

**Keywords:** EPIPHANY Index; cancer; clinical decision rule; mortality; prognostic scales; pulmonary embolism; incidental; risk

# Predicting serious complications in patients with cancer and pulmonary embolism using decision tree modelling: the EPIPHANY Index

A Carmona-Bayonas<sup>\*1</sup>, P Jiménez-Fonseca<sup>2</sup>, C Font<sup>3</sup>, F Fenoy<sup>4</sup>, R Otero<sup>5</sup>, C Beato<sup>6</sup>, J M Plasencia<sup>7</sup>, M Biosca<sup>8</sup>, M Sánchez<sup>9</sup>, M Benegas<sup>9</sup>, D Calvo-Temprano<sup>10</sup>, D Varona<sup>11</sup>, L Faez<sup>2</sup>, I de la Haba<sup>12</sup>, M Antonio<sup>13</sup>, O Madridano<sup>14</sup>, M P Solis<sup>2</sup>, A Ramchandani<sup>15</sup>, E Castañón<sup>16</sup>, P J Marchena<sup>17</sup>, M Martín<sup>14</sup>, F Ayala de la Peña<sup>1</sup> and V Vicente<sup>1</sup> on behalf of the Asociación de Investigación de la Enfermedad Tromboembólica de la Región de Murcia (the Region of Murcia's Association of Thromboembolic Disease Research)

<sup>1</sup>Department of Hematology and Medical Oncology, Hospital Universitario Morales Meseguer, Instituto Murciano de Investigación Biosanitaria (IMIB), Murcia, Spain; <sup>2</sup>Department of Medical Oncology, H. Universitario Central de Asturias, Oviedo, Spain; <sup>3</sup>Department of Medical Oncology, H. Universitario Clinic, Barcelona, Spain; <sup>4</sup>Department of Physiology, Facultad de Medicina, Universidad de Murcia, Murcia, Spain; <sup>5</sup>Department of Pneumology, H. Universitario Virgen del Rocío, CIBER de Enfermedades Respiratorias, Seville, Spain; <sup>6</sup>Department of Medical Oncology, Hospital Nisa-Aljarafe, Seville, Spain; <sup>7</sup>Department of Radiology, Hospital Universitario Morales Meseguer, Murcia, Spain; <sup>8</sup>Department of Medical Oncology, H. Universitario Vall d'Hebron, Barcelona, Spain; <sup>9</sup>Department of Radiology, H. Universitario Clinic, Barcelona, Spain; <sup>10</sup>Department of Radiology, H. Universitario Central de Asturias, Oviedo, Spain; <sup>11</sup>Department of Radiology, H. Universitario Vall D' Hebron, Barcelona, Spain; <sup>12</sup>Department of Internal Medicine, ICO Duran i Reynals, Barcelona, Spain; <sup>13</sup>Department of Medical Oncology, ICO Duran i Reynals, Barcelona, Spain; <sup>14</sup>Department of Internal Medicine, Hospital Universitario Infanta Sofía, San Sebastián de los Reyes, Madrid, Spain; <sup>15</sup>Department of Medical Oncology, Hospital Universitario Insular de Gran Canaria, Las Palmas, Spain; <sup>16</sup>Department of Medical Oncology, Clínica Universidad de Navarra, Pamplona, Spain and <sup>17</sup>Department of Internal Medicine, Parc Sanitari Sant Joan of Déu, Esplugues de Llobregat, Barcelona, Spain

**Background:** Our objective was to develop a prognostic stratification tool that enables patients with cancer and pulmonary embolism (PE), whether incidental or symptomatic, to be classified according to the risk of serious complications within 15 days.

**Methods:** The sample comprised cases from a national registry of pulmonary thromboembolism in patients with cancer (1075 patients from 14 Spanish centres). Diagnosis was incidental in 53.5% of the events in this registry. The Exhaustive CHAID analysis was applied with 10-fold cross-validation to predict development of serious complications following PE diagnosis.

**Results:** About 208 patients (19.3%, 95% confidence interval (CI), 17.1–21.8%) developed a serious complication after PE diagnosis. The 15-day mortality rate was 10.1%, (95% CI, 8.4–12.1%). The decision tree detected six explanatory covariates: Hestia-like clinical decision rule (any risk criterion present vs none), Eastern Cooperative Group performance scale (ECOG-PS; <2 vs ≥2), O<sub>2</sub> saturation (<90 vs ≥90%), presence of PE-specific symptoms, tumour response (progression, unknown, or not evaluated vs others), and primary tumour resection. Three risk classes were created (low, intermediate, and high risk). The risk of serious complications within 15 days increases according to the group: 1.6, 9.4, 30.6%;  $P < 0.0001$ . Fifteen-day mortality rates also rise progressively in low-, intermediate-, and high-risk patients: 0.3, 6.1, and 17.1%;  $P < 0.0001$ . The cross-validated risk estimate is 0.191 (s.e. = 0.012). The optimism-corrected area under the receiver operating characteristic curve is 0.779 (95% CI, 0.717–0.840).

**Conclusions:** We have developed and internally validated a prognostic index to predict serious complications with the potential to impact decision-making in patients with cancer and PE.

\*Correspondence: Dr A Carmona-Bayonas; E-mail: alberto.carmonabayonas@gmail.com

Received 7 October 2016; revised 6 February 2017; accepted 7 February 2017; published online 7 March 2017

© 2017 Cancer Research UK. All rights reserved 0007–0920/17

Pulmonary embolism (PE) is one of the most common and feared complications in cancer patients, given its frequency and the suffering it entails (Sørensen *et al*, 2000). There is a 15–30% prevalence rate of PE in necroscopic series for the most thrombogenic tumours, owing to interactions between the mechanisms of tumorigenesis, haemostatic activation, and other factors (Svendsen and Karwinski, 1989). The introduction of multidetector computed tomography (CT) has increased detection rates of incidental PE, present in 2–8% of the studies performed in cancer patients (Dentali *et al*, 2010). In some recent series, incidentally diagnosed PE accounted for a ~50% of embolic events (Font *et al*, 2014, 2016). On the other end of the spectrum of severity, PE is also a common cause of fatal events in daily practice, as well as in trials with new targeted therapies (Ranpura *et al*, 2011; Den Exter *et al*, 2013).

