Original Article

Symptom Intensity of Hospice Patients: A Longitudinal Analysis of Concordance Between Patients' and Nurses' Outcomes

Everlien de Graaf, RN, MSc, Daniëlle Zweers, RN, MSc, Alexander de Graeff, MD, PhD, Rebecca K. Stellato, MSc, and Saskia C.C.M. Teunissen, RN, PhD

Department of General Practice Center of Expertise in Palliative Care (E.d.G., D.Z., S.C.C.M.T.), Julius Center for Health Sciences and Primary Care, Utrecht University Medical Center Utrecht, Utrecht; Department of Medical Oncology (A.d.G.), Cancer Center University Medical Center Utrecht, Utrecht; and Department of Biostatistics and Research Support (R.K.S.), Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract

Context. Nearing death, hospice patients are increasingly unable or unwilling to self-report their symptom intensity and rely on nurses' assessments.

Objectives. We hypothesized that concordance between patients' and nurses' assessments of symptom intensity improves over time.

Method. A prospective longitudinal study was conducted from January 2012 to June 2015 using dyads of patient- and nurse-reported outcome measures, collected in daily hospice practice in the first three weeks after admission. Main outcomes were symptom intensity and well-being, measured using the Utrecht Symptom Diary (USD) and USD-Professional. Absolute concordance was the proportion of dyads with no difference in scores between USD and USD-Professional per week after admission. For agreement beyond chance, the squared weighted Kappa for symptom intensity and the one-way agreement intraclass correlation coefficient for well-being were used.

Results. The most prevalent symptoms, fatigue, dry mouth, and anorexia also had the highest intensity scores assessed by patients and nurses. Symptom intensity was underestimated more frequently than overestimated by the nurses. The absolute concordance was fair to good (35%–69%). Agreement beyond chance was low to fair (0.146–0.539) and the intraclass correlation for well-being was low (0.25–0.28). Absolute concordance and agreement beyond chance did not improve over time.

Conclusion. Concordance between patients' and nurses' assessment of symptom prevalence is good, and both patients and nurses reveal identical symptoms as most and least prevalent and intense. However, nurses tend to underestimate symptom intensity. Concordance between patients and nurses symptom intensity scores is poor and does not improve over time. J Pain Symptom Manage 2018;55:272–281.

Key Words

Hospice, symptom management, symptom intensity, patient-reported outcomes, nurse, concordance

Background

Hospices aim to optimize the quality of life of patients with a short-life expectancy by diminishing physical, psychological, social, and spiritual suffering.[1,2] Patients with advanced diseases suffer from multisymptomatology and complex symptom patterns.[3–6] When admitted to a hospice, patients suffer from a mean of six to eight symptoms concurrently, of which four symptoms are graded as

Address correspondence to: Everlien de Graaf, RN, MSc, Department of General Practice Center of Expertise in Palliative Care, Julius Center for Health Sciences and Primary Care Utrecht University Medical Center, Utrecht, PO Box 85500, Utrecht, 3508 GA, The Netherlands. E-mail: e.degraaf@umcutrecht.nl

Accepted for publication: September 4, 2017.

0885-3924/$ - see front matter
https://doi.org/10.1016/j.jpainsymman.2017.09.005

moderate to severe.[7,8] Therefore, symptom management is a major focus of hospice care. Patient-reported outcome measures are vital in symptom management. Outcomes are used in daily care to assess symptom burden, to monitor symptom burden over time, to evaluate the effect of interventions, and to support communication between patients and health care providers and between members of the multiprofessional team.[9,10] Moreover, patient-reported outcomes are increasingly used to evaluate the quality of services provided and to benchmark between services.[11–13]

Self-assessment is the gold standard for assessing symptom intensity.[14] However, not all hospice patients are able or willing to self-report their symptoms. Patients who have a low-performance status, patients with cognitive impairment, and those who are very old are less able to self-report symptoms.[7]

If patients are unable or unwilling to report their symptom burden, proxy measures are used to assess symptom intensity. Professionals and family caregivers can be the patients' proxy. Although studies are inconclusive, it can be stated that professionals tend to underestimate symptoms and family caregivers tend to overestimate symptoms, in particular psychological symptoms. Furthermore, nurses underestimate less than physicians and family caregivers were closer to the patients' experience than nurses.[15–24] However, most studies were performed in hospitals, during palliative oncology treatment, and none in an inpatient hospice setting.

In many hospices, nurses are available 24/7 for inpatients and are responsible for symptom assessment in daily hospice care.[25] In the hospice where this study was performed, a Dutch-adapted translation of the Edmonton Symptom Assessment System is used to self-report symptom intensity in daily care.[14,26] For patients unable or unwilling to report their symptom intensity, the Utrecht Symptom Diary for healthcare professional (USD-P) was developed in collaboration with hospice nurses, to be used as a proxy measure for assessment of symptoms. The USD-P entails identical items as the USD for patients.

