Original Article

Successful Validation of the Palliative Prognostic Score in Terminally Ill Cancer Patients

Marco Maltoni, MD, Oriana Nanni, PhD, Marco Pirovano, MD, Emanuela Scarpi, PhD, Monica Indelli, MD, Cinzia Martini, MD, Massimo Monti, MD, Ermenegildo Arnoldi, MD, Laura Piva, MD, Alberto Ravaioli, MD, Giorgio Cruciani, MD, Roberto Labianca, MD, Dino Amadori, MD, for the Italian Multicenter Study Group on Palliative Care

Divisione di Oncologia Medica (M.M., D.A.), Ospedale Pierantoni, Forlì, Italy; Unità di Biostatistica (O.N., E.S.), Istituto Oncologico Romagnolo, Forlì, Italy; Divisione di Oncologia Medica (M.P., R.L.), Ospedale S. Carlo Borromeo, Milan, Italy; Divisione di Oncologia Medica (M.I.), Ospedale S. Anna, Ferrara, Italy; Divisione Terapia del Dolore e Cure Palliative (C.M.), Istituto Nazionale Tumori, Milan, Italy; Divisione di Oncologia Medica (M.M.), Pio Albergo Trivulzio, Milan, Italy; Divisione di Oncologia Medica (E.A.), Ospedale di Trescore, Seriate, Italy; Divisione Terapia del Dolore e Cure Palliative (L.P.), Istituti Clinici di Perfezionamento, Milan, Italy; Divisione Oncologia Medica (A.R.) Ospedale degli Infermi, Rimini Italy; Divisione di Oncologia Medica, (G.C.) Ospedale Civile Lugo, Lugo, Italy

Abstract

The aim of this work was to validate a previously constructed prognostic score for terminally ill cancer patients in order to determine its value in clinical practice. The Palliative Prognostic Score (PaP Score) was tested on a population of 451 evaluable patients consecutively entered in the hospice programs of 14 Italian Palliative Care Centers. The score subdivided patients into three specific risk classes based on the following six predictive factors of death: dyspnea, anorexia, Karnofsky Performance Status (KPS), Clinical Prediction of Survival (CPS), total white blood count (WBC), and lymphocyte percentage. The performance of the PaP Score index in the training and testing sets was evaluated by comparing mortality rates in the 3 prognostic risk categories. The score was able to subdivide the validation-independent case series into three risk groups. Median survival was 76 days in group A (with a 86.6% probability of 30-day survival), 32 days in group B (with a 51.6% probability of 30-day survival), and 14 days in group C (with a 16.9% probability of 30-day survival). Survival medians were remarkably similar to those of the training set (64 days in group A, 32 days in group B, and 11 days in group C).

In the complex process of staging terminally ill patients, the PaP Score is a simple instrument which permits a more accurate quantification of expected survival. It has been validated on an independent case series and is thus suitable for use in clinical practice. J Pain Symptom Manage 1999;17:240–247. *© U.S. Cancer Pain Relief Committee, 1999*

Key Words

Palliative care, prognostic score, staging, training-testing, validation

Address reprint requests to: Marco Maltoni, M.D., Divisione di Oncologia Medica, Ospedale Pierantoni, Via C. Forlanini n. 34, 47100 Forlì, Italy

Accepted for publication: July 29, 1998.

0885-3924/99/$–see front matter
PII S0885-3924(98)00146-8

Introduction

Studies on new prognostic factors form an extensive part of the literature in oncology. Unfortunately, the results of these studies are often inconsistent and contradictory. Whereas widely accepted methodological principles guide the design, conduct, analysis, and reporting of clinical trials, no such guidelines exist for prognostic factor studies.

In a manner similar to clinical studies, Simon and Altman[1] classify prognostic studies into three groups: phase I study or early exploratory study; phase II or exploratory and generating hypotheses study; and phase III or confirmatory study. Phase I studies examine the association of individual factors with diagnosis and disease characteristics or the development of reproducible assays. Phase II studies determine whether prognostic factors provide improved means of identifying patients at particularly high or low risk of outcome (disease progression or death). Phase III studies are large confirmatory studies at prestated hypotheses and allow for more precise quantification of the magnitude of the effect.

