Brief Report

The Relationship Between Two Performance Scales: New York Heart Association Classification and Karnofsky Performance Status Scale

Miriam J. Johnson, MD, J. Martin Bland, PhD, Patricia M. Davidson, PhD, Phillip J. Newton, PhD, Stephen G. Oxberry, PhD, Amy P. Abernethy, MD, and David C. Currow, FRACP

Palliative Medicine (M.J.J.), Hull York Medical School, University of Hull, Hull, United Kingdom; Department of Health Sciences (J.M.B.), University of York, Heslington, York, United Kingdom; Cardiovascular and Chronic Care Centre (P.M.D., P.J.N.), University of Technology Sydney, St. Vincent's Hospital Sydney, Sydney, New South Wales, Australia; Palliative Medicine (S.G.O.), Kirkwood Hospice, Huddersfield, United Kingdom; Division of Medical Oncology (A.P.A.), Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA; and Discipline, Palliative and Supportive Services (A.P.A., D.C.C.), Flinders University, Daw Park, South Australia, Australia

Abstract

Context. Performance status is used to quantify the well-being and functional status of people with illness. Clinicians and researchers from differing fields may not instinctively understand the scales, typically disease specific, used in other disciplines.

Objectives. To provide a preliminary description of the relationship between the Karnofsky Performance Status Scale (KPS) and the New York Heart Association Classification (NYHA) and to stimulate discussion in research and clinical practice.

Methods. Simultaneous KPS and NYHA data (172 observations) from three studies of people with chronic heart failure were pooled. Linear regression was used to predict the mean KPS from NYHA. The strength of association between the scales was investigated using a Kendall's Tau-b correlation coefficient. The agreement between the predicted and observed KPS scores was investigated using weighted kappa with quadratic weights.

Results. Linear regression demonstrated a relationship between KPS and NYHA ($P < 0.0001$; $R^2 = 0.3$). Predicted KPS from NYHA class rounded to the nearest 10 gave the following values: Class I, predicted KPS 90%; Class II, predicted KPS 80%; Class III, predicted KPS 70%; and Class IV, predicted KPS 60%. A moderate strength of association between KPS and NYHA (Kendall's Tau-b correlation

Address correspondence to: Miriam J. Johnson, MD, Hull York Medical School, Hertford Building, The University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom. E-mail: miriam.johnson@hyms.ac.uk

Accepted for publication: May 8, 2013.

0885-3924/$ - see front matter
http://dx.doi.org/10.1016/j.jpainsymman.2013.05.006

coefficient of −0.49; $P < 0.0001$) and agreement between observed and predicted KPS (kappa coefficient = 0.52) was shown.

Conclusion. We suggest that the NYHA discriminates poorly between clinically important performance states in people with advanced disease (NYHA III and IV; KPS <50%). The KPS, used in conjunction, would provide useful additional information in research and clinical practice. J Pain Symptom Manage 2014;47:652−658.

Key Words

Performance status, Karnofsky Performance Status Scale, New York Heart Association Classification

Introduction

Performance status (PS) is used to quantify quickly the general well-being of people with illness and their ability to perform activities of daily living.[1−3] It is usually a proxy measure estimated by the clinician or researcher and influences the decision to apply treatment regimens, particularly with regard to conservative vs. nonconservative care, and to plan capacity for self-care. There are several well-established measures, although routine clinical use is uncommon except as an aid to decision making in oncology (e.g., "Is this patient fit enough to receive chemotherapy?") and as an estimate of disease severity in heart failure. In the research setting, PS often forms part of eligibility criteria and provides an intuitive description of the population studied.[4]

The choice of PS instruments typically aligns with disease-specific disciplines (e.g., oncology, cardiology, geriatrics). Little is known how the different discipline-specific scores relate to each other, and clinicians and researchers may not instinctively understand the descriptions used in other fields. This is an issue in the context of multiple chronic illnesses and a particular problem for palliative care clinicians and researchers, as palliative care serves as a single catchment for people with progressive severe disease irrespective of the diagnosis. Palliative care clinicians must intuitively create an "equals sign" to align severity of illness across diseases (e.g., metastatic cancer, heart failure, obstructive lung disease) to match clinical interventions with needs and define research populations.

A common understanding of regularly used PS measures in varying disease groups is needed. Furthermore, a common measure would help clinicians to appraise the relevance of research reports to their own practice and would aid pooling of data sets for secondary analyses.