In previous studies, concordance between patient- and nurse-reported outcomes was studied in a clinical setting mostly for patients with active treatment with an unknown life expectancy or survival time and short admission times or treatment contacts. In contrast, patients admitted to hospice have a life expectancy of less than three months, are predominantly 70 years and older, and the median admission time is three weeks.[7,25] Although at admission, a large proportion of patients are able to self-report symptoms, along the illness trajectory, toward death, this ability decreases rapidly. Insight into the concordance between patient and hospice nurses gives information about the reliability of assessment by nurses of symptom intensity of patients unable or unwilling to assess it themselves. This knowledge will help the multiprofessional team to improve symptom management and the quality of life and dying of hospice patients.

We hypothesized that concordance increases during admission because nurses can learn from observing the patients' experience. This learning curve should be established within the first three weeks after admission, in relation to the median hospice admission time of approximately three weeks.[7,25]

The aim of this study was to establish whether concordance between hospice inpatients' self-report symptom burden and nurses' proxy measures increases during the first three weeks of admission.

Methods

A prospective longitudinal study was conducted from January 2012 to June 2015 using dyads of patient and nurse outcome measures, prospectively collected in daily hospice practice. For the report of this study, the STrengthening the Reporting of OBservational studies in Epidemiology statement was used.[27]

Setting and Patients

For this study, patients admitted to a professional-driven seven-bed hospice in the center of the Netherlands were enrolled between January 2012 and June 2015. Patients had to be able and willing to assess their symptom intensity, and dyads of patient self-report and nurse's assessed symptom intensity on the same day had to be available. For patients with more than one dyad per week available, the first dyad per week was selected.

Patient Anonymity and Ethics Approval

Admitted patients were informed by the hospice nurse about the study and their right to decline. Patients were asked consent to use their data for the study. After verbal consent was obtained, written consent was recorded in the patient records. Data were collected from the patient records and anonymized by the principal investigator, using an electronic database, SYMPAL, coding the individual patient data. The principal investigator was able to link data to individual cases. The methods of consent, data collection, and use of the SYMPAL database for research queries were approved by the local medical ethics committee of the University Medical Center Utrecht, The Netherlands (11–113/C).

Outcomes and Measures

Patient Self-report of Symptom Intensity and Well-being

Symptom intensity and well-being were assessed twice a week using the Utrecht Symptom Diary (USD). The

USD is a Dutch-adapted version of the Edmonton Symptom Assessment System, a self-report symptom intensity scale.[28,29] The USD contains 11 symptoms: pain, sleeping disturbance, dry mouth, dysphagia, anorexia, constipation, nausea, dyspnea, fatigue, anxiety, and depressed mood and a one-item well-being measure. All symptoms are assessed using an 11-point numerical scale (0 = no symptom, best possible to 10 = worst intensity, worst possible). The recall period of the USD is now/at this moment. Patients usually completed the USD in the late afternoon.

Patient-assessed symptom prevalence is described as the percentage of patients scoring over 0 on the USD.

Nurses' Assessment of Symptom Intensity and Well-being

The nurse assessment of symptom intensity and well-being was performed by nurses using the USD-P on a daily basis at the end of the day shift. The USD-P is the USD-related proxy assessment tool, entailing the same 11 symptoms. Symptom intensity is measured on a five-point verbal rating scale (0 = no symptom to 4 = overwhelming), in concordance with the Palliative care Outcome Scale.[30] Well-being is measured on a 0−10 scale (0 = best possible −10 = worst possible) in concordance with the USD.

Data Analysis

To study concordance between patient and nurse assessments of symptom intensity, both the absolute concordance and agreement beyond chance were used. Complete concordance is defined as the proportion of dyads with no difference between patient and nurse measures. To compare USD and USD-P scores, the USD scores were categorized. Cutoffs were used to categorize the symptom items of the USD into five categories: none (USD score = 0), mild (USD score 1−3), moderate (USD score 4−6), severe (USD score 7−9), and very severe (USD score = 10).[31,32]

For all symptoms and well-being, the USD score was subtracted from the USD-P score, where 0 indicated absolute concordance. The USD difference was analyzed using descriptive statistics.

Agreement beyond chance is defined as the measure of agreement adjusting for chance, reducing the measure of agreement. Agreement beyond chance was analyzed using the weighted Kappa statistic for the categorical outcomes and the intraclass correlation for the numerical scales. Because the categorized USD and USD-P entail five categories, the squared weighted Kappa was used to correct for the chance of disagreement because of the large number of categories.[33] Kappa value of 0 or lower was considered poor, 0.01−0.2 slight, 0.21−0.4 fair, 0.41−0.6 moderate, 0.61−0.8 substantial, and 0.81−1 almost perfect.[34] The one-way agreement intraclass correlation was calculated for the well-being scores on USD and USD-P because the patients and nurses are from a larger pool of persons and agreement in measures was of interest.[35] Because the Kappa statistic can over- or under-correct the agreement between measurements based on the distribution of responses,[36] the absolute concordance and agreement beyond chance are both presented as well as the distribution of the differences between USD-P and the categorized USD scores.

To study agreement over time, the USD and USD-P differences, the squared weighted Kappa, and the intraclass correlation (ICC) were calculated and described for the first three weeks after admission. In addition, a secondary analysis was performed with dyads from patients with dyads in all three weeks after admission.