Classifying prognosis in PE is important, in as much as episodes classified as low risk might be eligible for support reductions (e.g., outpatient management or early discharge, etc.), thereby lowering costs and enhancing patient comfort without compromising safety. In contrast, subjects at higher risk should receive stepped up care or monitoring (Streiff *et al*, 2013). Different studies have identified several prognostic factors for cancer-associated symptomatic PE, most decisive among them being the presence of metastasis, immobilisation, low weight, or altered vital signs (Kline *et al*, 2012; Den Exter *et al*, 2013; Font *et al*, 2014, 2016). Several prospective, cohort studies have based selection of low-risk patients eligible for outpatient treatment on pragmatic clinical decision rules (CDR), such as the HESTIA study eligibility criteria, which are based prominently on altered vital signs and risk of bleeding (Siragusa *et al*, 2005; Zondag *et al*, 2011; Font *et al*, 2014; Weeda *et al*, 2016).

On the other hand, prognostic multivariate models have been created, such as the RIETE registry scale and POMPE-C score, that predict 30-day mortality probability following PE (Kline *et al*, 2012; Den Exter *et al*, 2013); although, at best, they are marginally superior to other classifications developed for PE in the general population (e.g., PESI or sPESI; Carmona-Bayonas *et al*, 2016). Nevertheless, using any of them to assist in decision-making involves problems, not least of which is that their suitability for incidental PE has yet to be proven (Kline *et al*, 2012; Den Exter *et al*, 2013). Furthermore, they are not sensitive to competitive risks, such as increased bleeding, responsible for some 10% of early mortality, or cancer progression, which accounts for 50% of 30-day mortality after a PE event (Den Exter *et al*, 2013; Carmona-Bayonas *et al*, 2016).

Consequently, there is no adequate prognostic stratification method for incidental and symptomatic PE. In this study, we have attempted to refine the classification of the entire spectrum of cancer-associated PE by combining an adaptation of the HESTIA criteria with other explanatory covariates and modelling a decision tree procedure.

## MATERIALS AND METHODS

**Patients.** The source of information is an observational registry of consecutive cases of cancer-associated PE, who received care at 14 Spanish hospitals between 2004 and 2015 (*Registro de Embolia Pulmonar en Pacientes con Neoplasias*, EPIPHANY registry for its Spanish acronym). This registry's design, methods, and characteristics have been previously reported in depth (Carmona-Bayonas *et al*, 2016; Font *et al*, 2016; Plasencia-Martínez *et al*, 2016). Briefly put, the basic eligibility criteria required that patients be adults ( $\geq 18$  years) with a PE diagnosis confirmed by means of objective imaging (CT angiography scans, high probability scintigraphy, or CT scheduled to assess tumour response or for other reasons). In order to choose a truly oncological population, subjects were

withdrawn from the study if the PE had occurred more than 1 month prior to the diagnosis of cancer, or if more than 1 month had elapsed since completing adjuvant chemotherapy. Patients were also excluded if they had not received anticoagulant therapy without justification according to international clinical practice guidelines (Streiff *et al*, 2013).

Given that the study included a prospective observation component until closure, in case of multiple events, only one was considered to be the index PE, defined as the evaluable PE closest to the time of recruitment. The remaining PEs in the same patient were considered 'previous history' if they took place prior to the index PE, or 'recurrence', if subsequent to it. The registry was approved by the local Ethics Committees at each centre; informed consent was obtained from all living participants.

**Study design.** The main objective of this study was to develop a prognostic model, the EPHIPANY index, for cancer patients and both incidental, as well as symptomatic PE. Given that it was a non-intervention database, the data reflect genuine clinical profiles and the decisions physicians make in line with their clinical practice. The data were collected from medical records or directly from the patients, together with clinicians with experience in cancer support treatment and radiologists who are subspecialised in diseases of the chest. All the investigators were trained in the study protocol requirements and the data were monitored *in situ* or by phone. The data were gathered by means of an electronic capture system, designed to refine inconsistencies and resolve data errors in real time. Data acquisition was not blinded. The minimum observation period was 3 months from the time PE was diagnosed, although longer follow-up was required whenever possible. The variables were collected during routine or unscheduled medical appointments.

**Variables.** The main outcome measure in this study was the occurrence of a serious medical condition between PE diagnosis on imaging and 15 days later. Serious complications are events that lead to serious clinical deterioration or death; for example, systolic blood pressure <90 mm Hg, acute respiratory failure, right-side heart failure, acute kidney failure, major bleeding, or any other event the investigator deems serious (Supplementary Table 1).

Other secondary end points were all-cause 30-day mortality, the cause of 30-day mortality, and 30-day venous or arterial rethrombosis. 'Rethrombosis' was defined as a second thrombotic event after appropriate PE treatment or progression of a previous venous thromboembolism (VTE) despite proper anticoagulant therapy. Rethrombosis was not considered to be a serious complication in the absence of the afore-named criteria. An autopsic diagnosis notwithstanding, researchers attributed the cause of death on the basis of a clinical history review and findings from complementary testing. Demise was deemed to be due exclusively to PE when there was a direct causal relationship through a concatenation of events associated to the thrombosis pathophysiology. 'Mixed' deaths were defined by the presence of a temporal relationship between patient demise and PE, although other intercurrent complications (e.g., infection or tumour progression) might plausibly play a relevant role. Death was considered unrelated to PE if there was no temporal relationship or concatenation of clear events. 'Multiple' was accepted as a response when there was a resumption of overlapping causes.

The potential explanatory covariates were selected after a bibliographic review and consultation with experts, taking into account their availability at patients' bedside. Data recording did not allow for lost data for outcomes, survival times, and basic demographic and clinical characteristics (vital signs, tumour status, performance status, etc.).

The 'CDR variable' was defined as adaptation of Hestia's study eligibility criteria used in previous studies (Zondag *et al*, 2011; Weeda *et al*, 2016). These criteria are typified by the presence of at

least one of the following: systolic blood pressure <100 mmHg, arterial oxygen saturation <90%, respiratory rate  $\geq 30$  breaths per minute, pulse  $\geq 110$  beats per minute, sudden or progressive dyspnoea, other serious complications, constituting admission criteria in and of themselves, and clinically relevant bleeding, high risk of bleeding, or platelets  $< 50\,000\text{ mm}^{-3}$ . The CDR was assessed immediately prior to the time of radiological diagnosis of PE.

