



Inter-rater reliability in performance status assessment among healthcare professionals: an updated systematic review and meta-analysis

Ronald Chow^{1,2} · Eduardo Bruera³ · Jennifer S. Temel⁴ · Monica Krishnan⁵ · James Im¹ · Michael Lock²

Received: 21 October 2019 / Accepted: 20 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Introduction Survival prediction for patients with incurable malignancies is invaluable information during end-of-life discussions, as it helps the healthcare team to appropriately recommend treatment options and consider hospice enrolment. Assessment of performance status may differ between different healthcare professionals (HCPs), which could have implications in predicting prognosis. The aim of this systematic review and meta-analysis is to update a prior systematic review with recent articles, as well as conduct a meta-analysis to quantitatively compare performance status scores.

Methods A literature search was carried out in Ovid MEDLINE, Embase, and Cochrane Central Register of Controlled Trials, from the earliest date until the first week of August 2019. Studies were included if they reported on (1) Karnofsky Performance Status (KPS), Eastern Cooperative Oncology Group (ECOG) Performance Status, and/or Palliative Performance Scale (PPS) and (2) assessment of performance status by multiple HCPs for the same patient sets. The concordance statistics (Kappa, Krippendorff's alpha, Kendall correlation, Spearman rank correlation, Pearson correlation) were extracted into a summary table for narrative review, and Pearson correlation coefficients were calculated for each study and meta-analyzed with a random effects analysis model. Analyses were conducted using Comprehensive Meta-Analysis (Version 3) by Biostat.

Results Fourteen articles were included, with a cumulative sample size of 2808 patients. The Pearson correlation coefficient was 0.787 (95% CI: 0.661, 0.870) for KPS, 0.749 (95% CI: 0.716, 0.779) for PPS, and 0.705 (95% CI: 0.536, 0.819) for ECOG. Four studies compared different tools head-to-head; KPS was favored in three studies. The quality of evidence was moderate, as determined by the GRADE tool.

Conclusions The meta-analysis's Pearson correlation coefficient ranged from 0.705 to 0.787; there is notable correlation of performance status scores, with no one tool statistically superior to others. KPS is, however, descriptively better and favored in head-to-head trials. Future studies could now examine the accuracy of KPS assessment in prognostication and focus on model-building around KPS.

Keywords Performance status · Inter-rater reliability · Healthcare professionals

✉ Ronald Chow
rchow48@uwo.ca

¹ Department of Epidemiology and Biostatistics, Schulich School of Medicine & Dentistry, University of Western Ontario, London, Ontario, Canada

² Division of Radiation Oncology, Department of Oncology, London Regional Cancer Program, University of Western Ontario, London, Ontario, Canada

³ Department of Palliative, Rehabilitation and Integrative Medicine, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

⁴ Division of Hematology and Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

⁵ Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Harvard University, Boston, MA, USA

Introduction

Survival prediction for patients with incurable malignancies is invaluable information during end-of-life discussions, as it helps the healthcare team to appropriately recommend treatment options and consider hospice enrolment [1]. A good estimate of survival, for example, can help avoid aggressive medical care that may lead to lower quality of life, worse caregiver bereavement outcomes, and a greater economic burden from healthcare [2–4].

Many studies have reported that performance status is a good prognostic indicator in patients with advanced cancer [5–10]. A literature review by Krishnan et al. in 2013 identified that 10 of 13 studied prognostic models incorporated performance status scores, namely, the Karnofsky Performance Status (KPS), Palliative Performance Scale (PPS), or Eastern Cooperative Oncology Group (ECOG) Performance Status [11]. These scales have also been used for enrolment eligibility into clinical trials as well as stratification factors in

oncology trial analysis [1]. These validated tools, however, are subjective and hence can be subject to bias and high interobserver variability [12].

Performance status may differ between different healthcare professionals (HCPs), which could have implications in predicting prognosis. It would therefore be of interest to determine and subsequently use the performance status with greatest interrater reliability during prognostication, to ensure minimal variation between clinicians. In 2016, a systematic review by Chow et al. studied interrater reliability of HCPs and reported that KPS had better agreement than both ECOG and PPS, based on narrative synthesis [1]. A meta-analysis of study data may be able to provide a better comparison of performance scales. Since then, several papers have also been published in the field, which could provide further insight into the interrater concordance.