Statistical analyses were performed using IBM SPSS 23 for descriptive statistics and Kappa. To calculate the weighted Kappa and intraclass correlation coefficient, R version 3.1.1 (2014−07-10) complemented with irr version 0.84 and psy 1.1 package. The level of significance was set at $\alpha = 0.05$.

Results

In total, 263 patients were admitted in the study period. After selection of the first dyad per week per patient, 295 dyads from 147 unique patients were included, of whom 45 patients had dyads in all three weeks and were included for the secondary analysis (Fig. 1).

The mean age of patients was 69.6 and 58% were women (Table 1). The primary diagnosis of most patients was cancer (88%) and patients were admitted predominantly for last resort (84%) (Table 1).

USD-Ps were completed by 21 nurses. All nurses were women, with a mean age of 44 years. At the start of this study, 15 nurses worked in the hospice for 1.7 (95% CI 0.7−2.8) years on average. Nurses worked

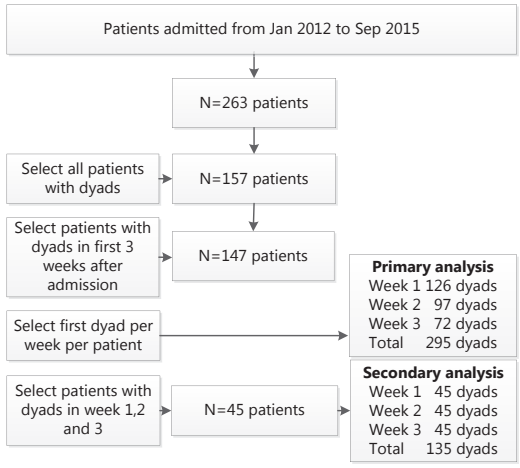

Fig. 1. Flowchart study enrollment.

Table 1
Patient Characteristics

		Patients in the Study ($N = 147$)	
Characteristics	Specified	Mean	SD
Age	Yrs	69.6	13.39
		n	%
Gender	Female	85	58
Primary diagnosis	Cancer	130	88
	Organ failure	9	6
	ALS	2	1
	Other	6	4
Prognosis at admission	<7 d	2	1
	7 d- 4 wk	43	29
	4 wk–3 mo	77	52
	>3 mo	13	9
	Unknown	12	8
Performance score (WHO)	1	2	1
	2	26	18
	3	80	54
	4	34	23
	Unknown	5	4
Marital status	Married/living together	69	47
	Widowed	33	22
	Divorced	11	8
	Unmarried	32	22
	Unknown	2	1
Reason for admission	Last resort	124	84
	Respite	23	16
End admission	Death	123	84
	Transfer to home	18	12
	Transfer to other care setting	6	4
		Median	SD
Admission	Days	24	40.89

three shifts per week on average, mostly consecutive shifts to ensure continuity of care. During the study period, two nurses left the hospice team and six nurses were added to the team.

Symptom Prevalence and Intensity

During the first three weeks after admission, fatigue, dry mouth, and anorexia were the most prevalent symptoms and had the highest mean intensity scores. Nausea, anxiety, and dysphagia were the least prevalent symptoms and depressed mood, anxiety, and nausea had the lowest mean intensity scores (Table 2). The mean well-being scores were 4.5, 4.5, and 3.9 in the first three weeks.

Fatigue, dry mouth, and anorexia were the most prevalent symptoms according to the nurses. In addition, nausea, dysphagia, and anxiety were the least prevalent symptoms. Nurses only scored fatigue as moderate to severe for most patients and dry mouth was scored predominantly as mild. The intensity of anorexia shifted from predominantly mild in the first week to predominantly none in the third week after admission. For all other symptoms, nurses scored none for the majority of patients. In Table 3, the categorized USD scores and USD-P scores are displayed. Well-being increased over time, scoring 3.9, 3.8, and 3.6, respectively.

Concordance

The difference between the USD-P and the categorized USD scores is displayed in Figure 2. The green bars (indicating complete concordance) are shifted to the right, indicating an underestimation of symptom intensity by the nurses.

Table 2
Patient-Reported Symptom Intensity

USD	Week 1, $n = 126$	Week 2, $n = 97$	Week 3, $n = 72$
Pain	2.18	2.20	2.17
Sleeping problems	2.67	2.89	1.86
Dry mouth	4.62	4.16	3.81
Dysphagia	1.71	1.73	1.22
Anorexia	4.58	3.80	2.88
Constipation	3.96	2.82	3.13
Nausea	1.08	1.08	0.85
Dyspnea	2.17	2.05	1.48
Fatigue	6.20	5.49	5.23
Anxiety	1.43	1.38	1.31
Depressed mood	1.79	1.84	1.74

Intensity = Mean USD Score.

