**The Prognosis in Palliative care Study II (PiPS2)**

**Statistical Analysis Plan**

V 1.0

Started: 19th Oct 2017

Edited: 13 Nov 2018

**Version History Log**

|  |  |  |
| --- | --- | --- |
| Version | Date Implemented | Details of change |
| 0.1 | Oct 2017 |  |
| 0.2 | 10 Jan 2018 | Rumana Omar and Paddy Stone provided feedback |
| 0.3 | 29 Jan 2018 | Paddy Stone and Anastasia Kalpakidou provided feedback |
| 0.4 | 06 Aug 2018 | Susan Charman provided feedback. |
| 1.0 | 13 Nov 2018 | Signed and authorised by team |

**SAP Authorisation**

|  |  |
| --- | --- |
| Author | |
| Name: | Ms Victoria Vickerstaff |
| Role: | Trial statistician |
| Signature: | **\\ad.ucl.ac.uk\homev\ucakvhv\Documents\signature nearly.png** |
| Date: | 13/11/2018 |

|  |  |
| --- | --- |
| Approver | |
| Name: | Prof. Rumana Omar |
| Role: | Lead statistician |
| Signature: | sig |
| Date: | 16/11/2018 |

|  |  |
| --- | --- |
| Approver | |
| Name: | Prof. Paddy Stone |
| Role: | Chief investigator |
| Signature: |  |
| Date: | 13-Nov-2018 |



Contents

[1. Study Summary 3](#_Toc529967092)

[2. Analysis Plan Summary 4](#_Toc529967093)

[3. Introduction 5](#_Toc529967094)

[3.1. Purpose and scope of the statistical analysis plan 5](#_Toc529967095)

[3.2. Analysis organisation 5](#_Toc529967096)

[3.3. Data checking 5](#_Toc529967097)

[4. Data collection 5](#_Toc529967098)

[4.1. Primary outcome measures 5](#_Toc529967099)

[4.2. Secondary outcome measures 5](#_Toc529967100)

[5. Data analysis plan 6](#_Toc529967101)

[5.1. Descriptive analysis 6](#_Toc529967102)

[5.2. Primary analysis 6](#_Toc529967103)

[5.3. Secondary analysis 7](#_Toc529967104)

[5.4. Sensitivity and other planned analyses 8](#_Toc529967105)

[6. Software 8](#_Toc529967106)

[7. References 8](#_Toc529967107)

# Study Summary

|  |  |
| --- | --- |
| For full details see the protocol (version 3.0 21.08.2017). | |
| **Title** | The Prognosis in Palliative care Study II (PiPS2): A multicentre prospective, observational, validation cohort study |
| **Short title** | The Prognosis in Palliative care Study II (PiPS2) |
| **Chief Investigator** | Patrick Stone |
| **Statisticians** | Senior statistician: Prof Rumana Omar  Study statistician: MsVictoria Vickerstaff |
| **Phase of study** | Phase III prognostic validation study |
| **Primary objective** | To validate PiPS-A&B and to compare PIPS-B against clinicians’ predictions of survival |
| **Primary Outcome Measure** | The primary outcomes will be the survival of the participants (measured from date of study entry) and the predictions of the PiPS-B prognostic model. |
| **Population** | Patients with locally advanced or metastatic, incurable cancer recently referred to the participating palliative care service. Patients with or without capacity are included in the study. |
| **Sample size** | To show at least a 5% improvement in correct predictions (in terms of overall agreement with observed patient survival giving an odds ratio of 1.28) when using the model compared to clinicians’ predictions, assuming 80% power and 5% significance level and using a McNemar’s test, a total of 1,267 patients will be required. We will recruit patients until we have obtained 1,267 participants with blood results (the validation set for PiPS-B). We estimate that in order to obtain 1,267 complete data sets we will need to recruit 1,334 patients with capacity (this assumes 5% missing data). In order to recruit 1,334 patients with capacity (without restricting our focus to those who are able to provide blood results or whom have blood results available) we will need to recruit 1,778 patients in total (this assumes 25% of total patients included in the study will lack capacity and will be unable to provide a blood test). |
| **Clinical Trials Gov** | ISRCTN13688211 |
|  |  |

# Analysis Plan Summary

**Descriptive analysis**

Predictors and the outcomes will be summarised using descriptive analysis. Categorical predictors will be reported as raw numbers and percentages. Continuous variables will be summarised using mean or median and standard deviation or interquartile (IQ) range as appropriate. The percentage of values missing for each predictor will also be presented. The survival times of patients will be summarised using median and IQ ranges, and Kaplan Meier graphs.