Other explanatory covariates included: age, gender, tumour stage, type of cancer, use of targeted cancer therapies, tumour response at the time of PE based on radiological criteria, Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 (Eisenhauer *et al*, 2009), Eastern Cooperative Group Performance Status scale (ECOG-PS; Oken *et al*, 1982), chronic obstructive pulmonary disease, prior cardiovascular disease, chronic kidney failure, concurrent deep vein thrombosis or a history of VTE, development of PE during treatment for a previous VTE, troponin levels (normal or high), creatinine clearance (normal, low), incidental or symptomatic diagnosis of the PE, presence of PE-specific symptoms, right ventricular diameter, additional radiological findings, Qanadli index (Qanadli *et al*, 2001), interventricular septal anomalies, presence of a single or multiple PE, oxygen saturation, blood pressure, heart and respiratory rates, previous tumour bleeding, prior use of antiaggregants, and major surgery in the previous month. Standardised definitions were used for each variable (Supplementary Table 1).

**Development of a decision tree model.** Classifications based on decision tree modelling seek to discover how the outcome variable is linked to the potential explanatory factors and, specifically, the configuration of these factors. This method is considered appropriate as the contribution of the explanatory covariates cannot be assumed to be necessarily additive or linear (Yohannes and Hodinott, 1999; Lewis, 2000). The Exhaustive CHAID algorithm builds a decision tree by means of repeated partitions of each subset into two or more child nodes, beginning with the full database (Biggs *et al*, 1991). This methodology was used to determine the strength of association between the presence of serious complications within 15 days and the previously mentioned potential predictors. To determine the best split in each node, the categories of each predictor were merged into pairs until statistically significant differences were no longer observed within each component of the pair in comparison to the target variable. The predictors that produced the most significant partitions were then recursively chosen. Thus, the algorithm identified the main interactions and built subgroups defined by the different sets of independent variables. The level of significance for splitting nodes ( $\alpha_{\text{split}}$ ) was 0.05. The Bonferroni method was used to adjust the value of significance.

To cope with the overfitting and instability inherent to the decision tree, a 10-fold cross-validation procedure was applied. Thus, the data were randomly divided into 10 equal subsets. Trees were systematically built in 9 of those partitions (training subsets) and then tested in the remaining group (testing subset). Cross-validation produces a single, final model. 'Risk' is defined as the proportion of cases incorrectly classified by each of the individual trees; whereas the cross-validated risk estimate is the average of the risks of all the trees. This analysis was performed with the SPSS 23.0 software (SPSS Inc., Chicago, IL, USA). The bootstrap (1000 replications) optimism-corrected area under the receiver operating characteristic curve (ROC) was estimated using R software with the rms package (Harrell *et al*, 2015).

Finally, the percentage of serious complications was calculated in each of the terminal nodes of the tree and this was used to create the EPHIPANY rule, a simplified classification based on three categories: low, intermediate, and high risk. The predictive value of this classification was estimated by means of the odds ratio (OR) and its 95% confidence interval (95% CI) between the groups

thereby generated and the appearance of the study end point. Cumulative hazards curves were calculated to establish the changes in hazard over time for each prognostic category.

## RESULTS

**Patient characteristics.** Patient characteristics are summarised in Table 1. Pulmonary embolism was incidentally diagnosed in 53% of the cases. Twenty-eight percent ( $n = 302$ ) of the episodes were

**Table 1. Baseline demographic and clinical characteristics**

Category	All patients, n = 1075
Age in years Mean (s.d.)	64 (12)
Males	492 (45.8%)
Active smokers	147 (13.7%)
COPD	128 (11.9%)
Chronic cardiovascular disease	56 (5.2%)
ECOG-PS $\geq 2$	506 (47.1%)
Malignant disease	
Breast	126 (11.7%)
Lung	272 (25.3%)
Colorectal	190 (17.7%)
Prostate	39 (3.6%)
Ovarian	41 (3.8%)
Esophagus-gastric	77 (7.2%)
Other	330 (30.7%)
TNM tumour stage IV at the time of PE	791 (73.6%)
Tumour response assessment at the time of PE	
Complete-partial response/stable disease/no evidence of disease	428 (39.8%)
Progressive disease/unknown/not evaluated	647 (60.2%)
Previous VTE	126 (11.7%)
Therapy in the previous month	
Major surgery	79 (7.3%)
Chemotherapy	576 (53.6%)
Targeted therapy	142 (13.2%)
Hormone therapy	90 (8.4%)
ESA	52 (4.8%)
Antiplatelet therapy	100 (9.3%)
Location of treatment	
Hospital admission	692 (64.4%)
Intensive care unit	33 (3.1%)
Home therapy	294 (27.4%)
Early discharge (<72 h)	47 (4.4%)
Home hospitalisation	8 (0.7%)
Initial therapy with LMWH	991 (92.2%)
Clinical findings	
Heart rate > 110 beats per minute	231 (21.5%)
Systolic pressure < 100 mm Hg	109 (10.1%)
Respiratory rate $\geq 30$ times per minute	80 (7.4%)
Body temperature < 36 °C	24 (2.2%)
Altered mental status	29 (2.7%)
Arterial oxygen saturation value < 90%	166 (15.4%)
Incidental PE diagnosis	564 (53.5%)
PE-specific symptoms	517 (48.1%)
Surgery of primary tumour	473 (44.0%)

Abbreviations: COPD = chronic obstructive pulmonary disease; ECOG-PS = Eastern Cooperative Oncology Group Performance status scale; ESA = erythropoiesis-stimulating agent; LMWH = low-molecular weight heparin; PE = pulmonary embolism; TNM = tumour node metastasis; VTE = venous thromboembolism. The data entries represent percentages calculated with respect to columns, with the exception of age. *P*-values were calculated with the Linear-by-Linear association test, except for the continuous variable 'age' for which the Kruskal-Wallis test was used. Risk categories are defined according to the end nodes of the decision tree in Figure 2.

treated at home. All patients received anticoagulation (initial therapy with low-molecular weight heparin in 92%). At the time of PE diagnosis, 73.6% of the patients had a metastatic tumour and 53.6% were receiving chemotherapy. The most common tumours were breast, lung, and colon cancer, accounting for 54.7% of the series. The recruitment process is illustrated in Figure 1.

**Outcomes.** The main end point of this study, serious complications within 15 days, occurred in 208 patients (19.3%; 95% CI, 17.1–21.8%). The 15-day mortality rate was 10.1% (95% CI, 8.4–12.1) and of the 109 patients who died within that period, 45 (41%) did so as a result of tumour progression and not PE (Table 2). The rates of embolic recurrence and major bleeding were 4% and 2%, respectively.

**Decision tree.** Figure 2 shows the decision tree model with the 15-day, serious complications data for each end node. The Exhaustive CHAID method selected six explanatory covariates from the initial 39: the Hestia-like CDR variable (any risk factor present *vs* none), ECOG-PS (<2 *vs* ≥2), oxygen saturation (<90 *vs* ≥90%), presence of PE-specific symptoms, previous tumour response evaluation (tumour progression, unknown, or not evaluated *vs* others), and prior surgical resection of primary tumour. While other additional nodes involving more variables could be generated, they did not provide any incremental risk discrimination.

The best predictor in the root node was the Hestia-like CDR variable; the episodes that did not meet any of these criteria had a lower risk of serious complications within 15 days, in comparison with episodes that satisfied at least one of them (4.7 *vs* 29.7%; OR 0.11, 95% CI, 0.07–0.18; *P*<0.0001) and 15-day mortality (2.5 *vs* 15.6%; OR 0.13; 95% CI, 0.07–0.25; *P*<0.0001), respectively. The decision tree makes it possible to elaborate on the prognostic stratification in seven terminal nodes. For purposes of practicality, they are summarised into three risk categories: high, intermediate, and low. Supplementary Table 2 outlines the demographic and clinical characteristics of each subgroup.

Low risk encompasses patients without any Hestia-like CDR criteria, and with controlled tumours or resected primary tumours, with a risk of serious complications of between 1.4–3.4% and 0.3% 15-day mortality. Tumours with any of the CDR risk factors were

at high risk (complications rate, 20–55%), with the exception of the group consisting of patients with a good performance status, and no PE-specific symptoms, who had an intermediate level of risk. This risk group would also include all stable patients having uncontrolled or unevaluated tumours, and without surgery for the primary tumour, with a 10.6% risk of complications. The cross-validated risk estimate is 0.191 (s.e.=0.012); the optimism-

**Table 2. Main outcomes**

Outcome	Overall data set n = 1075	%, (95% CI)
15-day serious complications <sup>a</sup>		
Overall	208	19.3%, (17.1–21.8)
Systolic BP <90 mm Hg	50	4.7%, (3.5–6.1)
Acute respiratory failure	86	8.0%, (6.4–9.8)
Fibrinolysis	7	0.7%, (0.2–1.4)
Major bleeding	54	5.0%, (3.8–6.5)
Acute right ventricular failure	22	2.0%, (1.3–3.1)
Acute renal failure	17	1.6%, (0.9–2.5)
Admission to ICU	21	2.0%, (1.2–3.0)
Cardiopulmonary resuscitation	4	0.4%, (0.1–0.9)
Non-invasive ventilation	9	0.8%, (0.4–1.6)
Oro-tracheal intubation	4	0.4%, (0.1–1.0)
Death	109	10.1%, (8.4–12.1)
Other	42	3.9%, (2.8–5.3)
Causes of 15-day mortality		
Death caused exclusively by PE	19	1.8%, (1.1–2.8)
PE-related death, mixed	17	1.6%, (0.9–2.5)
Fatal bleeding	6	0.6%, (0.2–1.2)
Cancer	45	4.2%, (3.1–5.6)
Infection	8	0.7%, (0.3–1.5)
Arterial thrombosis	0	0
Other/unknown	14	1.3%, (0.7–2.2)
15-day venous rethrombosis	8	0.7%, (0.3–1.5)
All-cause 30-day mortality	153	14.2%, (12.2–16.5)
30-day serious bleeding	59	5.5%, (4.2–7.0)
30-day venous rethrombosis	12	1.1%, (0.6–2.0)

Abbreviations: BP = blood pressure; CI = confidence interval; ICU = intensive care unit; PE = pulmonary embolism.  
<sup>a</sup>These events are not mutually exclusive.

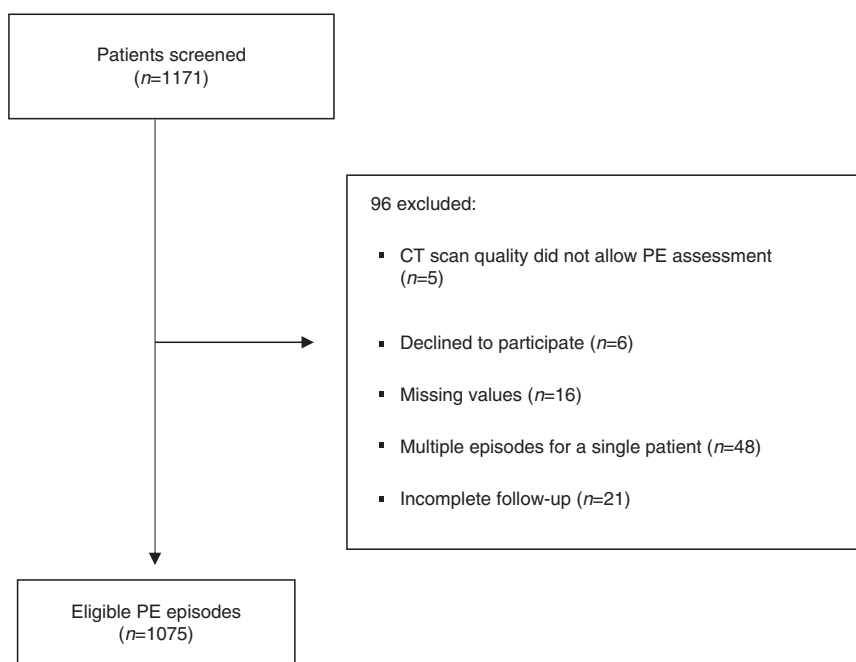


Figure 1. Study flow diagram. CT, computed tomography; PE, pulmonary embolism.



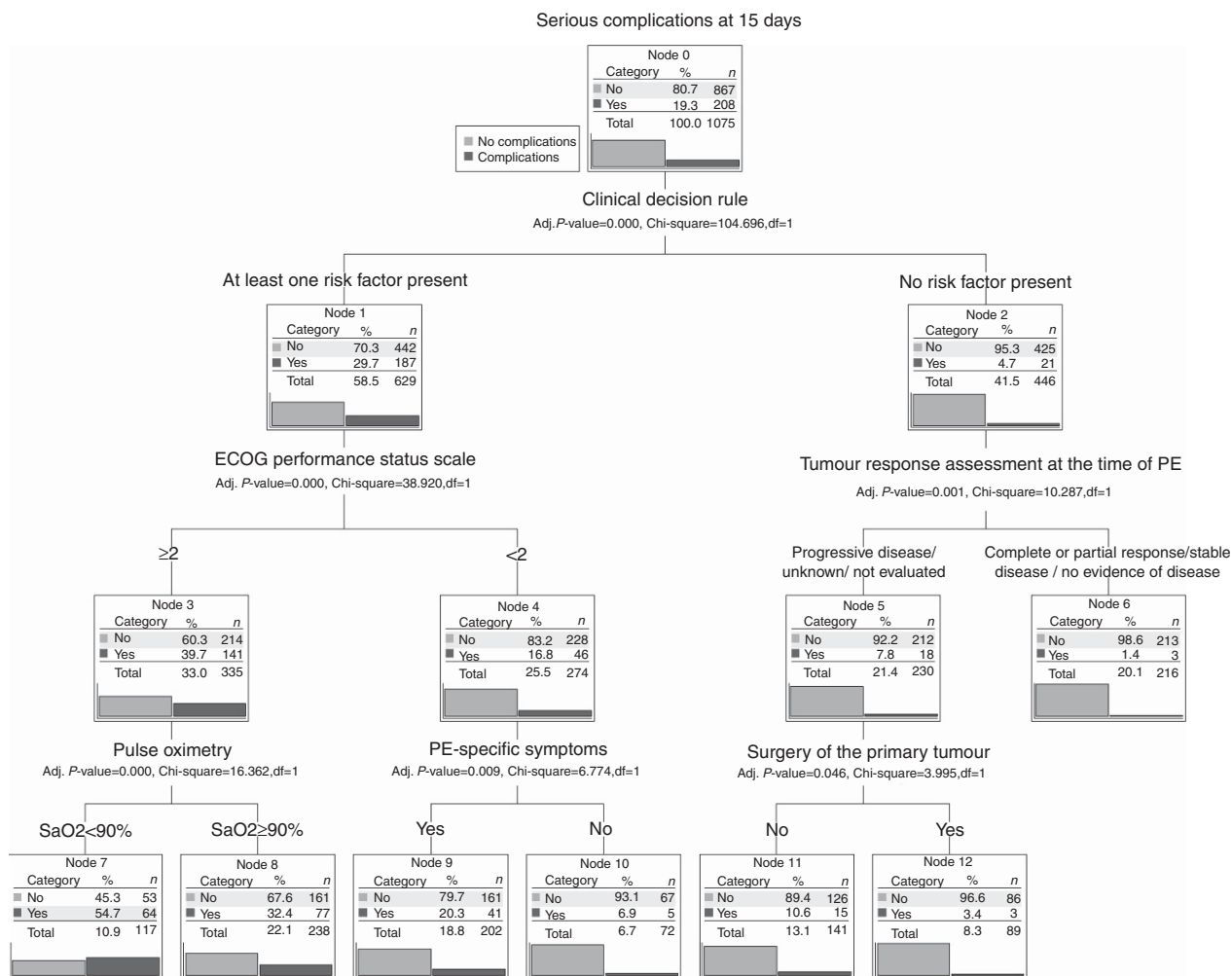


Figure 2. EIPHANY Index for the prediction of serious complications. The bars show the percentage of patients with no complications (light gray) or with complications (dark gray) within each node. The 'Clinical Decision Rule' variable encompasses the following characteristics: (1) systolic blood pressure < 100 mm Hg, (2) arterial oxygen saturation < 90%, (3) respiratory rate ≥ 30 breaths per minute, (4) pulse ≥ 110 beats per minute, (5) sudden or progressive dyspnoea, and (6) clinically relevant haemorrhage, high risk of bleeding, or platelets < 50 000 mm<sup>-3</sup>. The patient is classified as low or high risk according to whether they exhibit none of these characteristics or at least one of them. CDR, clinical decision rule; ECOG-PS, Eastern Cooperative Group Performance Status scale; PE, pulmonary embolism; SaO<sub>2</sub>, arterial oxygen saturation.

corrected value of the area under the ROC curve is 0.779 (95% CI, 0.717–0.840).

Outcomes according to risk groups are reported in Table 3. The risk of serious complications within 15 days increases with the group: 1.6, 9.4, 30.6%; *P*<0.0001. The risk of 15-day mortality also raises progressively, in patients of low, intermediate, and high risk: 0.3, 6.1, and 17.1%; *P*<0.0001. It is worth noting that high-intermediate-risk patients had increased risk of complications, with OR of 17.2 (95% CI, 7.7–40.3), *P*<0.0001, and death OR of 49.5 (95% CI, 6.8–356.9), *P*<0.0001.

Figure 3 illustrates the cumulative hazard function. Events are seen to be evenly distributed throughout the 15 days and do not cluster in the first hours following diagnosis of PE. The log-rank test reveals that the survival functions factored by prognostic categories are significantly different (*P*<0.0001).

**DISCUSSION**

This study reports on the development of a decision tree model to stratify any cancer patient with PE according to the risk of serious

complications within 15 days. Unlike other prognostic tools, the EIPHANY index is applicable across the entire spectrum of PE severity, including both incidental and symptomatic events (Wicki *et al*, 2000; Aujesky *et al*, 2006; Uresandi *et al*, 2007; Jiménez *et al*, 2010; Kline *et al*, 2012; Den Exter *et al*, 2013). The model is a validation and extension of the CDR proposed in several clinical trials with the aim of pragmatically selecting low-risk patients eligible for outpatient care (Siragusa *et al*, 2005; Zondag *et al*, 2011; Font *et al*, 2014; Weeda *et al*, 2016). These decision-making rules are based on the combination of altered vital signs (e.g., hypotension, hypoxaemia, tachycardia, etc.) and factors that point toward a high risk of bleeding or other contraindications to receiving treatment in the home. Moreover, the EIPHANY rule incorporates another five covariates that include discriminatory characteristics typical in cancer patients that are easily accessible at patients' bedside, such as ECOG-PS, evaluation of tumour response prior to PE using RECIST 1.1 criteria, previous primary tumour resection, oxygen saturation, and the presence or absence of PE-specific symptoms. All these variables have been widely used in various contexts to predict clinical outcome and there is good reason to think that they are also important in PE (Wicki *et al*, 2000; Aujesky *et al*, 2006;