Table 3
Patient-Reported and Nurse-Reported Symptom Intensity and Well-being, Absolute Concordance, and Agreement Beyond Chance

	Week 1, $n = 126$		Week 2, $n = 97$		Week 3, $n = 72$	
Symptom	USDcat	USD-P	USDcat	USD-P	USDcat	USD-P
Pain						
None *n* (%)	54 (43)	63 (50)	37 (38)	46 (47)	30 (42)	36 (50)
Mild *n* (%)	42 (33)	41 (33)	39 (40)	33 (34)	22 (31)	23 (32)
Moderate *n* (%)	20 (16)	15 (12)	13 (13)	14 (14)	11 (15)	10 (14)
Severe *n* (%)	8 (6)	7 (6)	8 (8)	4 (4)	8 (11)	1 (1)
Very severe *n* (%)	1 (1)	—		—		1 (1)
Concordance %	50.4		45.7		52.9	
Agreement Kappa 95% CI	0.446	0.26–0.63	0.343	0.19–0.5	0.435	0.23–0.64
Sleeping problems						
None *n* (%)	50 (40)	81 (64)	26 (37)	57 (59)	32 (44)	56 (78)
Mild *n* (%)	30 (24)	27 (21)	20 (21)	22 (23)	22 (31)	13 (18)
Moderate *n* (%)	24 (19)	11 (9)	25 (26)	12 (12)	13 (18)	2 (3)
Severe *n* (%)	16 (13)	4 (3)	15 (16)	4 (4)	3 (4)	1 (1)
Very severe *n* (%)	2 (2)	—	1 (1)	1 (1)	1 (1)	—
Concordance %	46.9		39.6		49.3	
Agreement Kappa 95% CI	0.321	0.17–0.47	0.36	0.21–0.51	0.328	0.12–0.53
Dry mouth						
None *n* (%)	18 (14)	34 (27)	18 (19)	27 (28)	10 (14)	27 (38)
Mild *n* (%)	28 (22)	64 (51)	24 (25)	49 (51)	23 (32)	33 (46)
Moderate *n* (%)	39 (31)	18 (14)	31 (32)	17 (18)	25 (35)	11 (15)
Severe *n* (%)	35 (28)	7 (6)	20 (21)	3 (3)	11 (15)	1 (1)
Very severe *n* (%)	3 (2)	1 (1)	3 (3)	—	2 (3)	—
Concordance %	36.5		35.2		31.9	
Agreement Kappa 95% CI	0.367	0.24–0.49	0.367	0.23–0.5	0.176	0.05–0.3
Dysphagia						
None *n* (%)	77 (61)	93 (74)	65 (67)	77 (79)	44 (61)	60 (83)
Mild *n* (%)	16 (13)	20 (16)	6 (6)	15 (16)	15 (21)	7 (10)
Moderate *n* (%)	17 (14)	6 (5)	14 (14)	4 (4)	10 (14)	4 (6)
Severe *n* (%)	8 (6)	2 (2)	11 (11)	1 (1)	2 (3)	
Very severe *n* (%)	3 (2)	1 (1)		—		
Concordance %	63.1		64.8		63.2	
Agreement Kappa 95% CI	0.484	0.29–0.68	0.414	0.23–0.6	0.214	−0.01–0.44
Anorexia						
None *n* (%)	19 (15)	36 (29)	31 (32)	37 (38)	23 (32)	46 (64)
Mild *n* (%)	30 (24)	52 (41)	12 (12)	27 (28)	21 (29)	17 (24)
Moderate *n* (%)	33 (26)	16 (13)	32 (33)	23 (24)	16 (22)	6 (8)
Severe *n* (%)	28 (22)	12 (10)	14 (14)	6 (6)	9 (13)	2 (3)
Very severe *n* (%)	9 (7)	5 (4)	6 (6)	3 (3)	1 (1)	—
Concordance %	34.5		47.2		44.1	
Agreement Kappa 95% CI	0.457	0.31–0.6	0.592	0.43–0.75	0.366	0.21–0.52
Constipation						
None *n* (%)	31 (25)	66 (52)	34 (35)	65 (67)	18 (25)	50 (69)
Mild *n* (%)	31 (25)	29 (23)	21 (22)	19 (20)	25 (35)	16 (22)
Moderate *n* (%)	29 (23)	14 (11)	27 (28)	6 (6)	16 (22)	3 (4)
Severe *n* (%)	10 (8)	6 (5)	9 (9)	2 (2)	10 (14)	—
Very severe *n* (%)	17 (14)	2 (2)	2 (2)	1 (1)	1 (1)	—
Concordance %	30.8		45.2		38.5	
Agreement Kappa 95% CI	0.233	0.08–0.38	0.29	0.12–0.46	0.198	0.05–0.35
Nausea						
None *n* (%)	87 (69)	97 (77)	67 (69)	79 (81)	48 (67)	62 (86)
Mild *n* (%)	19 (15)	16 (13)	17 (18)	13 (13)	17 (24)	8 (11)
Moderate *n* (%)	14 (11)	9 (7)	7 (7)	5 (5)	4 (6)	2 (3)
Severe *n* (%)	3 (2)	1 (1)	4 (4)	—	1 (1)	—
Very severe *n* (%)		2 (22)		—		—
Concordance %	72.4		67.8		66.2	
Agreement Kappa 95% CI	0.539	0.37–0.71	0.293	0.05–0.53	0.293	0.02–0.57
Dyspnea						
None *n* (%)	70 (56)	80 (64)	51 (53)	70 (72)	41 (57)	54 (75)
Mild *n* (%)	18 (14)	18 (14)	23 (24)	14 (14)	20 (28)	8 (11)
Moderate *n* (%)	18 (14)	17 (14)	11 (11)	8 (8)	6 (8)	6 (8)
Severe *n* (%)	15 (12)	9 (7)	10 (10)	4 (4)	3 (4)	3 (4)
Very severe *n* (%)	3 (2)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
Concordance %	58.1		58.2		62.3	
Agreement Kappa 95% CI	0.648	0.53–0.76	0.524	0.35–0.7	0.609	0.42–0.8