The aim of this systematic review and meta-analysis is to update the aforementioned systematic review with recent articles, as well as conduct a meta-analysis to quantitatively compare performance status scores.

Table 1 Study demographics

<i>Study</i>	<i>Country</i>	<i>Sample size</i>	<i>Performance scale</i>	<i>Assessors</i>
Hutchinson et al., 1979 [19]	Canada	60	KPS	Emergency room physician and senior medical resident ($n = 29$) Two renal physicians ($n = 31$)
Schag et al., 1984 [20]	United States of America	75	KPS	Physicians/oncologists and mental health professionals/psychologists/psychiatrist
Conill et al., 1990 [21]	Spain	100	KPS ECOG	Two oncologists
Roila et al., 1991 [22]	Italy	209	KPS ECOG	Two oncologists
Sorensen et al., 1993 [23]	Denmark	100	ECOG	Three oncologists
Fantoni et al., 1999 [24]	Italy	657	KPS	Experienced physician, young physician, and experienced nurse
Taylor et al., 1999 [25]	Australia	100	KPS ECOG	Oncologist, ward resident medical officer, and principal treating nurse
Ando et al., 2001 [26]	Japan	206	ECOG	Oncologist and nurse
Liem et al., 2002 [27]	United States of America	117	KPS	Attending and resident radiation oncologist
de Borja et al., 2004 [28]	Canada	36	KPS ECOG	Attending radiation oncologist, primary care nurse, staff radiation therapist
Campos et al., 2009 [29]	Canada	102	PPS	Attending radiation oncologist and palliative radiation therapist
Zimmermann et al., 2010 [30]	Canada	457	KPS ECOG PPS	Physician and nurse
Kim et al., 2015 [31]	United States of America	278	ECOG	Palliative care specialist, palliative care nurse, and medical oncologist
Neeman et al., 2019 [32]	United States of America	311	ECOG	Oncologist and nurse

ECOG Eastern Cooperative Oncology Group Performance Status

KPS Karnofsky Performance Status

PPS Palliative Performance Scale

Table 2 Interrater concordance of performance status scales

<i>Study</i>	<i>Concordance statistic</i>	<i>Interpretation</i>
Karnofsky Performance Status (KPS)		
Hutchinson et al., 1979 [19]	Kappa coefficient = 0.50 (emergency room physician and resident) Kappa coefficient = 0.46 (two renal physicians)	Good Good
Schag et al., 1984 [20]	Pearson correlation = 0.89 Kappa coefficient = 0.53	Good Good
Conill et al., 1990 [21]	Kendall's correlation = 0.76	Good
Roila et al., 1991 [22]	Kappa coefficient = 0.921	Good
Fantoni et al., 1999 [24]	Kendall's correlation = 0.82 (experienced and young physician) Kendall's correlation = 0.77 (experienced physician and nurse) Kendall's correlation = 0.76 (young physician and nurse)	Good Good Good
Taylor et al., 1999 [25]	Spearman correlation = 0.60–1.00	Good
Liem et al., 2002 [27]	Kappa coefficient = 0.29 Pearson correlation = 0.85 Spearman correlation = 0.76 Kendall's correlation = 0.67	Poor Good Good Moderate
de Borja et al., 2004 [28]	Spearman correlation = 0.74 (physician and nurse) Spearman correlation = 0.67 (physician and radiation therapist)	Good Good
Zimmermann et al., 2010 [30]	Kappa coefficient = 0.74	Moderate
Eastern Cooperative Oncology Group Performance Status (ECOG)		
Conill et al., 1990 [21]	Kendall's correlation = 0.75	Good
Roila et al., 1991 [22]	Kappa coefficient = 0.914	Good
Taylor et al., 1999 [25]	Spearman correlation = 0.60–1.00	Good
de Borja et al., 2004 [28]	Spearman correlation = 0.77 (physician and nurse) Spearman correlation = 0.57 (physician and radiation therapist)	Good Good
Zimmermann et al., 2010 [30]	Kappa coefficient = 0.67	Moderate
Kim et al., 2015 [31]	Kappa coefficient = 0.26 (palliative care physician and medical oncologist) Kappa coefficient = 0.23 (palliative care nurse and medical oncologist) Kappa coefficient = 0.61 (palliative care physician and palliative care nurse)	Poor Poor Moderate
Neeman et al., 2019 [32]	Kappa coefficient = 0.486 Spearman correlation = 0.612	Moderate Good
Palliative Performance Scale (PPS)		
Campos et al., 2009 [29]	Spearman correlation = 0.69 (physician and radiation therapist) Spearman correlation = 0.83 (physician and research assistant) Spearman correlation = 0.76 (radiation therapist and research assistant)	Good Very good Good
Zimmermann et al., 2010 [30]	Kappa coefficient = 0.72	Moderate

Note: The interpretation is based on criteria established in the Methods section

Methods

Search strategy

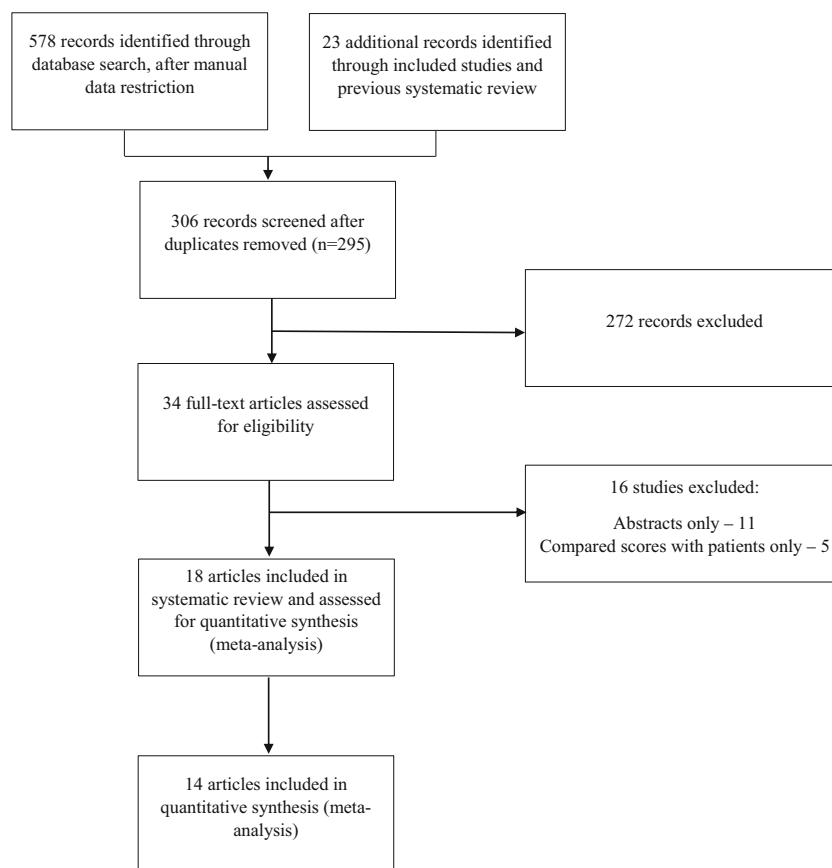
A literature search was carried out in Ovid MEDLINE, Embase, and Cochrane Central Register of Controlled Trials, from the earliest date until the first week of August 2019. The search was based off of that conducted by Chow et al. [1] and incorporated phrases such as “Karnofsky Performance Status,” “Eastern Cooperative Oncology Group Performance Status,” “Palliative Performance Status,” and “prognostic tool or prognostic instrument.” The search was restricted to English-language articles. As the prior review assessed

articles up until July 2015, a manual restriction was subsequently imposed on the search results, identifying only articles published in 2015 or beyond for screening in this systematic review [Appendix 1].

Screening

Following duplication removal, titles and abstracts were reviewed (Level 1 screening) to determine whether papers (1) used either KPS, PPS, or ECOG and (2) had performance status assessment conducted by multiple individuals. These studies were then eligible for full-text review (Level 2 screening) and included for data extraction if

Fig. 1 PRISMA flow diagram



studies reported multiple HCPs who rated the same patients and reported on inter-rater concordance. Inter-rater data between students and healthcare professionals were not eligible for inclusion. Abstracts without accompanying full texts were excluded at this stage. Data was subsequently extracted from articles, to note the calculated concordance statistic as well as the inter-rater concordance table. For studies that did not have an inter-rater concordance table, the authorship team was contacted to request for the data. If no response was received, authors were contacted again at both 2 and 4 weeks after the original request. Studies were excluded from this review if authors could not be reached or could not provide the available data.