In this article, we focus on a commonly used PS measure in cancer—the Karnofsky Performance Status Scale (KPS), and one from cardiology—the New York Heart Association Classification (NYHA), to demonstrate the issue. We present a preliminary exploration of the direct relationship between the KPS and NYHA as a hypothesis-generating exercise intended to stimulate discussion and prospective research and outline a pathway for current and future clinical practice.

Methods

Description of Scales

Karnofsky Performance Status Scale. The KPS, first described in 1948, is regarded as the gold standard performance scale for cancer patients.[5] It correlates well with physical functioning across its 11 ordinal measures and with survival at lower levels. Modified versions based on functional ability rather than the original focus on the place of care have been devised. The most recent (Australian-modified KPS) is most predictive of survival and has better face validity at the lower end of the scale.[6] Importantly, patient and clinician rating have been assessed.[3]

New York Heart Association Classification. The NYHA was first developed in 1928 and has undergone several revisions, the latest in 1994.[7] However, the measure is highly dependent

on the physician's opinion, reflected by its relatively poor inter-rater reliability with only 56% concordance between physicians.[8] As symptoms may fluctuate in heart failure, the severity of symptoms correlates poorly with prognosis. Despite these shortcomings, the NYHA is widely used in clinical practice and in research (both as an eligibility criterion and as a study outcome) to assess response to treatment and to guide management. The relationship between NYHA class and KPS is not known.

Description of Contributing Studies

This analysis was conducted on individual patient data pooled from three studies in people with advanced heart failure where both the KPS and NYHA PS scales were scored simultaneously. The contributing studies are reported elsewhere, or will be reported on completion, but are summarized here to place this article in context. Table 1 describes participants and setting. Data were available from 55 participants in two currently recruiting studies (oxygen therapy in heart failure [Oxygen HF] clinical trial: trial registration number ACTRN12609000103268; and the LACE study [Living and in the community with symptom burden with chronic heart failure]: personal communication, Davidson PM and Newton PJ) and 35 participants from a randomized, controlled, three-arm, crossover trial investigating the effect of opioids on chronic breathlessness.[9]

In all studies, the assessments of KPS and NYHA were made by trained researchers. Table 2 provides the descriptions for the KPS and NYHA performance scales. Participants in the Oxberry et al.[9] crossover trial provided baseline KPS/NYHA measures for each crossover ($n = 35$, contributing 70 observations). The 30 Oxygen HF study participants provided a total of 70 observations because of repeat measures at baseline, six weeks, three months, and six months. The LACE participants provided baseline measures only.

Ethics and Registration

All participants gave written informed consent. All studies received the required ethical approval. The clinical trials were registered (Oxberry et al.: ISRCTN 85268059; ACTRNO12606000269538). Confirmation was sought that National Health Service ethics approval was not required for a secondary analysis of anonymized data for this current pooled analysis.

Statistical Analysis

To allow for the repeat observations in the Oxberry et al. and Oxygen HF studies, we used robust standard errors, with each person in those studies being defined as a cluster. This does not affect the regression estimate but does change the standard errors to allow for the design.

The bivariate relationship between the NYHA and KPS was plotted. Linear regression was used to predict KPS from NYHA and the predicted score rounded to an integer and then to tens. We predicted mean KPS score

Table 1
Contributing Studies and Key Participant Characteristics

Characteristic	Oxberry et al. ($N = 35$)	Davidson (Oxygen-HF) ($N = 30$)	Davidson (LACE) ($N = 25$)
Design	Crossover RCT: morphine/oxycodone/placebo	Parallel group: medical air/oxygen/best care	Prospective cohort study
Setting	Outpatients in a tertiary cardiology center in northeast England	Inpatient wards and outpatient clinics in tertiary referral cardiology unit, Australia	Inpatient wards and outpatient clinics in tertiary referral cardiology unit, Australia
Age (years)	Mean = 41 SD = 20 Range = 10–90	Mean = 67 SD = 17 Range = 37–96	Mean = 60 SD = 18 Range = 27–89
Sex, M/F	30/5	23/7	14/11
Baseline intensity breathlessness	(NRS) Mean = 4.1 SD = 20 Range = 10–90	(VAS) Mean = 40.8 SD = 22 Range = 2–81	(NRS) Mean = 5.2 SD = 2.8 Range = 0–9
KPS	Mean = 69.1 SD = 6.1 Range = 60–80	Mean = 70 SD = 10 Range = 50–100	Mean = 76 SD = 9 Range = 60–90