**Table 3. Outcomes in each risk group (n = 1075)**

Category	All patients, n = 1075	Low risk: nodes 6, 12 n = 305	Intermediate risk: nodes 10, 11 n = 213	High risk: nodes 7, 8, 9 n = 557	P-value
15-day complications					
Overall	208 (19.3%)	6 (2.0%)	20 (9.4%)	182 (32.7%)	<0.0001
Hypotension	50 (4.7%)	0	0	50 (9.0%)	<0.0001
Acute respiratory failure	86 (8.0%)	1 (0.3%)	1 (0.5%)	84 (15.1%)	<0.0001
Fibrinolysis	7 (0.7%)	0	0	7 (1.3%)	0.019
Major bleeding	54 (5.0%)	3 (1.0%)	5 (2.3%)	46 (8.3%)	<0.0001
Right-side heart failure	22 (2.0%)	0	2 (0.9%)	20 (3.6%)	<0.0001
Acute kidney failure	17 (1.6%)	0	1 (0.5%)	16 (2.9%)	0.001
ICU admission	21 (2.0%)	0	0	21 (3.8%)	<0.0001
Cardiopulmonary resuscitation	4 (0.4%)	0	0	4 (0.7%)	0.075
Non-invasive mechanical ventilation	9 (0.8%)	0	0	9 (1.6%)	0.008
Oro-tracheal intubation	4 (0.4%)	0	0	4 (0.7%)	0.076
Demise	109 (10.1%)	1 (0.3%)	13 (6.1%)	95 (17.1%)	<0.0001
Other serious complication	42 (3.9%)	1 (0.3%)	2 (0.9%)	39 (7.0%)	<0.0001
Day first complication appeared, median (range)	3 (0–15)	13 (5–15)	7 (0–15)	3 (0–15)	0.020
Cause of death at 15 days					
PE	19 (1.8%)	0	2 (0.9%)	17 (3.1%)	0.001
Mixed cause	17 (1.6%)	0	0	17 (3.1%)	<0.0001
Bleeding	6 (0.6%)	0	1 (0.5%)	5 (0.9%)	0.088
Disease progression	45 (4.2%)	1 (0.3%)	10 (4.7%)	34 (6.1%)	<0.0001
Sepsis	8 (0.7%)	0	0	8 (1.4%)	0.012
Arterial thrombosis	0	0	0	0	—
Other reason	14 (1.3%)	0	0	14 (2.5%)	0.001
Venous rethrombosis within 15 days	8 (0.7%)	3 (1.0%)	0	5 (0.9%)	0.959
Recurrence of PE at 15 days	43 (4.0%)	16 (5.2%)	13 (6.1%)	14 (2.5%)	0.030
Arterial rethrombosis within 15 days	25 (2.3%)	2 (0.7%)	8 (3.8%)	15 (2.7%)	0.095
30-day mortality	153 (14.2%)	4 (1.3%)	19 (8.9%)	130 (23.3%)	<0.0001

Abbreviations: ICU = intensive care unit; PE = pulmonary embolism. Percentages refer to columns. It was possible to record more than one serious complication for the same patient; causes of death could overlap. The P-value was calculated using the Linear-by-Linear association test, except for the day the complication appeared (independent samples median test).

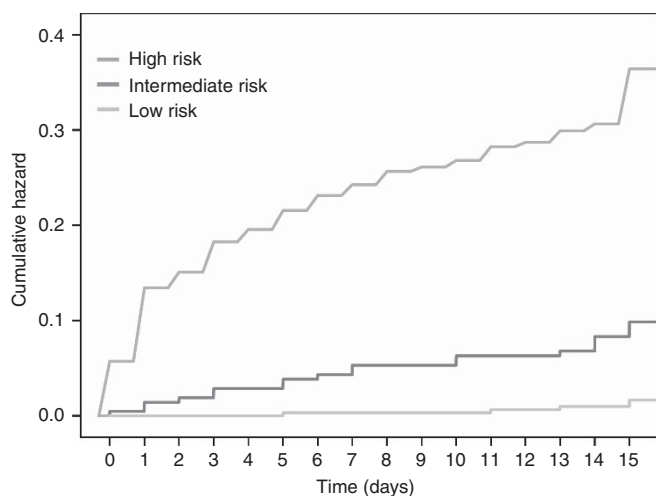


Figure 3. Cumulative hazard functions for serious complications. In this figure, cumulative hazard curves were plotted to show the change in hazards over time (days), for each prognostic category.

Uresandi *et al*, 2007; Jiménez *et al*, 2010; Kline *et al*, 2012; Den Exter *et al*, 2013).

The ECOG-PS has been largely acknowledged in oncology to predict toxicity and adverse clinical outcomes in various contexts (Oken *et al*, 1982). In general, functional worsening points to severe underlying pathology, poorer physiological reserve, and decreased mobility, thereby making patients more prone to thrombotic risk (Jiménez *et al*, 2010; Den Exter *et al*, 2013). Patients having any risk factor and poor functional status have a

worse prognosis than those with a good functional status, particularly when PE diagnosis is incidental. On the other hand, we have detected a new variable that should be incorporated into the CDR: evaluation of tumour response prior to PE based on RECIST radiological criteria, which determines short-term prognosis following PE. Tumours in progression or those at risk for progression because response could not be assessed are at higher risk for complications than those with controlled disease or with no evidence of disease, even in the absence of other prognostic factors. In fact, resection of the primary tumour appears to be the only thing that protects individuals with tumours in progression and no other risk factor facing a complicated clinical course. This variable likely improves prognosis as a consequence of decreasing local complications, such as serious bleeding (Lee *et al*, 2009). In fact, in this series, bleeding was located in the primary tumour in 43% of the cases, which rose to 50% in subjects who died due to haemorrhage.

Remarkably, some patients diagnosed incidentally displayed PE-specific symptoms upon meticulous anamnesis, as reported by other authors (O’Connell *et al*, 2006). Therefore, the absence of PE-specific symptoms does not correspond exactly with incidental PE.