The demonstration of an association between a factor and prognosis requires a univariate approach (without considering the role that other factors may play in such an association). Multivariate methods are required when the independent prognostic importance of several factors is investigated. Multiple regression models allow for the simultaneous assessment of each factor adjusted for all other factors included in the model.

When a model contains several prognostic variables, it is useful to construct a "prognostic score," which is a new variable combining the information from all the prognostic factors. Based on the range of the total risk score, it is possible to classify the patients in different risk groups. From the survival curves of different risk groups, the discriminatory power of the multivariate model used can be expressed. Usually, however, curves overestimate the model value because the same data used for selecting variables and estimating regression coefficients are then utilized to measure prognostic power. The same model may not have a predictive power in other similar but independent case series.

The assessment of the predictive value of a model, as part of the model building process, can be approached with a "training and testing" procedure.[2] In the training sample, the identification of prognostic factors and the construction of a prognostic index are developed with standard statistic analysis; in the testing sample (or validation sample), which may be either an independent data set or a sample splitting, the model is validated. With sample splitting, a portion of the data set is used for model development and another is used for evaluating discriminatory power for the model. This means that the data set must be large enough to afford sufficient precision for both model development and validation when split into two parts. When this is not possible, a new sufficiently large data set must be found for validation. This may be more convincing because an independently gathered set of data is used.[3]

With the aim of classifying patients with very advanced neoplastic diseases into homogeneous risk groups, our team constructed a prognostic score (Palliative Prognostic Score: PaP Score) on more than 500 terminally ill cancer patients (training set). As presented in the companion work,[4] this score is the result of the combination of the partial scores of the following prognostic factors investigated in multivariate analysis: Karnofsky Performance Status (KPS), Clinical Prediction of Survival (CPS), anorexia, dyspnea,[5] total white blood count (WBC), lymphocyte percentage.[6] On the basis of the PaP Score, three different risk groups were identified: group C had the worst prognosis (score >11.0), with a median survival of 11 days; group B (score 5.6–11.0) had a median survival of 32 days; and group A (score ≤ 5.5) a median survival of 64 days.

A different group of patients with the same eligibility criteria was recruited to verify the prognostic predictivity of the proposed score. The aim of this work is, therefore, to validate the PaP Score on a sufficiently large, independent, and consecutive set of terminally ill cancer patients.

Methods

The prognostic score to be validated in this work was constructed on a multicenter case series (training set) of more than 500 terminally ill cancer patients. The prognostic parameters used in the study are the result of a previous analysis of 34 parameters,[5,6] of which only KPS, CPS, anorexia, dyspnea, total WBC, and lym-

phocyte percentage were found to have an independent prognostic value. In order to obtain an easy-to-handle score, each factor was assigned a value proportional to its prognostic weight. The sum of these partial quantities for each individual patient was used to construct a total score (Table 1). All the patients were then subdivided into three different prognosis groups: A, B, and C. With the aim of validating the score, 451 evaluable patients from 14 Italian centers were consecutively enrolled in their respective centers over an 8-month period (January–August 1996). The same eligibility and exclusion criteria were adopted to put together a patient series which would be compatible with that of the training set. Patients with an advanced solid tumor for whom antiblastic therapy was no longer considered viable were accepted; palliative radiotherapy and anabolic hormone treatment were allowed, and renal cancer, multiple myeloma and other lymphatic pathologies were excluded.

Table 1
PaP Score and Classification of Patients in Three Risk Groups

	Partial score
Dyspnea	
no	0
yes	1
Anorexia	
no	0
yes	1.5
KPS	
≥50	0
30–40	0
10–20	2.5
CPS (weeks)	
>12	0
11–12	2.0
9–10	2.5
7–8	2.5
5–6	4.5
3–4	6.0
1–2	8.5
Total WBC	
normal (4,800–8,500 cell/mm^3)	0
high (8,501–11,000 cell/mm^3)	0.5
very high (>11,000 cell/mm^3)	1.5
Lymphocyte percentage	
normal (20.0–40.0%)	0
low (12.0–19.9%)	1.0
very low (0–11.9%)	2.5
Risk groups	**Total Score**
A 30-day survival probability >70%	0–5.5
B 30-day survival probability 30–70%	5.6–11.0
C 30-day survival probability <30%	11.1–17.5

PaP Score = Dyspnea score + Anorexia score + KPS score + CPS score + Total WBC score + Lymphocyte percentage score.

The following information was collected for each testing set patient at enrollment: CPS, based on the clinical experience of the physician; KPS; anorexia; dyspnea; WBC; and lymphocyte percentage. Moreover, personal details and data relative to the patient's stage of disease were reported.

As the study was a multicenter one, the difference in range of normal values for the biological parameters was overcome using formulas that permitted standardization into three categories. The WBC value was considered normal between 4,800 and 8,500 cells/mm^3. Leukocytosis was defined as mild when the WBC was >8,500 cell/mm^3 but ≤11,000 cell/mm^3 (equivalent to a ≤30% alteration of normal values). Leukocytosis was severe when the total WBC was found to be over 11,000 cells/mm^3 (equivalent to a >30% alteration of normal values). The lymphocyte percentage was considered normal between 20% and 40% of the total WBC and low for values <20% but ≥12% (equivalent to a ≤40% alteration of normal values). The lymphocyte percentage was considered very low for values <12% (equivalent to a >40% alteration of normal values).

To validate the PaP Score, the scoring procedure shown in Table 1 was applied to this new series of cases. Partial scores were assigned to each patient on the basis of his/her own prognostic factors, and the sum of these scores was then used to classify the patient into one of three different prognosis groups, A, B, and C. Curves for the three risk groups of the testing set were traced by the Kaplan and Meier method and the logrank test was calculated. All analyses were carried out using SAS Software.[7]

Results

Between January and August 1996, 451 patients were recruited from 14 Italian centers, some of which had also entered patients in the training sample. Patient characteristics are reported in Table 2. Both sexes were equally represented. The median age was 70 years (range 21–95). The most frequent pathologies were as follows: lung (18.6%), colorectal (14.9%), stomach (12.0%), and pancreas, liver, and gall bladder (11.5%). Almost two-thirds of the patients (65.4%) had locally advanced disease, and vis-

Table 2

Main Clinical-Biological Characteristics of 451 Patients

Variables	No. pts	%
Center of origin		
Forli	91	20.2
Ferrara	80	17.7
Milano (Istituto Nazionale Tumori)	48	10.6
Milano (Pio Albergo Trivulzio)	46	10.2
Seriate	37	8.2
Milano (Istituti Clinici Perfezionamento)	36	8.0
Rimini-Riccione	29	6.5
Lugo	27	6.0
Milano (S. Paolo)	16	3.5
Torino (Fondazione FARO)	14	3.1
Ancona	13	2.9
Milano (S. Carlo Borromeo)	8	1.8
Cesena	6	1.3
Sex		
Female	228	50.6
Male	223	49.4
Primary site of neoplasia		
Lung	84	18.6
Colorectal	67	14.9
Stomach	54	12.0
Pancreas, liver, gall bladder	52	11.5
Breast	41	9.1
Male urogenital	37	8.2
Female genital	28	6.2
Head and neck	26	5.8
Unknown primary site	15	3.3
Others	47	10.4
Metastatic sites		
Locally advanced disease	295	65.4
Viscera	251	55.7
Bone	124	27.5
Soft tissue	73	16.2
Central nervous system	59	13.1
KPS		
≥50	140	31.0
30–40	260	57.6
10–20	51	11.3
CPS week		
>12	49	10.9
11–12	47	10.4
9–10	32	7.1
7–8	64	14.2
5–6	65	14.4
3–4	114	25.3
1–2	80	17.7
Anorexia		
no	181	40.1
yes	270	59.9
Dyspnea		
no	302	67.0
yes	149	33.0
Palliative steroid treatment		
no	175	38.8
yes	276	61.2
Palliative progestinic treatment		
no	412	91.4
yes	39	8.6
Hospitalization		
no	371	82.3
yes	80	17.7
Blood transfusion in the last 15 days		
no	398	88.2
yes	53	11.8
Total WBC		
normal (4,800–8,500 cell/mm^3)	253	56.1
high (8,501–11,000 cell/mm^3)	107	23.7
very high (>11,000 cell/mm^3)	91	20.2
Lymphocyte percentage		
normal (20–40%)	162	35.9
low (12–19.9%)	180	39.9
very low (0–11.9%)	109	24.2

ceral metastases were present in 55.7%. A KPS of 30–40 was recorded in 57.6% of patients; only 31.0% had a KPS of ≥50. Anorexia was present in 60% of patients, and dyspnea was noted in 33.0%. Prior endocrine therapies included corticosteroids in 61.2% and progestin therapy in 8.6%. The low percentage of patients admitted to hospital (17.7%) indicates that most centers involved in the study had a Home Care Service. About 50% of patients had an abnormal WBC, and more than 64% had low lymphocyte values.

At the time of analysis (October 1996), 36 patients (8%) were still alive and were considered censored for the purposes of analysis. The median survival of the entire testing set was 33 days. Figure 1 shows the overall survival curves of the population, which highlights that the case series was made up of patients in the terminal stages of disease.

To validate the PaP Score, the scoring procedure shown in Table 1 was applied to the testing set. Partial scores were assigned to each patient on the basis of individual characteristics. These partial scores were than added together to form a total score, after which each patient was classified into one of the three risk groups: group A (28.2% or 127/451), group B (45.7% or 206/451), and group C (26.1% or 118/451).

Kaplan-Meier survival curves for the three testing set risk groups are shown in Fig. 2. It is evident that the three groups had different survival rates (logrank (= 203.8, $P < 0.0001$). The 30-day survival probability, as expected, was >70% (86.6%) for group A, from 30–70% (51.6%) for group B, and <30% (16.9%) for group C. Median survival and relative 95% confidence interval (95% CI) were as follows: 76 days (95% CI = 67–87 days) for group A, 32 days (95% CI = 28–39 days) for group B, and 14 days (95% CI = 11–18 days) for group C.

Discussion

The use of uniform and shared methodological standards has been recommended.[8] Our effort to construct a prognostic score for terminally ill cancer patients comprised five stages. Stage 1 involved the identification of statistically significant clinical variables by multivariate analysis.[5] Stage 2 was the identification of biological parameters by the same method.[6] Stage 3 consisted of construction of an integrated prognostic score.[4] Stage 4 comprised the validation of the score on an independent population, and Stage 5 involved using the score in clinical practice and verifying its acceptability by operators. Our work to date can be discussed from both a methodological and content point of view.

With regard to methodology, we believe that it is essential to validate a score before proposing it for use in clinical practice. By design, a prognostic model is predictive in the case series on which it was constructed. However, the same model might not possess predictive value in other similar, but independent populations. The validation stage must therefore be an integral part of the model building process.[2]

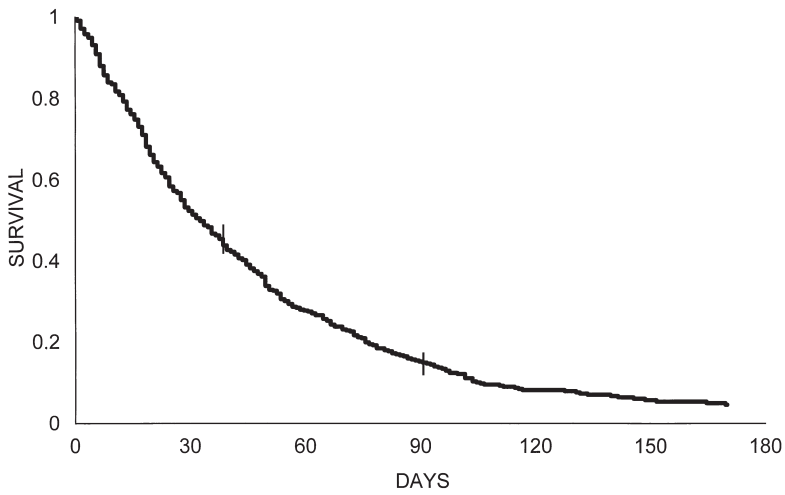

Fig. 1. Overall survival of 451 patients. Vertical bars represent 95% confidence interval.

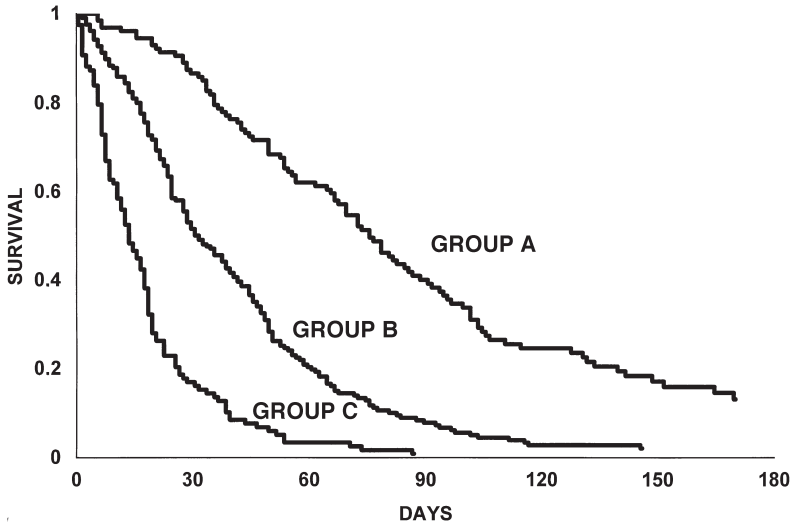

Fig. 2. Survival experience of the three groups of patients identified by the PaP Score of Table 1 in the testing series (Kaplan-Meier estimates). Logrank = 203.8 (2 df), $P < 0.0001$.

For this reason, our group decided to propose the score for clinical use only after its successful validation on an independent case series recruited using the same criteria as in the training set. The training and testing sets of the score were very similar to each other (median survival of 32 days in the former and 33 days in the latter). Moreover, they were almost identical to the sets presented in Christakis and Escarce's work[9] on the survival of patients after enrollment in hospice programs, in which a median survival of 36 days was reported for 6451 patients.

The latter observation would seem to indicate that eligibility and inclusion criteria for hospice programs are fairly standardized in most countries. However, survival rates vary among surveys. In Christakis and Escarce's case series, 15.6% of patients died within the first few weeks and 14.9% survived for more than 6 months. In both populations studied by us (training and testing set), a certain number of patients had a median survival of <2 weeks (26.2% and 26.1%, respectively) and, at the other extreme, a certain percentage of cases (34.3% and 28.2%, respectively) presented a median survival of >2 months. Therefore, within case series that are similar with regard to eligibility criteria, there are heterogeneous patient populations with varying survival characteristics and, consequently, different treatment and care needs. For this reason, we believe that the difficult process of treatment, care, and organizational decision making for this category of patients must not only take into account the quantitative aspects of life expectancy. On the contrary, careful consideration must also be made of all aspects linked to the patient's quality of life. It cannot be denied, however, that when offering the patient a treatment program, life expectancy does play an important part in helping to reduce the risk of over- and undertreatment. Our contribution to the overall staging of terminal patients,[10] which scientifically links the terminal phase to the previous phases of the disease, is the construction of a simple, tested tool to evaluate one quantitative aspect of terminal patients' classification.

In our study, CPS was used by clinicians in the hospices to which patients were referred by their doctors or by their families. It is interesting to note that even a variable as subjective and apparently unreproducible as the CPS proved predictive in the testing set, which contained several centers we had not recruited in the training set.

It might be objected that the physicians using CPS to calculate the PaP Score had both knowledge and sufficient experience in this field to make CPS predictive. An interesting point to be reflected on is whether the PaP Score also can be useful for personnel untrained in palliative care. For instance, can it

be used by referring physicians when suggesting hospice care? This might improve the timing of this care—not too early or too late—or make a patient's family not cherish unreasonable survival expectations.

In previous studies, palliative care experience is indeed a factor making CPS more predictive. This fact may be thought to constitute a weak spot in the score, implying the need for trained personnel in palliative care. Regardless, other variables are prognostically independent in the model containing CPS, and are able to provide more information on survival prediction. Thus, the remaining part of the PaP Score "corrects" possibly inaccurate CPS.

In the absence of PaP Score, CPS represents the only one instrument used by clinicians and families to predict the future progress of the illness and adopt the consequent care decisions. This may not be ideal. On the other hand, in this category of patients, there may be no need for biologically sophisticated prognostic factors. Easily obtainable "soft" prognostic factors may suffice. Nonetheless, the use of an integrated score is probably better. A validation study of a Pneumonia Prognostic Index carried out on a population of 14,199 patients[11] highlighted the need for a tried and tested tool to be "used in conjunction with, rather than in the place of, physician judgment". The PaP Score achieves this aim by incorporating physician judgment, corrected and integrated with a series of other objective parameters, into the framework of the score itself. The continuous integration of a subjective judgment with a series of more objective parameters, together with the attention to an evaluation on life expectancy by a comparison between CPS and actual survival, can provide ongoing education and training of clinicians in nonpalliative care settings, and contribute to improve prognostic abilities overall. Trials of the score in other settings such as acute care hospitals and in other countries are now indicated.

Acknowledgment

The co-authorship of the following group members is gratefully acknowledged: Luciano Frontini, MD and Anna Calcagno, MD (Servizio Ambulatoriale di Oncologia, Ospedale S. Paolo, Milano), Costanza Calia, MD (Fondazione FARO, Ospedale S. Giovanni Antica Sede, Torino), Attilio Gramazio, MD (Cattedra di Oncologia Medica, Università degli Studi di Ancona, Ancona), Andrea Rossi, MD, Paola Turci, MD (Servizio Oncologia Medica Ospedale Bufalini, Cesena [FO]) Stefania Derni, MD, Laura Fabbri, MD (Divisione di Oncologia Media, Ospedale Pierantoni, Forlì), Maria Vinci, MD and Gianfranco Giaccon, MD (Divisione di Oncologia Medica, Ospedale S. Carlo Borromeo, Milano), Raffaella Indelli, MD and Marina Marzola, MD (Divisione di Oncologia Medica, Ospedale S. Anna, Ferrara), Liliana Groff, MD, Augusto Caraceni, MD (Divisione Terapia del Dolore e Cure Palliative, Istituto Nazionale Tumori, Milano), Antonio Polselli, MD, Umberto Tonelli, MD (Divisione Oncologia Medica Ospedale degli Infermi Rimini); Luigi Montanari, MD (Divisione di Oncologia Medica, Ospedale Civile Lugo[RA].)

We are grateful to Mrs Grainne Tierney for the English translation of the manuscript. We also thank for their help in data collection: Rosanna Tedaldi, Luisa Guanella, Giuseppe Agazzi, Gianluigi Testa, Giovanna Zaninetta, and Roberto Pendola.

References

1. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer 1994; 69:979–985.

2. Marubini E, Valsecchi MG. Analysing survival data from clinical trials and observational studies. Chichester: John Wiley, 1995.

3. Fielding PL, Fenoglio-Preiser CM, Freedman LS. The future of prognostic factors in outcome prediction for patients with cancer. Cancer 1992;70: 2367–2377.

4. Pirovano M, Maltoni M, Nanni O, et al. A new palliative prognostic score (PaP Score): a first step for staging terminally ill cancer patients. (companion work).

5. Maltoni M, Pirovano M, Scarpi E, et al. Prediction of survival of patients terminally ill with cancer. Results of an Italian prospective multicenter study. Cancer 1995;75:2613–2622.

6. Maltoni M, Pirovano M, Nanni O, et al. Biological indices predictive of survival in 519 Italian terminally ill cancer patients. J Pain Symptom Manage 1997;13:1–10.

7. SAS Institute Inc. SAS/STAT User's Guide, version 6, 4th ed. Vol. 1. Cary, NC: SAS Institute Inc., 1989;943.

8. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. JAMA 1997;277:488–494.

9. Christakis NA, Escarce JJ. Survival of Medicare patients after enrollment in hospice programs. N Engl J Med 1996;335:172–178.

10. Toscani F. On behalf of the Italian Cooperative Research Group on Palliative Medicine: Classification and staging of terminal cancer patients: rationale and objectives of a multicentre cohort prospective study and methods used. Support Care Cancer 1996;4:56–60.

11. Fine MJ, Singer DE, Hanusa BH, et al. Validation of a pneumonia prognostic index using the Medisgroups Comparative Hospital Database. Am J Med 1993;94:153–159.