Meta-analysis

The primary endpoint was the reported or calculated concordance statistic, based on underlying data supplied by the authorship team of the published paper. The reported statistics were included in narrative synthesis. Kappa values range from -1 to $+1$, with values closer to $+1$ indicating higher concordance. Krippendorff's alpha has a range from 0 to 1, where higher values indicate better concordance. Kendall's tau coefficient, Spearman rank correlation, and Pearson correlation coefficient span from -1 to $+1$; absolute value of correlation statistics closer to 1 indicates greater agreement while the polarity of the statistic indicating the direction of agreement. The concordance statistics were interpreted as follows:

Fig. 2 Karnofsky Performance Status (KPS)

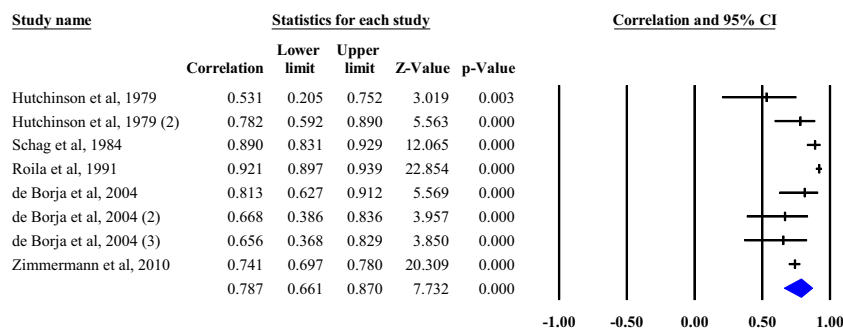
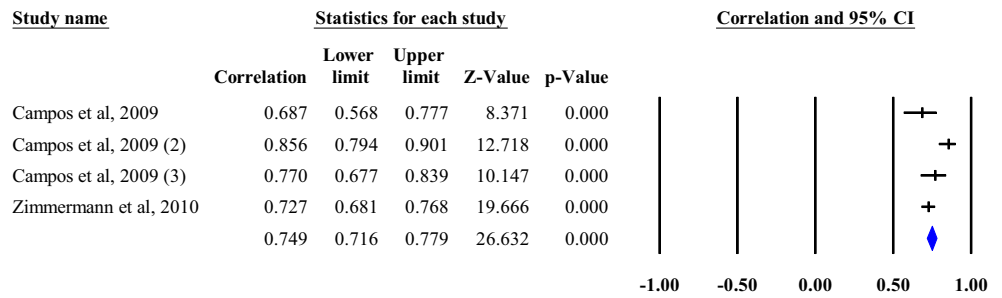


Fig. 3 Palliative Performance Scale (PPS)



- Kappa values greater than 0.40 indicate moderate agreement, and values above 0.75 reflect good agreement [13].
- Krippendorff’s alpha of 0.80 or higher is considered as good correlation [14].
- Kendall coefficient in excess of 0.7 is interpreted as good reliability [15].
- Spearman rank correlation values (specifically, absolute values) from 0.4 to 0.6 indicate moderate correlation, 0.6 to 0.8 indicates good correlation, and 0.8 to 1.0 indicates very good correlation [16].
- Pearson correlation coefficient greater than 0.80 signifies good correlation, while values less than 0.5 denote poor correlation [17].

Pearson correlation coefficients were recorded or calculated for each study and meta-analyzed with a random effect analysis model to yield a summary correlation coefficient per performance status scale. Analyses were conducted using Comprehensive Meta-Analysis (Version 3) by Biostat.

Assessment of study quality

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool [18] was used to assess the quality of evidence, for studies and the corresponding data, pertaining to each performance status.