The decision tree model classification method used after the Exhaustive CHAID procedure is one of the differences that distinguish the EPIPHANY index from other models (Wicki *et al*, 2000; Aujesky *et al*, 2006; Uresandi *et al*, 2007; Jiménez *et al*, 2010; Kline *et al*, 2012; Den Exter *et al*, 2013). This design was chosen given the interest in generating a classification that would reasonably imitate authentic decision-making. This means that, unlike a binary logistic regression, which postulates the existence of additive effects that contribute to explaining outcome, decision trees factor in the existence of strong interactions between variables

and are better suited to elaborating decision-making algorithms that follow the same structure (Yohannes and Hoddinott, 1999; Lewis, 2000). Thus, in the real world, decisions in subjects with PE are not generally made on the basis of the small additive contributions of several variables, but on the presence or absence of strong dichotomous predictors such as cardiogenic shock, acute respiratory failure, hypotension, etc. (Wicki *et al*, 2000; Aujesky *et al*, 2006; Uresandi *et al*, 2007; Jiménez *et al*, 2010; Kline *et al*, 2012; Den Exter *et al*, 2013). The presence of a single one of these variables indicates high risk and is fundamental in the clinical decision to intensify therapy, regardless of the contribution of the remaining covariates of a logistic regression model. Decision trees are also useful in situations having non-linear probable effects for some variables, as is assumed in a sample of patients diagnosed using different methods (CT-angiography scans vs conventional CT) and having dissimilar clinical characteristics, depending on if they are incidental or symptomatic events. Insofar as the previously developed scales are concerned (RIETE, POMPE-C, PESI, etc.) (Aujesky *et al*, 2006; Jiménez *et al*, 2010; Kline *et al*, 2012; Den Exter *et al*, 2013), we do not know for sure if they can complement these criteria, although a preliminary analysis performed by our group suggests that their use is not likely to be necessary after applying a clinical classification rule (Carmona-Bayonas *et al*, 2016).

Another striking difference between the EPIPHANY index and the afore-mentioned methods is that we propose beginning to use the probability of serious complications within 15 days as the primary end point and not all-cause 30-day mortality, which had been typically used in other studies (Aujesky *et al*, 2006; Jiménez *et al*, 2010; Kline *et al*, 2012; Den Exter *et al*, 2013). Of course we agree that mortality is a far more solid outcome; nonetheless, we believe that considering other end points in different clinical situations, as our group recently proposed (Carmona-Bayonas *et al*, 2016), is justified. One of the arguments is that the appearance of serious complications in individuals with PE treated as outpatients, far from medical supervision, can paradoxically turn low-risk patients into the most vulnerable, because of misclassification. In contrast, the probability of all-cause 30-day mortality will not necessarily affect decision-making regarding ambulatory treatment in some subgroups, as the cause of death is rarely the PE itself, and is often due to cancer progression (Den Exter *et al*, 2013; Carmona-Bayonas *et al*, 2016). In fact, patients on palliative care for advanced disease are those in whom it is even more important to prevent unnecessary hospitalisation at the end of their lives. The use of the all-cause 30-day mortality end point also entails the issue of proposing intensification of PE management (e.g., with fibrinolysis) in subjects at greater risk of early mortality due to cancer, who are precisely the ones who are less likely to benefit. For instance, when we examine the causes of death 15, 30, and 90 days after PE, cancer is the cause of death in 35%, 54%, and 65%, respectively. Although determining the cause of death in absence of autopsies has clear limitations, the same data appear in the RIETE registry, in which ~50% of the deaths resulted from the cancer itself and not PE (Den Exter *et al*, 2013).

This study has certain limitations that must be taken into account. First of all, it is a fundamentally retrospective registry of medical history data, with the intrinsic limitations in precision this entails. Nevertheless, most of the events contemplated are solid and are faithfully recorded in the histories (blood pressure, oxygen saturation, documentation of tumour response, ECOG-PS, exitus, etc.). Second, PE is a highly polymorphic pathology and more external validations are needed by other groups, being cognizant that these models offer a general overview of the main risk factors. However, some subjects have other particular factors with a definitive impact on prognosis. Third, decision tree modelling can be weak, unstable predictors in certain contexts. Thus, random

forest models that incorporate the prediction of multiple, individual decision trees may perform better, albeit they are also more complicated to interpret and use in daily practice (Breiman, 2001). It is also doubtful that they can achieve a better definition of 'low risk'. Finally, the assimilation and integration of radiological variables and/or biomarkers (e.g., troponin, pulmonary artery obstruction indices, right ventricular dilatation, etc.) would call for more in-depth studies.

In short, we have elaborated a decision tree to predict serious complications in cancer patients with PE that enables patients to be classified into groups of high, intermediate, and low risk for complications. This model validates and refines the classification rules previously used by other authors; it is based on variables that are easy to obtain; it's easy to use, and can have potential implications for clinical management.

## ACKNOWLEDGEMENTS

We thank Priscilla Chase Duran for editing the text. This project was funded in part by a restricted educational grant from Leo Pharma Spain and by support from the Asociación de Investigación de la Enfermedad Tromboembólica de la Región de Murcia (the Region of Murcia's Association of Thromboembolic Disease Research).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- Aujesky D, Obrosky DS, Stone RA, Auble TE, Perrier A, Cornuz J, Roy P-M, Fine MJ (2006) A prediction rule to identify low-risk patients with pulmonary embolism. *Arch Intern Med* **166**: 169–175.
- Biggs D, De Ville B, Suen E (1991) A method of choosing multiway partitions for classification and decision trees. *J Appl Stat* **18**: 49–62.
- Breiman L (2001) Random forests. *Mach Learn* **45**: 5–32.
- Carmona-Bayonas A, Font C, Fonseca PJ, Fenoy F, Otero R, Beato C, Plasencia J, Biosca M, Sánchez M, Benegas M, Calvo-Temprano D, Varona D, Faez L, Vicente MA, de la Haba I, Antonio M, Madridano O, Ramchandani A, Castañón E, Marchena PJ, Martínez MJ, Martín M, Marín G, de la Peña FA, Vicente V (2016) On the necessity of new decision-making methods for cancer-associated, symptomatic, pulmonary embolism. *Thromb Res* **143**: 76–85.
- Den Exter PLPL, Gómez V, Jiménez D, Trujillo-Santos J, Muriel A, Huisman MV, Monreal M, Registro Informatizado de la Enfermedad Tromboembólica (RIETE) Investigators Gómez V, Jiménez D (2013) A clinical prognostic model for the identification of low-risk patients with acute symptomatic pulmonary embolism and active cancer. *Chest* **143**: 138–145.
- Dentali F, Ageno W, Becattini C, Galli L, Gianni M, Riva N, Imberti D, Squizzato A, Venco A, Agnelli G (2010) Prevalence and clinical history of incidental, asymptomatic pulmonary embolism: a meta-analysis. *Thromb Res* **125**: 518–522.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancy J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**: 228–247.
- Font C, Carmona-Bayonas A, Beato C, Reig Ó, Sáez A, Jiménez-Fonseca P, Plasencia JM, Calvo-Temprano D, Sanchez M, Benegas M (2016) Clinical features and short-term outcomes of cancer patients with suspected and unsuspected pulmonary embolism: the EPIPHANY study. *Eur Respir J* **49**: 1600282.
- Font C, Carmona-Bayonas A, Fernández-Martínez A, Beato C, Vargas A, Gascon P, Otero R (2014) Outpatient management of pulmonary

