Prognostic models of survival in patients with advanced incurable cancer: the PiPS2 observational study

Patrick Stone,[1*] Anastasia Kalpakidou,[1] Chris Todd,[2] Jane Griffiths,[2] Vaughan Keeley,[3] Karen Spencer,[2] Peter Buckle,[1] Dori-Anne Finlay,[1] Victoria Vickerstaff[1] and Rumana Z Omar[4]

[1]Marie Curie Palliative Care Research Department, Division of Psychiatry, University College London, London, UK
[2]School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[3]Palliative Medicine Department, Derby Teaching Hospitals NHS Foundation Trust, Derby, UK
[4]Department of Statistical Science, University College London, London, UK

*Corresponding author p.stone@ucl.ac.uk

Declared competing interests of authors: Patrick Stone reports membership of the National Institute for Health Research Health Technology Assessment (HTA) End of Life Care and Add on Studies Board (2015–16). Rumana Omar is a member of the HTA General Committee (2018 to present).

Disclaimer: This report contains transcripts of interviews conducted in the course of the research and contains language that may offend some readers.

Published May 2021
DOI: 10.3310/hta25280

Scientific summary

The PiPS2 observational study

Health Technology Assessment 2021; Vol. 25: No. 28
DOI: 10.3310/hta25280

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

The Prognosis in Palliative care Study (PiPS) prognostic models were developed by members of our research team for patients with incurable cancer under specialist palliative care services. Separate prognostic models were created for patients without or with available blood results [called PiPS-A (Prognosis in Palliative care Study – All) and PiPS-B (Prognosis in Palliative care Study – Blood)]. Each version of PiPS consists of two prognostic models, one to predict 2-week (14-day) survival and one to predict 2-month (56-day) survival plus a 'decision rule' for determining how the model outputs should be interpreted. If the probability of surviving for 'days' (14 days) is > 50% and the probability of surviving for 'months+' (> 2 months) is < 50% then the patient is predicted to survive for 'weeks' (2 weeks to 2 months).

The primary purpose of this study was to validate PiPS-A and PiPS-B and to compare PiPS-B against clinicians' predictions of survival. The secondary purpose was to validate four other prognostic tools: the Palliative Prognostic Index, Palliative Performance Scale, Feliu Prognostic Nomogram and Palliative Prognostic score.

The Palliative Performance Scale is a measure of functional status that can be used to discriminate between groups with different survival prospects. The Palliative Prognostic Index predicts whether patients will live less than 3 or 6 weeks or more than 6 weeks. It is calculated using five clinical variables, one of which is the Palliative Performance Scale (the others are oral intake and presence or absence of dyspnoea, oedema and delirium). The Palliative Prognostic score stratifies patients into three risk groups depending on their probability of surviving 30 days (< 30%, 30–70% and > 70% probability). Palliative Prognostic score is calculated using six variables (clinician prediction of survival, Karnofsky performance status, presence of anorexia, dyspnoea, total white blood count and lymphocyte percentage). The Feliu Prognostic Nomogram uses five variables (Eastern Co-operative Oncology Group performance status, serum albumin concentration, lactate dehydrogenase concentration, lymphocyte counts and time from initial diagnosis to diagnosis of terminal disease) to predict survival at 15, 30 and 60 days.

It is also important to consider whether or not (and how) prognostic instruments will be incorporated into clinical practice. Current instruments are not widely used and it is therefore important to understand the potential barriers and facilitators in clinical practice.

Objectives

The overall aim of this research was the validation of models of survival in patients with advanced cancer.

The primary objective was to validate PiPS-A and PiPS-B and to compare PiPS-B with agreed multiprofessional estimates of survival.

1. The secondary objectives were to:

    - validate the Palliative Prognostic score, Feliu Prognostic Nomogram, Palliative Prognostic Index and Palliative Performance Scale
    - determine the acceptability of all prognostic models (including PiPS) to patients, carers and clinicians and to identify potential barriers to clinical use.

Methods

This was a national, multicentre, prospective, observational, cohort validation study of a prognostic tool with a nested qualitative substudy using face-to-face interviews with patients, carers and health-care professionals. The methods of the validation study and the embedded qualitative substudy are described separately.

Validation study

The validation study was a prospective observational cohort study of patients with advanced, incurable cancer. Both patients with and without capacity to consent were involved to maximise the clinical utility of the research. Patients with capacity provided written informed consent and patients without capacity were included if their next of kin provided written informed agreement.

The burden of data collection for participants was designed to be minimal. For patients without capacity to consent, all that was required was permission to access their notes and for the researcher to confer with the clinical team about their medical condition (i.e. primary diagnosis and extent of disease, performance status, presence or absence of key symptoms, general health and blood results). For patients with capacity to consent, we asked for an additional blood test and information about the severity and extent of their symptoms.

Clinical prediction of survival was also obtained. An attending clinician and nurse estimated survival independently. If estimates agreed about whether patients would survive for days (< 14 days), weeks (14–55 days) or months+ (≥ 56 days), then this was regarded as the agreed multiprofessional estimate of survival. If the estimates were different, then staff were asked to confer and reach a consensus.

These data were used to calculate the three prognostic scores that do not incorporate blood results (PiPS-A, Palliative Prognostic Index and Palliative Performance Scale) and the three prognostic scores that do require blood results (PiPS-B, Palliative Prognostic score and Feliu Prognostic Nomogram). At the end of data collection, dates of death for study participants were obtained from NHS Digital.

Nested qualitative substudy

Semistructured, face-to-face interviews were conducted with a purposive sample of patients, carers and clinicians. The patient and carer sample comprised patients and carers of patients with or without capacity who had agreed or who had declined to participate in the quantitative study. The final sample size was determined by data saturation.

Interviews used topic guides (see *Appendices 1* and *2*) developed in collaboration with service user representatives. Topic guides were based on reviews of the literature and the results of previous consultations with service users. Patients and carers were asked about perceived advantages and disadvantages of receiving prognostic information and how it should be presented to them. The clinician sample comprised health-care professionals who routinely make prognostic predictions: palliative care specialists, oncologists, nurses and general practitioners. Interviews were interactive and explored the acceptability of PiPS and other models. Clinicians were shown the prognostic models, experimented with them during the interview and commented on their perceived clinical usefulness (e.g. ease of completion and interpretability of outputs). They were asked about potential barriers to and facilitators of using the models and to discussing prognostic information with patients and carers.

Interviews were audio-recorded and fully transcribed. Interview data were entered into NVivo 10 (QSR International, Warrington, UK) and analysed using the five stages of framework analysis.

Results

The results of the validation study and the nested qualitative evaluation are presented separately.

Validation study

Between 1 August 2016 and 30 April 2018, 1833 patients (with capacity to consent, $n = 1610$; without capacity to consent, $n = 223$) were recruited at 27 sites in England and Wales. The mean age of patients was 70.2 years (standard deviation 11.9 years) and 938 (51.2%) were male. Patients came from inpatient palliative care units (1241; 67.7%), day hospices (169; 9.2%), community services (153; 8.4%), outpatients (146; 8.0%) and hospital palliative care teams (124; 6.8%).

PiPS-B

Discrimination of the PiPS-B 14-day model (PiPS-B14) and PiPS-B 56-day model (PiPS-B56) was evaluated using the *c*-statistic (PiPS-B14 0.837, 95% confidence interval 0.810 to 0.863; PiPSB-56 0.810, 95% confidence interval 0.788 to 0.832). Calibration was assessed using the calibration slope based on a logistic regression model fitted to the validation data using the predicted log-odds as the only predictor (PiPS-B14 0.781, 95% confidence interval 0.676 to 0.886; PiPS-B56 0.914, 95% confidence interval 0.808 to 1.02). The *c*-statistic for the PiPS-B decision rule to discriminate between participants who died within days (< 14 days) and those who survived ≥ 14 days was 0.631 (95% confidence interval 0.602 to 0.659). The *c*-statistic for the PiPS-B decision rule to discriminate between those who died within 56 days and those who survived ≥ 56 days was 0.735 (95% confidence interval 0.713 to 0.757). When PiPS-B14 and PiPS-B56 were combined using the decision rule, there was no significant difference ($p = 0.851$) in the accuracy of PiPS-B predictions (910/1484; 61.3%) and the agreed multiprofessional estimate of survival (914/1484; 61.6%) when compared with observed survival time of the patients.

PiPS-A

Discrimination of the PiPS-A 14-day model (PiPS-B14) and PiPS-A 56-day model (PiPS-B56) was evaluated using the *c*-statistic (PiPS-A14 0.825, 95% confidence interval 0.803 to 0.848; PiPS-A56 0.776, 95% confidence interval 0.755 to 0.797). Calibration was assessed using the calibration slope (PiPS-A14 0.981, 95% confidence interval 0.872 to 1.09; PiPS-A56 0.946, 95% confidence interval 0.842 to 1.05). The *c*-statistic for the PiPS-A decision rule to discriminate between participants who died within days (< 14 days) and those who survived ≥ 14 days was 0.680 (95% confidence interval 0.655 to 0.705). The *c*-statistic for the PiPS-A decision rule to discriminate between those who died within 56 days and those who survived ≥ 56 days was 0.687 (95% confidence interval 0.666 to 0.708). When PiPS-A14 and PiPS-A56 were combined using the decision rule, PiPS-A was significantly less accurate than the agreed multiprofessional estimate of survival [PiPS-A accuracy of 1012/1802 (56.2%) vs. agreed multiprofessional estimate of survival accuracy of 1117/1802 (62.0%; $p < 0.001$)] when compared with the observed survival time of the patients.

Palliative Prognostic score

The median survival for each Palliative Prognostic score risk group was 121 days (interquartile range 49–289 days), 28 days (interquartile range 14–60 days) and 7 days (interquartile range 4–19 days), respectively. The proportion of patients surviving 30 days in each risk category was 687 out of 794 (86.5%), 306 out of 654 (46.8%) and 22 out of 143 (15.4%), respectively. Owing to the nature of the probabilistic predictions provided by Palliative Prognostic score, it was not possible to make a direct comparison between its performance and that of the clinicians.

Feliu Prognostic Nomogram

The Feliu Prognostic Nomogram predicts probability of survival at 15, 30 and 60 days. The discriminatory ability of the Feliu Prognostic Nomogram was evaluated using the *c*-index (0.684, 95% confidence interval 0.669 to 0.700). The calibration of the Feliu Prognostic Nomogram was assessed using the calibration slope (1.049, 95% confidence interval 0.939 to 1.158). Owing to the nature of the probabilistic predictions provided by the Feliu Prognostic Nomogram, it was not possible to make a direct comparison between its performance and that of the clinicians.

Palliative Prognostic Index

The median survival for each Palliative Prognostic Index risk group was 16 days (interquartile range 5–52 days), 38 days (interquartile range 15–106 days) and 79 days (interquartile range 32–219 days), respectively. The Palliative Prognostic Index was significantly less accurate than the agreed multiprofessional estimate of survival [Palliative Prognostic Index accuracy of 990/1828 (54.2%) vs. agreed multiprofessional estimate of survival accuracy of 1143/1828 (62.5%); $p < 0.001$] when compared with the observed survival time in the data.

Palliative Performance Scale

The median survival of each Palliative Performance Scale risk group was as follows: Palliative Performance Scale-10, 2 days (interquartile range 1–4 days); Palliative Performance Scale-20, 6 days (interquartile range 3–16 days); Palliative Performance Scale-30, 20 days (interquartile range 7–42 days); Palliative Performance Scale-40, 24 days (interquartile range 10–69 days); Palliative Performance Scale-50, 40 days (interquartile range 19–97 days); Palliative Performance Scale-60, 65 days (interquartile range 28–172 days); Palliative Performance Scale-70, 99 days (interquartile range 44–284 days); Palliative Performance Scale-80, 186 days (interquartile range 85–477 days); and Palliative Performance Scale-90, 252 days (interquartile range 135–568 days). Palliative Performance Scale-100 was not calculated as numbers were too low.

Nested qualitative substudy

Semistructured interviews were conducted with a purposive sample of patients ($n = 29$), carers ($n = 20$) and clinicians ($n = 32$) between April 2017 and July 2018. Interview data were analysed using the five stages of framework analysis: familiarisation, developing a thematic framework, indexing, charting, and mapping and interpretation.

Patients and carers

The majority of patients (25/29) and carers (17/20) were recruited from two hospices based in the Greater Manchester area. Five of the patients interviewed had declined to take part in the clinical study and their reasons for refusal were explored during the interview. The three following themes were identified in interview analysis.

Desire for detailed patient prognostic information

The majority of participants expressed a desire for detailed information from doctors about patient life expectancy. Few received this information, however, and when they did it was not expressed clearly. A number of patients who had asked about their life expectancy commented that doctors tended to be vague, overoptimistic and unwilling to deliver news that was considered bad or uncertain.

Acceptability of PiPS predictor models

All participants considered that the PiPS models were acceptable for use in clinical practice and (if accurate) could help clinicians more accurately predict life expectancy. They also considered that PiPS could help doctors initiate sensitive conversations about prognostication.

Preferred presentation of sensitive information

Patients and carers agreed that the most appropriate way of receiving prognostic information was during face-to-face discussions and that this should be verbal rather than written. Participants preferred life expectancy to be presented in terms of days, weeks or months.

Clinicians

Clinicians included palliative care specialists, oncologists, general practitioners and nurses. Interviews were interactive and participants were shown the PiPS models and other prognostic tools. The following seven main themes arose from the interviews.

Challenges and difficulties of predicting length of survival

All clinicians commented on the complex nature of estimating length of survival for patients. Some commented on how challenging they found predicting length of survival for patients with advanced cancer.

Language used when discussing prognosis

The majority of clinicians explained that they tended to avoid giving detailed prognostication with specific time frames. Clinicians said that this was because they did not know the answer or they did not want the discussion to have a negative impact on the patient or carer.

Reasons for overestimating prognosis

Clinicians said that the main reason that they preferred to convey kinder, optimistic information was that it was perceived as less harmful. Clinicians considered that it was better not to challenge patients' or carers' perceptions about the disease trajectory and described not wanting to take away hope.

Acceptability of PiPS predictor models

The majority of clinicians considered PiPS to be a useful algorithm that could offer a more objective approach to estimating a patient's prognosis. Even if PiPS was no more accurate than clinicians' estimates, they stated that PiPS may be a beneficial tool that could help improve their confidence in making and discussing survival predictions.

Facilitators of PiPS use in clinical practice

All clinicians commented that the PiPS prognosticator was user friendly. Participants further commented on how PiPS could help to inform decision-making about treatment options and/or discharge planning. They suggested that PiPS could be helpful when commissioning care packages for patients.

Barriers to PiPS use in clinical practice

A minority of participants considered it inappropriate to use PiPS, especially if blood tests were needed to improve the accuracy of the prediction. Other barriers related to clinicians' preference for relying on their own clinical judgement and clinicians' avoidance of prognostic discussions with patients and carers.

Clinicians' attitudes to existing prognostic models

Clinicians considered the other prognostics tools to be not very user friendly, vague, cumbersome, difficult to use and dated.

Conclusions

Both PiPS-B14 and PiPS-B56 demonstrated excellent discrimination. PiPS-B56 was well calibrated but PiPS-B14 showed a degree of overfitting. PiPS-B14 and PiPS-B56, combined using the decision rule, perform as well as an agreed multiprofessional estimate of survival. PiPS-B has greater discriminatory ability than the Palliative Prognostic Index or the Feliu Prognostic Nomogram and is more accurate than the Palliative Prognostic Index when compared with clinicians. PiPS-A14 and PiPS-A56, combined using the decision rule, perform significantly worse than an agreed multiprofessional estimate of survival.

It is clear that PiPS is regarded by patients, carers and health-care professionals as an acceptable way to determine prognosis. Even if no more accurate than clinicians' estimates, health-care professionals stated that PiPS could help improve their confidence in making survival predictions.

Future research might improve the accuracy of PiPS by recalibrating PiPS-B14 and adjusting the thresholds for the decision rules for determining outcomes and by considering the inclusion of additional variables. An impact study will be required to evaluate whether or not use of PiPS can improve clinical outcomes.

Trial registration

This trial is registered as ISRCTN13688211.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 28. See the NIHR Journals Library website for further project information.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.370

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 13/20/01. The contractual start date was in May 2016. The draft report began editorial review in August 2019 and was accepted for publication in March 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).