(*Continued*)

Table 3
Continued

	Week 1, $n = 126$		Week 2, $n = 97$		Week 3, $n = 72$	
Symptom	USDcat	USD-P	USDcat	USD-P	USDcat	USD-P
Fatigue						
None *n* (%)	5 (4)	3 (2)	5 (5)	4 (4)	8 (11)	6 (8)
Mild *n* (%)	13 (10)	23 (18)	15 (16)	27 (28)	10 (14)	25 (35)
Moderate *n* (%)	43 (34)	62 (49)	40 (41)	34 (35)	24 (33)	25 (35)
Severe *n* (%)	54 (43)	31 (25)	30 (31)	23 (24)	26 (36)	15 (21)
Very severe *n* (%)	8 (6)	6 (5)	4 (4)	9 (9)	3 (4)	1 (1)
Concordance %	44.4		42.7		39.1	0.42
Agreement Kappa 95% CI	0.478	0.36–0.6	0.307	0.1–0.52		0.24–0.62
Anxiety						
None *n* (%)	79 (63)	97 (77)	59 (61)	74 (76)	45 (63)	59 (82)
Mild *n* (%)	21 (17)	22 (18)	18 (19)	16 (17)	15 (21)	13 (18)
Moderate *n* (%)	14 (11)	6 (5)	8 (8)	6 (6)	5 (7)	—
Severe *n* (%)	6 (5)	—	7 (7)	—	4 (6)	—
Very severe *n* (%)	1 (1)	—		—	1 (1)	—
Concordance %	64.9		64.0		64.7	
Agreement Kappa 95% CI	0.232	0.07–0.4	0.364	0.15–0.58	0.146	−0.02–0.31
Depressed mood						
None *n* (%)	71 (56)	97 (77)	46 (47)	69 (71)	36 (50)	54 (75)
Mild *n* (%)	24 (19)	20 (16)	24 (25)	22 (23)	17 (24)	13 (18)
Moderate *n* (%)	18 (14)	6 (5)	12 (12)	4 (4)	9 (13)	3 (4)
Severe *n* (%)	7 (6)	1 (1)	9 (9)	1 (1)	6 (8)	2 (3)
Very severe *n* (%)	1 (1)	—		1 (1)		—
Concordance %	57.9		47.7		47.0	
Agreement Kappa 95% CI	0.461	0.3–0.62	0.206	0.00–0.41	0.194	0.02–0.37

USD = Utrecht Symptom Diary; USD-P = USD-Professional; Concordance = absolute concordance; Agreement = agreement beyond chance; Kappa = weighted Kappa.

The complete concordance (Table 3) was more than 60% on average for nausea, anxiety, and dysphagia. For dyspnea and depressed mood, complete concordance was more than 50%. The complete concordance was lowest for dry mouth, constipation, and anorexia. The least intense symptoms, nausea, anxiety, and dysphagia, showed the highest complete concordance.

Agreement beyond chance (Table 3) was moderate for dyspnea, fair for anorexia, fatigue, nausea, and pain, and low for anxiety.

Complete Concordance Over Time

The complete concordance for anxiety and dysphagia were more than 60% and stable in the first three weeks after admission. Furthermore, complete concordance for fatigue, nausea, and dry mouth decreased slightly and absolute concordance for dyspnea and constipation increased over the first three weeks after admission.

Agreement Beyond Chance Over Time

The trajectory of weighted Kappa showed a stable pattern over time, during the first three weeks after admission. The secondary analysis of data from the patients with complete data ($n = 45$), neither absolute concordance nor agreement beyond chance showed improvement for any of the symptoms over time (data not shown).

Agreement beyond chance in well-being was poor and stable over time with comparable confidence intervals (ICC Week 1 0.27 [95% CI 0.09–0.44], Week 2 0.28 [95% CI 0.07–0.46], and Week 3 0.25 [95% CI 0.01–0.46]).

The secondary analysis ($n = 45$) showed a higher agreement in Week 3 (ICC 0.30, 95% CI 0.01–0.55) but with a wide confidence interval, confirming that agreement does not improve over time (data not shown).

Discussion

Concordance between patient- and nurse-reported symptom intensity was studied in the first three weeks after admission to a hospice. Fatigue, dry mouth, and anorexia were the most prevalent and intense symptoms according to both patients and nurses. Nausea, anxiety, and dysphagia were the least prevalent symptoms and depressed mood, anxiety, and nausea had the lowest mean intensity scores according to patients. Nurses scored the identical symptoms as least prevalent but in a different order. Both patients and nurses indicate a decreased perceived well-being, although nurses overestimated well-being. The difference between USD-P and USD showed that nurses predominantly underestimate symptom intensity. Absolute concordance was relatively high for low-intensity symptoms. The weighted Kappa analysis shows that only dyspnea reaches modest agreement, whereas well-being and depressed mood scores showed only slight agreement and anxiety only reaches poor agreement.

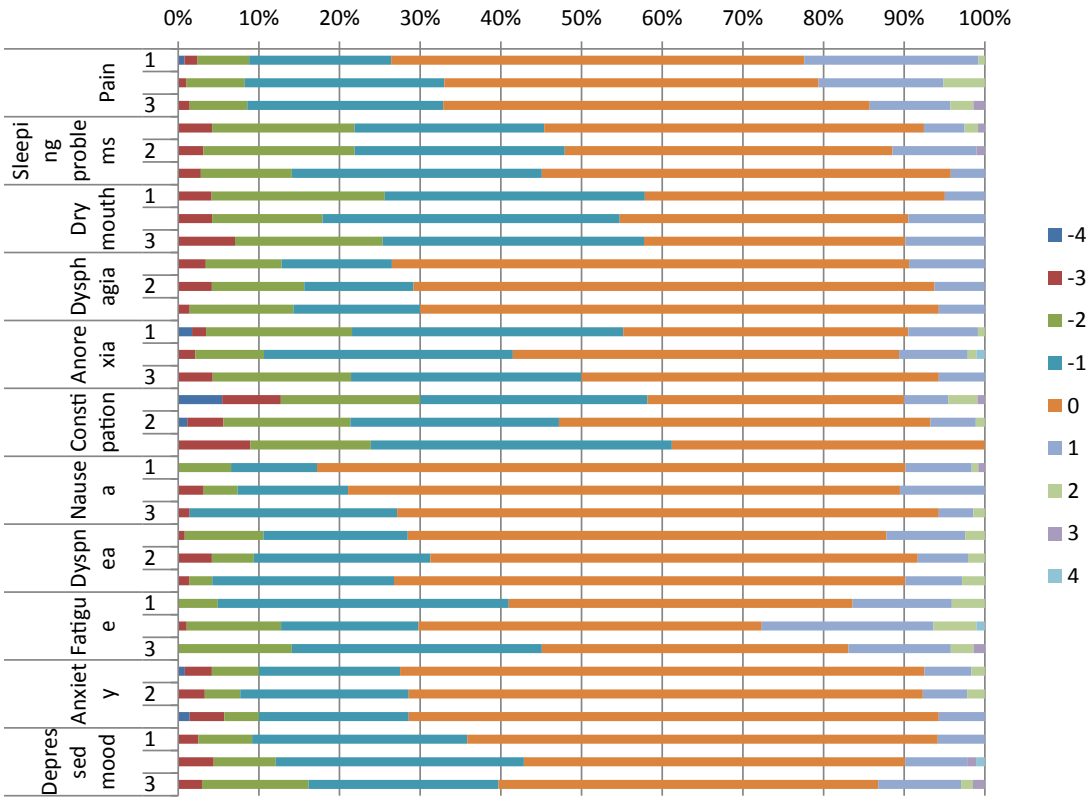

Fig. 2. Difference USD-P and USD (USD-P minus USD categorized score). USD = Utrecht Symptom Diary; USD-P = USD-Professional. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

There was little agreement with regard to well-being. Neither absolute concordance nor agreement beyond chance increased over time. An analysis restricted to patients for whom three subsequent dyads were available, confirmed this lack of improvement of concordance.

Strengths and Limitations

The primary strength of this study was that it was conducted in a real-life setting, using prospectively collected data. As a consequence, bias and confounding likely affect our results. Four considerations of the results are discussed.

First, patients in the last days before death are underrepresented in this study because they are less able to self-report their symptoms.[7] Because concordance is lowest in more intense symptoms, and symptom intensity increases and well-being decreases toward death, the absolute concordance in this study could be an overestimation of the true concordance in symptom intensity for all patients. Second, hospice inpatients are less able to self-report toward death and their ability to self-report fluctuates during admission. This could potentially have decreased the concordance over time. However, if the analysis was restricted to patients for whom three dyads were available, the concordance still did not improve over time. Third, there could be differences between nurses in their capacity to assess symptoms in their patients. As we were unable to link measurements to specific nurses, we were unable to test this hypothesis. Finally, to enable a comparison between USD and USD-P, USD cutoffs were used for categorization purposes. These cutoffs are not well established for most symptoms, except of pain and fatigue. This might have contributed to the low concordance scores.[31]

Concordance

The agreement beyond chance is mostly lower than the absolute concordance, except for anorexia and dyspnea. Although anorexia itself is not observable, appetite and eating patterns are. Meals are important daily routines and are considered important by patients, their families, and nurses. This seemed to have resulted in a better understanding of appetite or a lack of appetite and intake of food but requires an in-depth inquiry to be sure. Dyspnea, or shortness of breath, is observable for nurses and discussed because it is a known stressor for patients and their families. This could explain why both symptoms show a higher agreement although the symptom intensity was higher. For nausea, depressed mood, and

anxiety, a large degree of concordance was found but a small weighted Kappa. A large proportion of patients scoring zero on the USD could explain this difference. Apparently, nurses are able to observe the absence of a symptom but assessing the intensity is problematic.

Overall, nurses' underestimation occurs much more frequently than overestimation. This underestimation of both physical and psychological symptoms is congruent with most previous studies, although Dawber et al. found an overestimation of physical and psychological symptoms in an acute hospital population.[30] Although it is known that psychological symptoms are difficult to interpret and are more likely to be overlooked and undertreated in daily care,[24] our results indicate that both physical and psychological symptom are at risk for underestimation and thus undertreatment in the absence of self-report.

The fluctuations of symptom intensity could be an alternative explanation for the low concordance between patients and nurses. The USD and USD-P both assess symptom intensity "at this moment." As the symptoms are assessed on the same day but not necessarily at the same time, concordance might be lower. However, both patients and nurses usually assessed symptoms in the afternoon. An assessment where the timing is set at "in the last 24 hours" might be preferred for fluctuating symptoms.[37]

Concordance Over Time

The stability of complete concordance over time suggests that the ability of nurses to assess symptom intensity does not increase. This could be influenced by the fact that most nurses do not work full time, resulting in few consecutive workdays per period. In addition, nurses generally work day-, evening-, and nightshift. As a result, the number of comparable observations and contact with individual patients is low, and consequently, there is little possibility for nurses to learn from the patients' expressions of suffering from symptoms. Even in a hospice environment with solely specialized nurses, nurses tend to underestimate symptom burden of patients and should consequently emphasize to all patients and their families the importance of self-assessment. To level out the nurses' tendency to underestimate symptom intensity, a combined strategy of nurses' and family caregivers' assessments could be used when patients are unable or unwilling to self-report because family caregivers caring tend to overestimate patients' symptom intensity.[38] However, research is needed to establish if concordance is improved by a combined strategy and to study the feasibility of this strategy in daily practice.

To conclude, our results indicate that skilled hospice nurses are able to detect the absence of symptoms but are less competent to assess the intensity of symptoms, specifically severe symptoms. There did not seem to be a learning curve: the estimation of symptom severity did not improve during admission. Observable symptoms, such as dyspnea and dysphagia, have a better concordance than symptoms that are not easily observed.

Hospice care is multidimensional care aiming to optimize the quality of life of terminally ill patients. Symptom management is vital to an optimal quality of life and self-report is the gold standard to assess symptom intensity. However, patients and the multiprofessional hospice team have to rely on proxy assessment when patients become unwilling or unable to self-assess their symptoms. In daily practice, nurses should be aware of the likelihood of underestimation of symptom intensity, specifically for symptoms that are difficult to observe. Nurses could develop strategies to overcome their underestimation, by reflecting on their estimates using concurrent patients' self-report measures and the use of dyads of family members' symptom intensity scores and nurses' symptom intensity scores concurrently for patients who are unable or unwilling to self-report. These strategies may increase concordance and decrease the chance of under-assessment and as a result undertreatment for these symptoms.

Symptom management by a multiprofessional team is founded on an impeccable assessment of symptom prevalence and intensity. Nurses have a major responsibility to assess symptom intensity. Therefore, the assessment of symptom intensity and the integration of these results in daily practice should be key in the nursing basic education and specialized palliative care courses.

Disclosures and Acknowledgments

The authors would like to express our thanks to the patients who took part in this study and the dedicated hospice caregivers who made it possible. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Hui D, de La Cruz M, Mori M, et al. Concepts and definitions for "supportive care," "best supportive care," "palliative care," and "hospice care" in the published literature, dictionaries, and textbooks. Support Care Cancer 2013; 213:659–685.

2. Saunders C. Hospice care. Am J Med 1978;65:726–728.

3. Teunissen SCCM, Wesker W, Kruitwagen C, De Haes HCJM, Voest EE, De Graeff A. Symptom prevalence

in patients with incurable cancer: a systematic review. J Pain Symptom Manage 2007;341:94–104.

4. Kirkova J, Rybicki L, Walsh D, Aktas A. Symptom prevalence in advanced cancer: age, gender, and performance status interactions. Am J Hosp Palliat Care 2012;29:139–145.

5. Shah AB, Udeoji DU, Baraghoush A, Bharadwaj P, Yennurajalingam S, Schwarz ER. An evaluation of the prevalence and severity of pain and other symptoms in acute decompensated heart failure. J Palliat Med 2013;16:87–90.

6. Opasich C, Gualco A. The complex symptom burden of the aged heart failure population. Curr Opin Support Palliat Care 2007;14:255–259.

7. de Graaf E, Zweers D, de Graeff A, Daggelders G, Teunissen S. Does age influence symptom prevalence and intensity in hospice patients, or not. A retrospective cohort study. J Geriatr Palliat Care 20147.

8. Potter J, Hami F, Bryan T, Quigley C. Symptoms in 400 patients referred to palliative care services: prevalence and patterns. Palliat Med 2003;17:310–314.

9. Pereira J, Green E, Molloy S, et al. Population-based standardized symptom screening: cancer care Ontario's Edmonton symptom assessment System and performance status initiatives. J Oncol Pract 2014;10:212–214.

10. Pereira JL, Chasen MR, Molloy S, et al. Cancer care professionals' attitudes toward systematic standardized symptom assessment and the Edmonton symptom assessment System after large-scale population-based implementation in Ontario, Canada. J Pain Symptom Manage 2016;514:662–672.e8.

11. Currow DC, Allingham S, Yates P, Johnson C, Clark K, Eagar K. Improving national hospice/palliative care service symptom outcomes systematically through point-of-care data collection, structured feedback and benchmarking. Support Care Cancer 2015;232:307–315.

12. Eagar K, Watters P, Currow DC, Aoun SM, Yates P. The Australian Palliative Care Outcomes Collaboration (PCOC)—measuring the quality and outcomes of palliative care on a routine basis. Aust Health Rev 2010;34:186–192.

13. Etkind SN, Daveson BA, Kwok W, et al. Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: does it make a difference? A systematic review. J Pain Symptom Manage 2015;493:611–624.

14. Hui D, Bruera E. The Edmonton Symptom Assessment System 25 years later: past, present and future developments. J Pain Symptom Manage 2017;53:630–643.

15. Steindal SA, Sorbye LW, Bredal IS, Lerdal A. Agreement in documentation of symptoms, clinical signs, and treatment at the end of life: a comparison of data retrieved from nurse interviews and electronic patient records using the Resident Assessment Instrument for Palliative Care. J Clin Nurs 2012;21:1416–1424.

16. Jones JM, McPherson CJ, Zimmermann C, Rodin G, Le LW, Cohen SR. Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings. J Pain Symptom Manage 2011;423:354–365.

17. Horton R. 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. Palliat Med 2002;16:488–494.

18. Higginson IJ, Gao W. Caregiver assessment of patients with advanced cancer: concordance with patients, effect of burden and positivity. Health Qual Life Outcomes 2008;6:42.

19. Hisamura K, Matsushima E, Nagai H, Mikami A. Comparison of patient and family assessments of quality of life of terminally ill cancer patients in Japan. Psychooncology 2011;20:953–960.

20. McPherson CJ, Wilson KG, Lobchuk MM, Brajtman S. Family caregivers' assessment of symptoms in patients with advanced cancer: concordance with patients and factors affecting accuracy. J Pain Symptom Manage 2008;351:70–82.

21. Palma A, Del Rio I, Bonati P, et al. Frequency and assessment of symptoms in hospitalized patient with advanced chronic diseases: is there concordance among patients and doctors? Rev Med Chil 2008;136:561–569.

22. Stromgren AS, Groenvold M, Sorensen A, Andersen L. Symptom recognition in advanced cancer. A comparison of nursing records against patient self-rating. Acta Anaesthesiol Scand 2001;45:1080–1085.

23. Nekolaichuk CL, Bruera E, Spachynski K, MacEachern T, Hanson J, Maguire TO. A comparison of patient and proxy symptom assessments in advanced cancer patients. Palliat Med 1999;13:311–323.

24. Rhondali W, Hui D, Kim SH, et al. Association between patient-reported symptoms and nurses' clinical impressions in cancer patients admitted to an acute palliative care unit. J Palliat Med 2012;15:301–307.

25. de Graaf E, van Klinken M, Zweers D, Teunissen S. From concept to practice, is multidimensional care the leading principle in hospice care? An exploratory mixed method study. BMJ Support Palliat Care 2017; https://doi.org/10.1136/bmjspcare-2016-001200. [Epub ahead of print].

26. Claessens P, Menten J, Schotsmans P, Broeckaert B. Development and validation of a modified version of the Edmonton Symptom Assessment Scale in a Flemish palliative care population. Am J Hosp Palliat Care 2011;28:475–482.

27. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med 2007;45:247–251.

28. Bruera E, Kuehn N, Miller MJ, Selmser P. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care 1991;7:6–9.

29. Zeng L, Zhang L, Culleton S, et al. Edmonton Symptom Assessment Scale as a prognosticative indicator in patients with advanced cancer. J Palliat Med 2011;14:337–342.

30. Dawber R, Armour K, Ferry P, Mukherjee B, Carter C, Meystre C. Comparison of informal caregiver and named nurse assessment of symptoms in elderly patients dying in hospital using the palliative outcome scale. BMJ Support Palliat Care 2016; https://doi.org/10.1136/bmjspcare-2015-000850. [Epub ahead of print].

31. Oldenmenger WH, de Raaf PJ, de Klerk C, van der Rijt CC. Cut points on 0-10 numeric rating scales for symptoms included in the Edmonton Symptom Assessment Scale in cancer patients: a systematic review. J Pain Symptom Manage 2013;456:1083–1093.

32. Hearn J, Higginson I. Development and validation of a core outcome measure for palliative care: the palliative

care outcome scale. Palliative Care Core Audit Project Advisory Group. Qual Health Care 1999;8:219–227.

33. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977;33:363–374.

34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–174.

35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420.

36. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–549.

37. Hannon B, Dyck M, Pope A, et al. Modified Edmonton Symptom Assessment System including constipation and sleep: validation in outpatients with cancer. J Pain Symptom Manage 2015;495:945–952.

38. McMillan SC, Moody LE. Hospice patient and caregiver congruence in reporting patients' symptom intensity. Cancer Nurs 2003;26:113–118.