Results

A total of 578 records were identified through search strategy, following manual restriction by date of

publication. From the prior systematic review and also through searching the reference lists of the included studies, an additional 23 records were identified. After duplicates were removed, 306 records underwent Level 1 screening; 34 were subsequently assessed during Level 2 screening. Eighteen were deemed relevant for data extraction, of which 14 [19–32] were included in our review (Fig. 1). Two studies included in the prior review were excluded, as only a conference proceeding was published [33, 34].

A total of 2808 patients were studied across all 15 studies; sample size per study ranged from 36 to 657. KPS was studied in 9 studies, ECOG in 9, studies and PPS in 2 studies [Table 1]. Studies reported moderate to good inter-rater reliability for all three tools [Table 2]. Only one study, by Zimmermann et al. [30], studied all three tools; KPS had the best inter-rater concordance. Three other studies studied both KPS and ECOG, two (Conill et al., 1990 [21] and Roila et al., 1991 [22]) reported better inter-rater agreement for KPS, and one (Taylor et al., 1999 [25]) presented inconclusive data.

The meta-analysis yielded Pearson correlation coefficients of 0.787 (95% CI: 0.661, 0.870) for KPS, 0.749 (95% CI: 0.716, 0.779) for PPS, and 0.705 (95% CI: 0.536, 0.819) for ECOG [Figs. 2, 3, and 4]; there is no tool that is statistically superior in terms of reliability, given the overlapping confidence intervals. The quality of evidence was assessed to be moderate for each performance status tool [Table 3].

Discussion This is the first meta-analysis of this nature to our knowledge with a total sample size of 2808. The meta-analysis did not indicate any tool to be statistically superior in terms of inter-rater reliability, as noted by overlapping

Fig. 4 Eastern Cooperative Oncology Group Performance Status (ECOG)

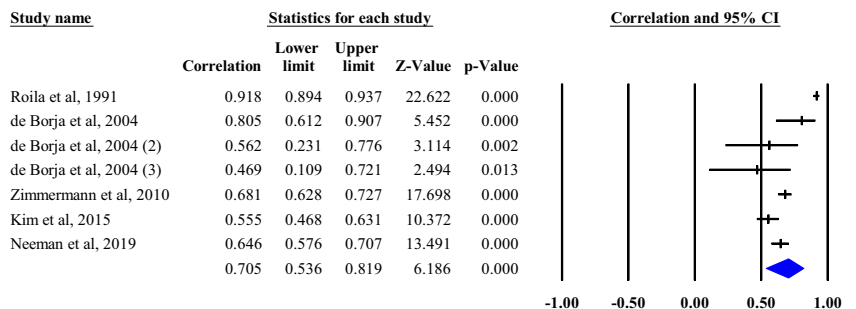


Table 3 GRADE summary of findings table (adapted version): performance status tools

Performance status tool	Correlation (95% CI)	Number of participants	Number of studies	Quality of evidence (GRADE)
Karnofsky Performance Status	0.787 (0.661, 0.870)	882	4	⊕⊕⊕○ Moderate due to imprecision ^a
Palliative Performance Scale	0.749 (0.716, 0.779)	763	2	⊕⊕⊕○ Moderate due to imprecision ^a
ECOG Performance Status	0.705 (0.536, 0.819)	1336	5	⊕⊕⊕○ Moderate due to imprecision ^{a,b}

^a Some studies had small sample size or large variability, leading to wide confidence intervals

^b Some studies had extra segments of scores (i.e., ECOG score of 3.0, 3.5, 4.0), leading to varying variability

confidence intervals. The point interval estimates of the Pearson correlation coefficients suggest that KPS has the best concordance, which is similarly echoed in the single study that compared all three tools [30]. Given that two of the three studies that examined KPS and ECOG similarly echo KPS to have superior inter-rater reliability, KPS may be slightly better, descriptively, than other tools. This conclusion supports the prior review [1].

Despite the reported differences in inter-rater concordance, multiple studies report good correlation of PS scores. However, the correlation is not perfect, and this may be due to different assessment techniques. While medical oncologists may place emphasis on documenting efficacy and toxicity of treatment, palliative care physicians may focus more on symptom distress and daily function and thereby downgrade a patient's performance status. Coupled with the fact that oncologists only continue chemotherapy for patients with better performance status, they may have a more optimistic bias in determining performance status than palliative care specialists [31]. To improve inter-rater reliability, training programs may be useful, to refocus HCPs towards common assessment techniques.