**Validation of PiPS models**

The discriminatory ability of the models will be assessed using the C-statistic. Separate C-statistics will be calculated for the “two weeks” and the “two months” models. We will also assess model performance by plotting Kaplan-Meier survival curves for each of the three risk groups identified by the PiPs models (“days,” “weeks,” and “months+”). Model calibration will be assessed by comparing observed and predicted probabilities.

**Comparison between PiPS model and clinician predictions**

To compare the accuracy of the model and clinicians’ predictions, the primary analysis will focus on the PiPS-B model. McNemar’s test will be used to compare the proportion of overall patient deaths predicted correctly by PIPS-B with the corresponding proportion predicted correctly by clinicians.

# Introduction

## Purpose and scope of the statistical analysis plan

This document describes the main statistical analyses to be applied to the data from PIPSII study. This Statistical Analysis Plan was written by Victoria Vickerstaff and Rumana Omar. The plan was agreed by the Study Management Group.

## Analysis organisation

The final analysis will be performed after the dates of death have been obtained from NHS Digital three months after completion of the study. Prior to the final analysis, all relevant data will have been entered, checked and locked, the analysis plan will have been finalised and approved, and the analysis programs prepared as much as possible. The primary analysis will be performed independently by two statisticians (VV and RO) to ensure its accuracy.

## Data checking

Before analysis, basic checks will be performed to confirm the quality of the data. Incomplete or inconsistent data include:

* Missing data
* Data outside expected range
* Other inconsistencies between variables e.g., in the dates the questionnaires were completed

If any inconsistencies are found, the corresponding values will be double checked with the researchers and corrected if necessary. All changes will be documented by the study statistician.

# Data collection

## Primary outcome measures

The primary outcomes of interest will be the survival of the participants (measured from date of study entry) and the predictions of the Prognosis in Palliative care Score - A (PiPS-A) and the Prognosis in Palliative care Score - B (PiPS-B) prognostic models. Both models provide a prediction about whether a patient is likely to live for “days” (less than 14-days), “weeks” (2 to 7 weeks), or “months +” (2 months or more).

## Secondary outcome measures

The secondary outcomes will be the predictions produced by:

* The Palliative Prognostic Index, PPI (less than 3 week survival, 3 to 6 week survival, and greater than 6 week survival);
* The Palliative Performance Scale, PPS (probability of dying within 7, 14 or 28 days);
* Feliu Prognostic Nomogram, FPN (risk of dying within 15, 30 or 60 days);
* Palliative Prognostic score, PaP (risk of dying within 30 days).

# Data analysis plan

## Descriptive analysis

Initially the predictors and the outcome will be summarised using descriptive analysis. Categorical predictors shall be reported as raw numbers and percentages. Reports of continuous variables shall include mean or median and standard deviation or interquartile (IQ) range as appropriate. The percentage of values missing for each predictor will also be presented. The survival times of patients will be summarised using median and IQ ranges and Kaplan Meier graphs.

## Primary analysis

**Validation of PiPS models**

We will validate the PiPS models as they were presented for use in the original study by Gwilliam et. al [1]. For both PiPS-A and PiPS-B, two separate models have been developed to predict the two week (14 day) and two month (56 day) survival of patients (thus generating three prognostic categories; less than two weeks, two weeks to two months and greater than two months). The week and month models include different sets of predictors. For both models (weeks and months), if the predicted probability of the event exceeded 50% for a patient, then the patient was classified to have the event. Otherwise, it was assumed that the patient did not have the event. Thus if, for example, the models predicted that a patient would survive two weeks, but predicted that the patient would die within two months, then the PiPS model outcome would be that the patient was predicted to die in “weeks”.

For the primary analysis, patients with complete data sets for validation of PiPS-B will be used, if the proportion of missing data is less than 10%. Otherwise, multiple imputation based on chained equations [9] will be used to impute missing predictor values. The imputation model will include the predictors in the PiPS-B risk model, predictors of missingness and possibly the centre effect (if there is considerable variation across centres as measured by the ICC or the descriptive statistics of the estimate of the between centre variance). The centre effect will be incorporated either including fixed effects for centres, or creating groups of centres with similar characteristics and including these groups as fixed effects, or using the estimate of the random effects obtained from fitting the PiPS-B model to the data. We will either use Rubin’s rule or use the stacked approach to combine the performance measures from the imputed datasets (Wood et al . 2015).

The discriminatory ability of the models will be assessed using the C-statistic. Separate C-statistics will be calculated for the “two weeks” and the “two month” models. The C-statistic will be estimated by forming all patient pairs and calculating the proportion of patient pairs where the patient who has the event has the higher predicted value. The PiPS online calculator provides (see www.pips.sgul.ac.uk) a prediction as to whether a patient will survive for days, weeks or months. Additional details of the models used to create the survival predictions are provided in the Appendix. The model calibration will be assessed by comparing the observed and the predicted proportions for each of these categories. The calibration of the prognostic models will be further assessed using the calibration intercept (calibration at large) and slope based on a logistic regression model fitted to the validation data using the predicted log-odds as the only predictor [2]. This will also be done separately for the “two weeks” and the “two month” models. The calibration intercept and slope, and the C-statistic will initially be estimated without taking account of potential patient clustering within centres. In a second analysis, these performance measures will be calculated for each centre separately (assuming most centres have sufficient number of events to allow such calculations) and the estimates pooled across centres using a weighted average [3]. The calibration intercept and slope, and the C-statistic will be presented as estimates with confidence intervals.