- embolism in cancer: data on a prospective cohort of 138 consecutive patients. *J Natl Compr Canc Netw* **12**: 365–373.
- Harrell Jr F, Frank E, Maintaner Frank E (2017) 'Package 'rms.' Vanderbilt University, pp 229. Available at: <https://cran.r-project.org/web/packages/rms/rms.pdf>.
- Jiménez D, Aujesky D, Moores L, Gómez V, Lobo JL, Uresandi F, Otero R, Monreal M, Muriel A, Yusen RD (2010) Simplification of the pulmonary embolism severity index for prognostication in patients with acute symptomatic pulmonary embolism. *Arch Intern Med* **170**: 1383–1389.
- Kline JA, Roy PM, Than MP (2012) Derivation and validation of a multivariate model to predict mortality from pulmonary embolism with cancer: the POMPE-C tool. *Thromb Res* **129**: e194–e199.
- Lee JA, Lim DH, Park W, Ahn YC, Huh SJ (2009) Radiation therapy for gastric cancer bleeding. *Tumori* **95**: 726–730.
- Lewis RJ (2000) An introduction to classification and regression tree (CART) analysis. In Annual Meeting of the Society for Academic Emergency Medicine; San Francisco, CA, USA, pp 1–14.
- Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, Carbone PP (1982) Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* **5**: 649–655.
- O'Connell CL, Boswell WD, Duddalwar V, Caton A, Mark LS, Vigen C, Liebman HA (2006) Unsuspected pulmonary emboli in cancer patients: clinical correlates and relevance. *J Clin Oncol* **24**: 4928–4932.
- Plasencia-Martínez JM, Carmona-Bayonas A, Calvo-Temprano D, Jiménez-Fonseca P, Fenoy F, Benegas M, Sánchez M, Font C, Varona D, de la Haza DM (2016) Prognostic value of computed tomography pulmonary angiography indices in patients with cancer-related pulmonary embolism: data from a multicenter cohort study. *Eur J Radiol* **87**: 66–75.
- Qanadli SD, El HM, Vieillard-Baron A (2001) New CT index to quantify arterial obstruction in pulmonary embolism: comparison with angiographic index and echocardiography. *AJR Am J Roentgenol* **176**: 1415–1420.
- Ranpura V, Hapani S, Wu S (2011) Treatment-related mortality with bevacizumab in cancer patients: a meta-analysis. *JAMA* **305**: 487–494.
- Siragusa S, Arcara C, Malato A, Anastasio R, Valerio MR, Fulfaro F, Lo Coco L, Grimaudo S, Bajardi G, Abbadessa V, Gebbia N (2005) Home therapy for deep vein thrombosis and pulmonary embolism in cancer patients. *Ann Oncol* **16**(Suppl 4): iv136–iv139.
- Streiff MB, Bockenstedt PL, Cataland SR, Chesney C, Eby C, Fanikos J, Fogarty PF, Gao S, Garcia-Aguilar J, Goldhaber SZ, Hassoun H, Hendrie P, Holmstrom B, Jones KA, Kuderer N, Lee JT, Millenson MM, Neff AT, Ortel TL, Smith JL, Yee GC, Zakarija A, Fogerty AE, Gao S, Goldhaber SZ, Hassoun H, Hendrie P, Holmstrom B, Kuderer N, Lee JT, Millenson MM, Neff AT, Ortel TL, Siddiqi T, Smith JL, Yee GC, Zakarija A, McMillian N, Naganuma M (2013) Venous thromboembolic disease. *J Natl Compr Canc Netw* **11**: 1402–1429.
- Svendsen E, Karwinski B (1989) Prevalence of pulmonary embolism at necropsy in patients with cancer. *J Clin Pathol* **42**: 805–809.
- Sorensen HT, Mellekjær L, Olsen JH, Baron JA (2000) Prognosis of cancers associated with venous thromboembolism. *N Engl J Med* **343**: 1846–1850.
- Uresandi F, Otero R, Cayuela A, Cabezedo MA, Jimenez D, Laserna E, Conget F, Oribe M, Nauffal D (2007) A clinical prediction rule for identifying short-term risk of adverse events in patients with pulmonary thromboembolism. *Arch Bronconeumol* **43**: 617–622.
- Weeda ER, Kohn CG, Peacock WF, Fermann GJ, Crivera C, Schein JR, Coleman CI (2016) External validation of the hestia criteria for identifying acute pulmonary embolism patients at low risk of early mortality. *Clin Appl Thromb Hemost* doi:10.1177/1076029616651147.
- Wicki J, Perrier A, Perneger TVV, Bounameaux H, Junod AF (2000) Predicting adverse outcome in patients with acute pulmonary embolism: a risk score. *Thromb Haemost* **84**(4): 548–552.
- Yohannes Y, Hoddinott J (1999) Classification and regression trees: an introduction. *International Food Policy Research Institute*. Washington, DC, USA. Available at: [http://pdf.usaid.gov/pdf\\_docs/PNACH725.pdf](http://pdf.usaid.gov/pdf_docs/PNACH725.pdf).
- Zondag W, Mos ICM, Creemers-Schild D, Hoogerbrugge ADM, Dekkers OM, Dolsma J, Eijsvogel M, Faber LM, Hofstee HMA, Hovens MMC, Jonkers GJPM, van Kralingen KW, Kruip MJHA, Vlasveld T, de Vreede MJM, Huisman MV (2011) Outpatient treatment in patients with acute pulmonary embolism: the Hestia Study. *J Thromb Haemost* **9**: 1500–1507.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)