Scores may also vary between HCPs based on clinic flow. HCPs who typically bring patients into the examination room tend to have a greater understanding of patients' disabilities [30]. Nurses and research assistants, therefore, may have a greater understanding of functionality and mobility of patients; an oncologist may not be aware of a patient's disabilities and may grade patients differently. To date, however, there is no investigation into whether nurses or research assistants rate performance status with greater accuracy.

This systematic review was not without limitations. As inherent in all systematic review methodologies, individual study bias is not minimized due to simply amalgamating study data in a meta-analysis. Additionally, multiple studies that had raw data on inter-rater reliability were not included in this review due to lack of data available, even after contacting the authorship teams. The applicability of Pearson correlation coefficient as a stand-alone summary metric generated by meta-analysis may not be as applicable/robust as Kappa

coefficient and Spearman rank correlation coefficients, which may be more suitable for summarizing inter-rater concordance data. The Pearson correlation therefore should be used for predominantly illustrative purposes for comparison, rather than a definitive estimate of concordance.

In conclusion, KPS may be the best performance status tool, albeit not statistically superior to others. There is, however, a paucity of studies directly comparing different performance status tools, and future studies could compare multiple tools head-to-head. Given this study conclusion and agreement with the prior review, future studies could now examine the accuracy of KPS assessment in prognostication and focus on model-building around KPS. Furthermore, training programs may potentially be useful in standardizing assessment techniques, to increase inter-rater reliability.

Compliance with ethical standards

Conflict of interest None.

Appendix 1

Search Strategy

Database: Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed.

Citations, Daily and Versions(R) <1946 to August 05, 2019 > Search Strategy:

1 exp. Karnofsky Performance Status/ or Karnofsky performance status.mp. (4669).

2 Eastern Cooperative Oncology Group Performance Status.mp. (2204).

3 (KPS or ECOGPS or ECOG or ECOG or PPS).mp. (16049).

4 ((Karnofsky or Eastern Cooperative Oncology Group or palliative) adj4 (scale or status or score)).mp. (11320).

5 (performance status or performance scale or performance score).mp. (27973).

6 or/1–5 (40324).

7 ((evaluat* or assess* or compar* or choose or choice or select or pick or prefer* or inter-rater or interrater or rate or

rating or difference*) adj5 (“performance status” or “performance score” or “performance scale” or KPS or Karnofsky or ECOG* or prognostic tool* or prognostic.

instrument*).mp. (3265).

8 (physician* or doctor* or nurse* or oncologist* or research assistant* or clinician* or practitioner* or specialist*).mp. (1250590).

9 ((physician* or doctor* or nurse* or oncologist* or research assistant* or clinician* or practitioner* or specialist*) adj5 (“performance status” or KPS or Karnofsky or ECOG* or prognostic tool* or prognostic instrument*).mp. (218).

10 6 and (9 or (7 and 8)) (357).

11 limit 10 to English language (346).

Database: Embase Classic+Embase <1947 to 2019 Week 31 > Search Strategy:

1 exp. Karnofsky Performance Status/ or Karnofsky performance status.mp. (11358).

2 Eastern Cooperative Oncology Group Performance Status.mp. (3254).

3 (KPS or ECOGPS or ECOG or ECOG or PPS).mp. (40720).

4 ((Karnofsky or Eastern Cooperative Oncology Group or palliative) adj4 (scale or status or.

score)).mp. (19654).

5 (performance status or performance scale or performance score).mp. (54278).

6 or/1–5 (82709).

7 ((evaluat* or assess* or compar* or choose or choice or select or pick or prefer* or inter-

rater or interrater or rate or rating or difference*) adj5 (“performance status” or “performance score” or “performance scale” or KPS or Karnofsky or ECOG* or prognostic tool* or prognostic.

instrument*).mp. (6119).

8 (physician* or doctor* or nurse* or oncologist* or research assistant* or clinician* or practitioner* or specialist*).mp. (1768578).

9 ((physician* or doctor* or nurse* or oncologist* or research assistant* or clinician* or practitioner* or specialist*) adj5 (“performance status” or KPS or Karnofsky or ECOG* or.

prognostic tool* or prognostic instrument*).mp. (450).

10 6 and (9 or (7 and 8)) (845).

11 limit 10 to english language (832).

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <June 2019 > Search.

Strategy:

1 exp. Karnofsky Performance Status/ or Karnofsky performance status.mp. (1445).

2 Eastern Cooperative Oncology Group Performance Status.mp. (1102).

3 (KPS or ECOGPS or ECOG or ECOG or PPS).mp. (12098).

4 ((Karnofsky or Eastern Cooperative Oncology Group or palliative) adj4 (scale or status or.

score)).mp. (4804).

5 (performance status or performance scale or performance score).mp. (13980).

6 or/1–5 (20035).

7 ((evaluat* or assess* or compar* or choose or choice or select or pick or prefer* or inter-

rater or interrater or rate or rating or difference*) adj5 (“performance status” or “performance.

score“ or “performance scale” or KPS or Karnofsky or ECOG* or prognostic tool* or prognostic.

instrument*).mp. (1526) (physician* or doctor* or nurse* or oncologist* or research assistant* or clinician* or practitioner* or specialist*).mp. (97075)((physician* or doctor* or

nurse* or oncologist* or research assistant* or clinician* or practitioner* or specialist*) adj5 (“performance status” or KPS or Karnofsky or ECOG* or prognostic tool* or prognostic instrument*).mp. (59) 6 and (9 or (7 and 8)) (156)

References

1. Chow R, Chiu N, Burera E et al (2016) Inter-rater reliability in performance status assessment among health care professionals: a systematic review. *Ann Palliat Med* 5:83–92
2. Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, Dahlin CM, Blinderman CD, Jacobsen J, Pirl WF, Billings JA, Lynch TJ (2010) Early palliative care for patients with metastatic non-small-cell lung cancer. *New Engl J Med* 363:733–742
3. Wright AA, Zhang B, Ray A, Mack JW, Trice E, Balboni T, Mitchell SL, Jackson VA, Block SD, Maciejewski PK, Prigerson HG (2008) Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA* 300:1665–1673
4. Zhang B, Wright AA, Huskamp HA, Nilsson ME, Maciejewski ML, Earle CC, Block SD, Maciejewski PK, Prigerson HG (2009) Health care costs in the last week of life: associations with end-of-life conversations. *Arch Intern Med* 169:480–488
5. Evans C, McCarthy M (1985) Prognostic uncertainty in terminal care: can the Karnofsky index help? *Lancet* 1:1204–1206
6. Pirovano M, Maltoni M, Nanni O, Marinari M, Indelli M, Zaninetta G, Petrella V, Barni S, Zecca E, Scarpi E, Labianca R, Amadori D, Luporini G (1999) A new palliative prognostic score: a first step for the staging of terminally ill cancer patients. Italian multicenter and study group on palliative care. *J Pain Symptom Manag* 17:231–239
7. Bachelot T, Ray-Coquard I, Catimel G et al (2000) Multivariable analysis of prognostic factors for toxicity and survival for patients enrolled in phase I clinical trials. *Ann Oncol* 11:115–116
8. Maltoni M, Caraceni A, Brunelli C, Broeckaert B, Christakis N, Eychmueller S, Glare P, Nabal M, Viganò A, Larkin P, de Conno F, Hanks G, Kaasa S, Steering Committee of the European Association for Palliative Care (2005) 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* 23:6240–6248
9. Lingjun Z, Jing C, Jian L, Wee B, Jijun Z (2009) Prediction of survival time in advanced cancer: a prognostic scale for Chinese patients. *J Pain Symptom Manag* 38:578–586

10. Chan EY, Wu HY, Chan YH (2013) Revisiting the palliative performance scale: changes in scores during disease trajectory predicts survival. *Palliat Med* 27:367–374
11. Krishnan M, Temel JS, Wright AA, Bernacki R, Selvaggi K, Balboni T (2013) Predicting life expectancy in patients with advanced incurable cancer: a review. *J Support Oncol* 11:68–74
12. Kelly CM, Shahrokni (2016) Moving beyond Karnofsky and ECOG performance status assessments with new technologies. *J Oncol*:6186543
13. Fleiss JL, Levin B, Paik MC (eds) (2003) *Statistical methods for rates and proportions*, 3rd edn. John Wiley & Sons, Inc., Hoboken
14. De Swert K (2002) Calculating inter-coder reliability in media content analysis using Krippendorff's alpha. Available online: <http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf>. Accessed 20 July 2015
15. Nunally JC (ed) (1978) *Psychometric theory*. McGraw-Hill, New York
16. Landis JR, Koch GC (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
17. Daniela S, Jantschi L. Pearson versus spearman, Kendall's tau correlation analysis on structure-activity relationships of biological active compounds. Available online: <http://ljs.academicdirect.org/A09/179-200.htm>.
18. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, GRADE Working Group (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336:924–926
19. Hutchinson TA, Boyd NF, Feinstein AR (1979) Scientific problems in clinical scales, as demonstrated in the Karnofsky index of performance status. *J Chronic Dis* 32:661–666
20. Schag CC, Heinrich RL, Ganz PA (1984) Karnofsky performance status revisited: reliability, validity and guidelines. *J Clin Oncol* 2: 187–193
21. Conill C, Verge E, Salamero M (1990) Performance status assessment in cancer patients. *Cancer* 65:1864–1866
22. Roila F, Lupattelli M, Sassi M, Basurto C, Bracarda S, Picciafuoco M, Boschetti E, Milella G, Ballatori E, Tonato M (1991) Intra- and inter-observer variability in cancer patients' performance status assessed according to Karnofsky and ECO scales. *Ann Oncol* 2: 437–439
23. Sorenson JB, Klee M, Palshof T et al (1993) Performance status assessment in cancer patients. An inter-observer variability study. *Br J Cancer* 159:1988–1992
24. Fantoni M, Izzi I, Del Borgo C et al (1999) Inter-rater reliability of a modified Karnofsky scale of performance status for HIV-infected individuals. *AIDS Patient Care STDs* 13:23–28
25. Taylor AE, Olver IN, Sivanthan T, Chi M, Purnell C (1999) Observer error in grading performance status in cancer patients. *Support Care Cancer* 7:332–335
26. Ando M, Ando Y, Hasegawa Y, Shimokata K, Minami H, Wakai K, Ohno Y, Sakai S (2001) Prognostic value of performance status assessed by patients themselves, nurses, and oncologists in advanced non-small cell lung cancer. *Br J Cancer* 85:1634–1639
27. Liem BJ, Holland JM, Kang MY et al (2002) Karnofsky performance status assessment: resident versus attending. *J Cancer Educ* 17:138–141
28. de Borja MT, Chow E, Bovett G et al (2004) The correlation among patients and health care professionals in assessing functional status using the Karnofsky and eastern cooperative oncology group performance status scales. *Support Cancer Ther* 2:59–63
29. Campos S, Zhang L, Sinclair E, Tsao M, Barnes EA, Danjoux C, Sahgal A, Goh P, Culleton S, Mitera G, Chow E (2009) The palliative performance scale: examining its inter-rater reliability in an outpatient palliative radiation oncology clinic. *Support Care Cancer* 17:685–690
30. Zimmermann C, Burman D, Bandukwala S, Seccareccia D, Kaya E, Bryson J, Rodin G, Lo C (2010) Nurse and physician inter-rater agreement of three performance status measures in palliative care outpatients. *Support Care Cancer* 18:609–616
31. Kim YJ, Hui D, Zhang Y et al. Difference in performance status assessments between palliative care specialists, nurses and oncologists. *J Pain Symptom Manag* 2015; 49: 1050–1058.e2
32. Neeman E, Gresham G, Ovasapians N et al (2019) Comparing physician and nurse eastern cooperative oncology group performance status (ECOG-PS) ratings as predictors of clinical outcomes in patients with Cancer. *Oncologist* 24:1–7
33. May CH, Lester JF, Lee S (2012) PS score discordance and why it matters. *Lung Cancer* 75:S18
34. Addy C, Sephton M, Suntharalingam J et al (2012) Assessment of performance status in lung cancer: do oncologists and respiratory physicians agree? *Lung Cancer* 75:S17–S18

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.