HF = heart failure; RCT = randomized controlled trial; M = male; F = female; NRS = 0–10 numerical rating scale; VAS = 0- to 100-mm visual analogue scale; KPS = Karnofsky Performance Status Scale.

from the NYHA class and estimated the error of this as a mean for each NYHA category. We also estimated the error for the KPS value that would be predicted from the NYHA class for any individual. The precision of a predicted KPS score from an individual's NYHA class is important to know, as the aim was to replace individual NYHA scores, where that was the only PS measurement available, in the pooled data set, so all PS measures were in the form of KPS measures for the pooled analysis. To calculate the approximate standard error for a predicted observation, we added the residual variance about the regression line to the square of the robust standard error estimate for the regression estimate and took the square root.

The strength of the association between the KPS score and NYHA class was investigated using a Kendall's Tau-b correlation coefficient, as the best measure of association between two ordered categorical variables. The agreement between the KPS score predicted from the NYHA class using the model above and the observed KPS score was investigated using weighted kappa with quadratic weights. These analyses did not allow for the replication.

Results

The contributing studies provided 172 observations with both the NYHA and Karnofsky scores (Table 3). No individual was observed to have a KPS score <50%.

The linear regression to predict KPS from each NYHA group category and as individual forecast KPSs from individual NYHA measures is shown in Fig. 1, and the values are given in Table 4. A relationship is demonstrated (*P*-value for trend <0.0001), although this is not strong (R^2= 0.3). Predicted KPS from the NYHA class rounded to the nearest 10 gave the following values: Class I, predicted KPS 90%; Class II, predicted KPS 80%; Class III, predicted KPS 70%; and Class IV, predicted KPS 60%.

Using the Kendall's Tau-b correlation coefficient, a relationship between the KPS score and NYHA class was seen (correlation coefficient of −0.49; a negative value reflecting the inverse relationship; $P < 0.0001$). A moderate strength of agreement was demonstrated between the observed and predicted KPS scores. We used a quadratic weighted kappa coefficient, because the categories are ordered and a high penalty is given to disagreements in

Table 2
Scoring Descriptions for the KPS, AKPS, and NYHA Performance Scales

Scale Score	Original KPS and AKPS	
100%	Normal; no complaints; no evidence of disease.	
90%	Able to carry on normal activity; minor signs or symptoms.	
80%	Normal activity with effort; some signs or symptoms of disease.	
70%	Cares for self; unable to carry on normal activity or to do active work.	
60%	Requires occasional assistance but is able to care for most of his needs.	
50%	Requires considerable assistance and frequent medical care.	
Scale Score	**Original KPS**	**AKPS-Modified Description**
40%	Disabled; requires special care and assistance.	In bed more than 50% of the time.
30%	Severely disabled; hospitalization necessary; active supportive treatment is necessary.	Almost completely bedfast.
20%	Very sick; hospitalization necessary; active supportive treatment is necessary.	Totally bedfast and requiring extensive nursing care by professionals and/or family.
10%	Moribund; fatal processes progressing rapidly.	Comatose or barely arousable.
0%	Dead	Dead
Scale Class	**NYHA**	
I	No limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, or dyspnea (shortness of breath).	
II	Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in fatigue, palpitation, or dyspnea.	
III	Marked limitation of physical activity. Comfortable at rest, but less than ordinary activity causes fatigue, palpitation, or dyspnea.	
IV	Unable to carry out any physical activity without discomfort. Symptoms of cardiac insufficiency at rest. If any physical activity is undertaken, discomfort is increased.	

KPS = Karnofsky Performance Status Scale; AKPS = Australian-modified KPS; NYHA = New York Heart Association classification.

Table 3
Karnofsky Performance Status Scale and NYHA Class Ratings From 172 Pairs of Observations

	NYHA Classification				
Karnofsky Score	I	II	III	IV	Total
50	0	0	2	1	3
60	0	3	20	9	32
70	0	13	64	4	81
80	0	22	22	1	45
90	3	7	0	0	10
100	0	1	0	0	1
Total	3	46	108	15	172

NYHA = New York Heart Association Classification.

the extreme values where there were few data points. Again, a moderate strength of agreement was found (kappa = 0.52).

For 98 participants (57.0%), the KPS score is predicted correctly and is more than one possible value out for only four participants (2.3%).

Discussion

In this preliminary exploration, we have demonstrated a relationship between the NYHA class and KPS score in a subset of research study participants and developed an empirical conversion scale to allow KPS scores to be derived from the NYHA class.

