management*, *14*(3), 136–146.

Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clinical and Experimental Pharmacology and Physiology*, *29*, 527–536.

Mann, H.R. & Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50–60.

McDaniel, R.W. & Rhodes, V.A. (1995). Symptom experience. *Seminars in Oncology Nursing*, *11*(4), 232–234.

McPherson, C.J. & Addington-Hall, J.M. (2003). Judging the quality of care at the end of life: Can proxies provide reliable information? *Social, Science & Medicine*, *56*, 95–109.

Mercadante, S., Casaccio, A., & Fulfaro, F. (2000). The course of symptom frequency and intensity in advanced cancer patients followed at home. *Journal of Pain and Symptom Management*, *20*, 104–112.

Nekolaichuk, C.L., Bruera, E., Spachynski, K., et al. (1999*a*). A comparison of patient and proxy symptom assessments in advanced cancer patients. *Palliative Medicine*, *13*, 311–323.

Nekolaichuk, C.L., Maguire, T.O., Suarez-Almazor, M., et al. (1999*b*). Assessing the reliability of patient, nurse and family caregiver symptom ratings in hospitalized advanced cancer patients. *Journal of Clinical Oncology*, *17*(11), 3621–3630.

Potter, J., Hami, F., Bryan, T., et al. (2003). Symptoms in 400 patients referred to palliative care services: Prevalence and patterns. *Palliative Medicine*, *17*, 310–314.

Radbruch, L., Nauck, F., Ostgathe, Ch., et al. (2003). What are the problems in palliative care? Results from a representative survey. *Supportive Care in Cancer*, *11*, 442–451.

Radbruch, L., Nauck, F., Fuchs, M., Neuwöhner, K., Schulenberg, D., Lindena, G., and the Working Group on the Core Documentation for Palliative Care Units in Germany. (2002). What is palliative care in Germany? Results from a representative survey. *Journal of Pain and Symptom Management*, *23*(6), 471–483.

Radbruch, L., Sabatowski, R., Loick, G., et al. (2000*a*). MIDOS. Validierung eines minimalen Dokumentationssystems für die Palliativmedizin. *Schmerz*, *14*, 231–239.

Radbruch, L., Sabatowski, G.L., Jonen-Thielemann, I., et al. (2000*b*). Cognitive impairment and its influence on pain and symptom assessment in a palliative care unit: Development of a Minimal Documentation System. *Palliative Medicine*, *14*, 266–276.

Rhodes, V.A., McDaniel, R.W., Homan, S.S., et al. (2000). An instrument to measure symptom experience. Symptom occurrence and symptom distress. *Cancer Nursing*, *23*(1), 49–54.

Sneeuw, K.C., Aaronson, N.K., Sprangers, M.A., et al. (1999). Evaluating the quality of life of cancer patients: Assessments by patients, significant others, physicians and nurses. *British Journal of Cancer*, *81*(1), 87–94.

Sprangers, M.A. & Aaronson, N.K. (1992). The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *Journal of Clinical Epidemiology*, *45*, 743–760.

Strömgren, A.S., Groenvold, M., Pedersen, L., et al. (2002). Symptomatology of cancer patients in palliative care: Content validation of self-assessment questionnaires against medical records. *European Journal of Cancer*, *38*, 788–794.

https://doi.org/10.1017/S1478951505050169 Published online by Cambridge University Press