We will also assess model performance by plotting Kaplan-Meier survival curves for each of the three risk groups identified by the PiPS models (“days,” “weeks,” and “months+”).

**Comparison between PiPS model and clinician predictions**

To compare the accuracy of the model and clinicians’ predictions, the primary analysis will focus on the PiPS-B model. McNemar’s test will be used to compare the proportion of overall patient deaths predicted correctly by PIPS-B with the corresponding proportion predicted correctly by clinicians. Table 1 (below) will form the basis of this comparison.

Table 1 – Blank table for analysis comparing PiPS-B predictions with clinicians’ predictions

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Clinician predictions compared to observed deaths** | |  |
| **Risk model predictions compared to observed deaths** | Number of patients when clinician’s prediction was correct | Number of patients when clinician’s prediction was incorrect | **Total** |
| Number of patients when PiPS prediction was correct |  |  |  |
| Number of patients when PiPS prediction was incorrect |  |  |  |
| Total |  |  |  |

## Secondary analysis

As part of the secondary analyses we will combine the models’ predictions for the two week and two month cut-off points to produce a categorical prediction of survival (“days,” “weeks,” or “months/years”) and compare with clinicians’ estimates and the corresponding observed values descriptively with respect to their accuracy . Table 2 (below) will be used for this descriptive comparison (the cells will contain the counts in each category).

Table 2 – Blank table for analysis of secondary outcomes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Actual Survival** | | | **Total** |
| **Clinician Predictions** | Days | Weeks | Months |  |
| Days |  |  |  |  |
| Weeks |  |  |  |  |
| Months |  |  |  |  |
|  | **Actual Survival** | | |  |
| **Model Predictions** | Days | Weeks | Months |  |
| Days |  |  |  |  |
| Weeks |  |  |  |  |
| Months |  |  |  |  |
|  |  |  |  |  |
| **Total** |  |  |  |  |

Linear weighted κ will be also used to compare the performance of the clinicians with that of the models. If appropriate we will also consider using the net reclassification index (NRI) as part of this secondary analysis to compare clinician and model predictions, noting that NRI needs to be used with caution, particularly when there are three or more risk categories [4-7].

As part of the secondary analyses, the other risk models (PaP, FPN, PPI and PPS) will also be validated. The calibration of these prognostic models will be assessed using the calibration slope [2] based on a logistic model for binary outcomes and Cox model for survival outcomes [8]. Graphical comparisons of the observed and predicted risks for clinically relevant patient risk groups will also be made. Clinically relevant time points will be used for comparisons for survival outcomes. Model discrimination will be assessed using the C-statistic for binary outcomes and C-index for survival outcomes [7]. The predictions made by the other prognostic models under evaluation in this project will also be compared with the corresponding observed outcomes and clinician predictions (where available). Potential missing data in predictor values will be handled as described.

## Sensitivity and other planned analyses

Characteristics of patients with missing data will be compared with those with complete information to investigate any bias.

We will also assess the model performance of the Glasgow Prognositc Score, GPS. The GPS categorises patients into three groups depending on prognosis (‘GPS0’, ‘GPS1’ and ‘GPS2’). Similar analyses will performed as those used for the secondary outcome measures.

# Software

The statisticians will download the data from the study specific online database provided by Sealed Envelope into a format suitable for Stata. All the statistical analysis will be performed using Stata version 14 (or above).

# References

1. Gwilliam, B., et al., *Development of prognosis in palliative care study (PiPS) predictor models to improve prognostication in advanced cancer: prospective cohort study.* BMJ, 2011. **343**: p. d4920.
2. Miller, M., et al., *Validation of probabilistic predictions.* Med Decis Making, 1993(Jan-Mar; 13(1)): p. 49-58.
3. Van Oirbeek, R. and E. Lesaffre, *An application of Harrell's C-index to PH frailty models. .* Stat Med, 2010. **29**(30): p. 3160-71.
4. Leening, M.J., et al., *Net reclassification improvement and integrated discrimination improvement require calibrated models: relevance from a marker and model perspective.* Statistics in Medicine, 2014. **33**(19): p. 3415-8.
5. Pencina, K.M., M.J. Pencina, and R.B. D'Agostino, Sr., *What to expect from net reclassification improvement with three categories.* Statistics in Medicine, 2014. **33**(28): p. 4975-87.
6. Kerr, K.F., et al., *Net reclassification indices for evaluating risk prediction instruments: a critical review.* Epidemiology, 2014. **25**(1): p. 114-21.
7. Steyerberg, E., et al., *Assessing the performance of prediction models: a framework for some traditional and novel measures.* Epidemiology, 2010. **21**(1): p. 128-138.
8. Steyerberg, E., *Clinical Prediction Models. A practical Approach to development, validation and updating.* . 2009, Rotterdam: Springer.
9. Wood AM, Royston P, White IR. *The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data*. Biometrical Journal. 2015 Jul 1;57(4):614-32
10. .White, I., P. Royston, and A. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Statistics in Medicine, 2011 **30**: p. 377 - 399.