Only three participants (1.7%) scored less than 60% on the KPS, despite 123 (75%) having the NYHA Class III or IV failure. Of the 15 with NYHA Class IV failure, representing those with advanced disease (indeed, NYHA Class III is often included in this definition), only one scored less than 60% on the KPS, which is considered moderately good PS in oncology and palliative care. Given that a KPS ≤30% relates particularly well to *poor* prognosis, it is unsurprising that the NYHA score is not a good prognostic marker. From these data, the NYHA appears to stop at a KPS of 60%; yet there are many gradations of *clinical meaningful* deterioration demonstrated by a KPS score <50%, indicating that the NYHA is not serving the clinician or researcher well at poorer levels of patients PS.

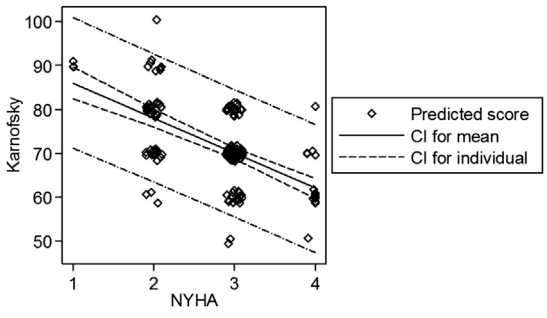

Fig. 1. Predicted Karnofsky Performance Status Scale score from the New York Heart Association Classification (NYHA) by category and individual with CIs.

What Are the Implications for Clinicians and Researchers?

Table 5 juxtaposes the descriptions of both scales according to scale category relationship. NYHA Classes I and II have similar descriptions as their KPS counterparts but are less concordant for worse PS categories; we suggest that the KPS discriminates better between different PS in people who are becoming less well.

Ezekowitz et al.[10] found a similar moderate association between the NYHA class and two other PS scales designed for palliative care populations (the Palliative Performance Scale and the Edmonton Symptom Assessment Scale) among a convenience sample cohort of 105 attendees at a heart function clinic. They concluded that the use of a palliative care PS scale in addition to the NYHA class may be helpful in the identification of people with heart failure who are appropriate for palliative or hospice care referral.

Similarly, the use of both the KPS score and NYHA class in studies with heart failure may add clarification to the description of the population. This may be particularly relevant for those with NYHA Class IV disease, but it is also notable that there is only moderate correlation at all levels of the NYHA. Using both scales even when KPS >60 may provide the clinician with additional useful information, and, for people with NYHA Class IV heart failure, routine application of the KPS in clinical practice may provide a better prompt for assessment of palliative care needs, given the wide range of PS apparently encompassed within this category.

Limitations

These are preliminary data including data from incomplete data sets (the Oxygen HF study and LACE are still recruiting). In addition, although the error for group mean values

Table 4
Predicted KPS Score From the NYHA by Category and Individual With CIs

NYHA Class	Predicted KPS	SE (Group Mean)	SE (Individual Forecast)	95% CIs	
				Group	Individual
I	86	1.84	7.66	82–90	71–100[a]
II	78	1.065	7.43	76–80	63–93
III	70	0.65	7.38	69–71	56–84
IV	62	1.16	7.44	60–64	47–77

KPS = Karnofsky Performance Status Scale; NYHA, New York Heart Association Classification; SE = standard error.
[a]Rounded down from the impossible estimated value of 101.

is relatively small, there is much less precision for the individual predicted values from any individual NYHA score.

In addition, the data available, as a subset of palliative care research study participant data, inevitably reflect the trial eligibility criteria. The Oxberry et al. and Oxygen HF eligibility criteria explicitly state the NYHA Class III/IV and stable clinical status for at least one month; thus, the participants' PSs are grouped in the poorer performance scales, but they were not so unwell as to be unable to participate in these prospective studies. Thus, there are few data for people at the NYHA class extremes. We have attempted to minimize this issue by using weighted kappa with quadratic weights, but conclusions about the relationship between the NYHA and KPS scales would be more robust and valid if a full range of scores were available for both.

Table 5
Scoring Descriptions Between the NYHA and KPS Using Group Mean Predictions

Scale Score	Original KPS and Australian-Modified KPS		Scale Class	NYHA
100%	Normal; no complaints; no evidence of disease.		0	
90%	Able to carry on normal activity; minor signs or symptoms.		I	No limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, or dyspnea (shortness of breath).
80%	Normal activity with effort; some signs or symptoms of disease.		II	Slight limitation of physical activity. Comfortable at rest, but ordinary physical activity results in fatigue, palpitation, or dyspnea.
70%	Cares for self; unable to carry on normal activity or to do active work.		III	Marked limitation of physical activity. Comfortable at rest, but less than ordinary activity causes fatigue, palpitation, or dyspnea.
60%	Requires occasional assistance but is able to care for most of his needs.		IV	Unable to carry out any physical activity without discomfort. Symptoms of cardiac insufficiency at rest. If any physical activity is undertaken, discomfort is increased.
50%	Requires considerable assistance and frequent medical care			
	Original KPS	AKPS-Modified Description		
40%	Disabled; requires special care and assistance.	In bed more than 50% of the time.		
30%	Severely disabled; hospitalization necessary; active supportive treatment is necessary.	Almost completely bedfast.		
20%	Very sick; hospitalization necessary; active supportive treatment is necessary.	Totally bedfast and requiring extensive nursing care by professionals and/or family.		
10%	Moribund; fatal processes progressing rapidly.	Comatose or barely arousable.		

The attempt to cross predominantly cancer-focused assessments (KPS) into non-cancer areas such as heart failure may be seen as unreasonable. However, multi-morbidity provides a challenge: people with cancer (cared for by oncologists) also have heart failure and people with heart failure also have cancer (cared for by cardiologists), and all these patients are seen in primary and palliative care settings.

What Further Research Is Needed to Understand the Relationship Between the NYHA and KPS?

In view of the limitations inherent in this secondary analysis of research study data, a prospective study designed to investigate the relationship between the two scores is needed to gain more information about patients with a PS at the more extreme ends of the KPS. It would also be useful to know whether a tool such as the KPS would perform better as a sole PS scale in people with heart failure than the NYHA, whether the use of the two together (the two scales may measure distinctly different processes) is better than either alone or whether a new measure designed for use irrespective of etiology should be developed.

Conclusions

Given the challenge of multi-morbidity in the care and study of people with advanced disease, it is important that the meaning of global performance assessments is understood across disciplines and the situations where we may need more than one assessment are defined. We suggest that the NYHA discriminates poorly between clinically important performance states in people with advanced disease (NYHA Class III and IV). The KPS, used in conjunction, would provide useful additional information in research and clinical practice.

Disclosures and Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that they have no financial or personal relationships with individuals, organizations, or companies that might be perceived to bias the work.

References

1. Dajczman E, Kasymjanova G, Kreisman H, et al. Should patient-rated performance status affect treatment decisions in advanced lung cancer? J Thorac Oncol 2008;3:1133–1136.

2. Maltoni M, Caraceni A, Brunelli C, et al. Prognostic factors in advanced cancer patients: evidence-based clinical recommendations—a study by the Steering Committee of the European Association for Palliative Care. J Clin Oncol 2005;23:6240–6248.

3. Jeyasingam L, Agar M, Soares M, Plummer J, Currow DC. A prospective study of unmet activity of daily living needs in palliative care inpatients. Aust Occup Ther J 2008;55:266–272.

4. Orr ST, Aisner J. Performance status assessment among oncology patients: a review. Cancer Treat Rep 1986;70:1423–1429.

5. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod CC, ed. Evaluation of chemotherapeutic agents in cancer. New York: Columbia University Press, 1949:191–205.

6. Abernethy AP, Shelby-James T, Fazekas BS, Woods D, Currow DC. The Australia-modified Karnofsky Performance Status (AKPS) scale: a revised scale for contemporary palliative care clinical practice [ISRCTN81117481]. BMC Palliat Care 2005;4:7.

7. The Criteria Committee of the New York Heart Association. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels, 6th ed. Boston, MA: Little, Brown & Co., 1994.

8. Goldman L, Hashimoto B, Cook EF, Loscalzo A. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new specific activity scale. Circulation 1981;64:1227–1234.

9. Oxberry SG, Torgerson DJ, Bland JM, et al. Short-term opioids for breathlessness in stable chronic heart failure: a randomized controlled trial. Eur J Heart Fail 2011;13:1006–1012.

10. Ezekowitz JA, Thai V, Hodnefield TS, Sanderson L, Cujec B. The correlation of standard heart failure assessment and palliative care questionnaires in a multidisciplinary heart failure clinic. J Pain Symptom Manage 2011;42:379–387.