**Appendix**

For both PiPS-A and PiPS-B, two separate models have been developed to predict the two week (14 day) and two month (56 day) survival of patients (thus generating three prognostic categories; less than two weeks, two weeks to two months and greater than two months). The week and month models include different sets of predictors. For both models (weeks and months), if the predicted probability of the event exceeded 50% for a patient, then the patient was classified to have the event. Otherwise, it was assumed that the patient did not have the event. Thus if, for example, the models predicted that a patient would survive two weeks, but predicted that the patient would die within two months, then the PiPS model outcome would be that the patient was predicted to die in “weeks”.

**PiPS-A two week (14 day)**

The PiPS-A two week (14 day) log odds are

where

|  |  |
| --- | --- |
| amts | AMTS score (If ≤3 then = 0 , if >3 then = 1) |
| pulse | Pulse rate |
| distant\_mets | Presence of distant metastases (No = 0, Yes = 1) |
| mets\_liver | Presence of liver metastases (No = 0, Yes = 1) |
| ecog | Eastern Co-operative Oncology Group score |
| overall\_health | Global Health Score |
| anorexia | Anorexia (No = 0, Yes = 1) |
| mets\_bone | Presence of bone metastases (No = 0, Yes = 1) |
| dyspnoea | Dyspnoea (No = 0, Yes = 1) |
| dyshpagia | Dyshpagia (No = 0, Yes = 1) |

The corresponding probability of survival for the PiPS-A two week (14 day) model is

**PiPS-A two month (56 day)**

The PiPS-A two month (56 day) log odds are

where

|  |  |
| --- | --- |
| amts | AMTS score (If ≤3 then = 0 , if >3 then = 1) |
| pulse | Pulse rate |
| distant\_mets | Presence of distant metastases (No = 0, Yes = 1) |
| mets\_liver | Presence of liver metastases (No = 0, Yes = 1) |
| ecog | Eastern Co-operative Oncology Group score |
| overall\_health | Global Health Score |
| Primary\_breast | Primary cancer breast (No = 0, Yes = 1) |
| Mgo\_cancer | Primary cancer MGO (No = 0, Yes = 1) |
| Lost weight | Lost weight (No = 0, Yes = 1) |

The corresponding probability of survival for the PiPS-A two month (56 day) model is:

**PiPS-B two week (14 day)**

The PiPS-B two week (14 day) log odds are:

where

|  |  |
| --- | --- |
| amts | AMTS score (If ≤3 then = 0 , if >3 then = 1) |
| pulse | Pulse rate |
| distant\_mets | Presence of distant metastases (No = 0, Yes = 1) |
| mets\_bone | Presence of bone metastases (No = 0, Yes = 1) |
| Anorexia | Anorexia (No = 0, Yes = 1) |
| Ecog | Eastern Co-operative Oncology Group score |
| Overall\_health | Global Health Score |
| Wbc | WBC (x10^9/L) |
| Platlet | Platelet count (x10^9/L) |
| urea | Urea (mmol/L) |
| alanine | ALT (U/L) |
| creactive | CRP (mg/L) |

The corresponding probability of survival for the PiPS-B two week (14 day) model is:

**PiPS-B two month (56 day)**

The PiPS-B two month (56 day) log odds are:

where

|  |  |
| --- | --- |
| Pulse | Pulse rate |
| Wbc | WBC (x10^9/L) |
| Platlet | Platelet count (x10^9/L) |
| Neutrophil | neutrophils (x10^9/L) |
| lymphocyte\_10exp9 | Lymphocytes (x10^9/L) |
| Urea | Urea (mmol/L) |
| Alkaline | Alk Phos (U/L) |
| Creactive | CRP (mg/L) |
| Mgo\_cancer | Primary cancer MGO (No = 0, Yes = 1) |
| fatigue | Fatigue (No = 0, Yes = 1) |
| Overall\_health | Global Health Score |

The corresponding probability of survival for the PiPS-B two month (56 day) model is: