Cancer Epidemiology 53 (2018) 111–118

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Development and validation of a prognostic index for survival in non-small cell lung cancer: Results from a Tunisian cohort study

Ghassen Soussi[a,*], Nissaf Ben Alaya[b], Nawel Chaouch[a], Hajer Racil[a]

[a] Department of Pulmonary Medicine II, Abderrahmen Mami Hospital, 2080 Ariana, Tunisia
[b] National Observatory of New and Emerging Diseases, 5-7, Khartoum Street, Diplomat Complex, Block 4, 13th Floor, Belvedere, 1002 Tunis, Tunisia

ARTICLE INFO

Keywords:
Non-small cell lung cancer
Prognosis
Survival
Cox proportional hazards
Modelling

ABSTRACT

Introduction: Despite the continuous efforts made with the TNM system, the issue of heterogeneity of prognosis within the stages of non-small cell lung cancer (NSCLC) could not be resolved. Our aim was to identify prognostic factors and develop an index to predict NSCLC survival with greater accuracy.

Methods: We conducted a survival study over 5 years on patients with NSCLC. Kaplan–Meier analysis followed by Cox regression modelling were used. Prognostic indices were derived, using either an additive or a multiplicative pattern, and were compared by their receiver operating characteristics (ROC) curves. We then proceeded to a risk stratification and validation of the index on the derivation cohort.

Results: Two hundred and sixty-two NSCLC patients were included. Two models were constructed, using the following nine variables as prognostic factors: age, performance status, haemoglobin level, leucocyte count, calcium, lactate dehydrogenase, alkaline phosphatase levels, histological type and TNM stage. Four prognostic indices were derived, and the best one was picked and validated on a population of five risk groups. The higher the risk group, the shorter the survival.

Conclusions: This novel and simple prognostic tool could predict survival more accurately in patients with NSCLC.

1. Introduction

Lung cancer remains the leading cause of cancer death in the world. In 2012, there were 1.8 million new cases and almost 1.6 million deaths due to lung cancer. Its prognosis is poor with a 5-year overall survival of 10–15%. Tobacco smoking is the predominant cause of lung cancer worldwide, bringing about 90% of lung cancer cases. Lung carcinomas are classified into two histological types, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC); the latter accounts for 85% of all lung cancer cases [1].

The management and therapeutic follow-up of NSCLC relies strongly on the International Association for the Study of Lung Cancer (IASLC) tumour–node–metastasis (TNM) status. Through many years, this staging system has made a giant leap towards addressing the issues of heterogeneity of stage and size categories. Nevertheless, an overlap in survival rates between stages remains because of substantial heterogeneity in prognosis within the TNM groupings. As a matter of fact, sketching outcomes for an individual patient with NSCLC is a hard task even though the general outlook for groups of patients can be anticipated. Shifting the paradigm to a specific risk stratification at the individual level seems more relevant than ever in an era in which personalised medicine in cancer patients is being promoted.

A plethora of factors affecting outcomes in NSCLC were brought to light by numerous clinical investigations in a matter of a few years. Since a completist approach going through large numbers of prognostic factors may yield results that are difficult to interpret, we believe that simple indices, based on the most relevant factors, should suffice to allow a firm grasp of the prognostication process and help as a first building block for a decision support system in NSCLC patients.

From this perspective, we aimed to identify the main clinical, biological and histological prognostic factors determining outcomes in NSCLC and to develop a simple prognostic index predicting survival in NSCLC patients.

2. Methods

2.1. Data collection

We conducted a longitudinal prognostic study of a monocentric cohort of consecutive patients with NSCLC.

* Corresponding author at: Department of Pulmonary Medicine II, Abderrahmen Mami Hospital, 2080 Ariana, Tunisia.
E-mail addresses: ghassen.soussi@gmail.com (G. Soussi), nissafba@yahoo.fr (N. Ben Alaya), besnawel@yahoo.fr (N. Chaouch), hajer.racil@gmail.com (H. Racil).

https://doi.org/10.1016/j.canep.2018.01.018
Received 8 October 2017; Received in revised form 9 January 2018; Accepted 27 January 2018

All case records of patients with lung cancer admitted to the Respiratory Medicine Department II, a ward for male patients, in Abderrahmen Mami Hospital, Ariana, Tunisia, from January 1st, 2008 to December 31st, 2012 were analysed. We included histologically confirmed NSCLC cases of all TNM stages. Exclusion criteria were: lack of histological confirmation, secondary NSCLC, and SCLC cases.

Data were collected using the patients' case records, including: age at diagnosis, history of tobacco smoking, body mass index (BMI), performance status (PS) according to the Eastern Cooperative Oncology Group (ECOG) criteria, as well as clinical symptoms at presentation (cough, chest pain, haemoptysis, dyspnoea, dysphonia, fever, lethargy, anorexia and weight loss). Biological data were provided by blood tests taken upon admission: haemoglobin level, white cell and platelet counts, and levels of C-reactive protein (CRP), calcium, lactate dehydrogenase (LDH), D-dimer, and alkaline phosphatase (ALP). Tumour-related factors were also collected: endoscopic site of lesion, histological type, TNM status (according to the AJCC 8th Edition Cancer Staging System for NSCLC), and type and number of metastatic sites.

Data needed for calculating survival were collected prospectively. Death was considered the outcome of interest. The starting point (time origin) was set to the date of presentation and the status of each patient (alive, deceased, or lost to follow-up) was determined on the study termination date (set to April 8th, 2014).

Data entry was performed using EpiData® version 3.1 (Odense, Denmark: The EpiData Association; 2003–2005).

2.2. Statistical analysis

Data were exported to IBM® SPSS® Statistics for Windows, version 21.0. (Armonk, NY: IBM Corp.; 2012) for the statistical analysis. Graphics were rendered using Adobe® Illustrator® CC software v19.1.1. (Adobe Systems Incorporated; 2015).

All continuous variables were dichotomised according to clinically relevant classification. The impact of each variable on survival was studied in univariable analysis using the Kaplan–Meier method. The log rank (Mantel Cox) test was used to compare survival distributions between groups. During this step, all hypothesis testing was two-tailed, and *P*-values < .05 were considered to indicate statistical significance.

Multivariable analysis was conducted using backward stepwise Cox regression, with the survival time as the dependent variable and the prognostic factors as the explanatory variables. Only variables with a significance threshold $P < .20$ in univariable analysis were included as candidate variables in the modelling procedure. Missing data were handled using median replacement as an imputation strategy.

Two models were anticipated. Probability for stepwise entry was set to $P < .05$ for both while probability for stepwise removal was set to $P < .05$ (model A) and $P < .10$ (model B).

We applied proportional weighting of the variable categories to the corresponding hazard ratio (HR) only if $P < .05$. The weighted points were determined by the relative magnitudes of the regression coefficients with the following weighting rule: number of allocated points ≈ 10*HR.

Prognostic indices (PIs) were derived from the models obtained, using the additive pattern PI = $\sum$ (HRi*10) and the multiplicative pattern PI = $\prod$ (HRi*10).

The discriminative ability of the PI was tested by comparing the area under the curve (AUC) of their receiver operating characteristics (ROC) curves pairwise, using Hanley and McNeil's test [2]. This test was performed using MedCalc® Statistical Software version 14.8.1. (Ostend, Belgium: MedCalc Software bvba; 2014). The Bonferroni correction was used to counteract the problem of multiple comparisons by controlling the familywise error rate. Literature findings as well as the practical aspect impacted the choice between the PIs whenever Hanley and McNeil's test could not discriminate between them.

The PI with the best predictive accuracy was picked. Patients were stratified into risk groups according to their PI. The cut-offs used for this categorisation were chosen in order to guarantee a good representativeness in sample size and number of events.

The predictive ability of the PI was eventually tested by the Kaplan–Meier method and Cox regression following the risk stratification procedure.

2.3. Ethical statement

The protocol of this study was approved by our institutional ethics committee (Comité d'éthique, Hôpital Abderrahmen Mami) which is listed in the UNESCO Global Ethics Observatory (GEObs2). Oral informed consent was obtained from all living patients before enrolment. Full confidentiality and anonymity of the research subjects during collection, storage and use of data were respected. Absence of harm to the research participants was guaranteed.

3. Results

3.1. Patient characteristics

We identified 333 patients with lung cancer. After applying exclusion criteria, 262 cases remained. All patients were male with a median age of 60.5 years at diagnosis. They were more likely to be smokers (95.0%) with a smoking mean of 52.4 pack-years. More than nine out of ten patients reported respiratory symptoms at presentation (mainly chest pain). Nearly three quarters of the cases complained about general symptoms (weight loss in 57.6%, lethargy in 49.6%, anorexia in 48.1%, fever in 16.4%). Mean body mass index (BMI) was 22.1 kg/m^2 and 80.9% had a PS of 0 or 1. Anaemia was found in 49.4% of cases, an elevation of white cell count in 47.7%, of CRP level in 64.1%, of D-dimers in 21.5%, of LDH in 36.3%, of ALP in 33.7% and of calcium in 4.2%. Nearly all patients (96.6%) underwent bronchoscopy, which was normal in 26.9% of cases. Adenocarcinoma was the leading histological type (50.8%), followed by squamous cell carcinoma (30.2%).

Clinical staging at diagnosis showed that 22.9% of patients were at an early stage (I, II or IIIA), 15.3% were at a locally advanced stage (IIIB), and 61.8% had metastatic disease. Metastases involved mostly the contralateral lung (50.6%) and the adrenal glands (30.2%). Only 8.4% of patients underwent curative surgery; 63.1% received chemotherapy, 21.8% underwent radiotherapy, and 11.1% were offered exclusive supportive care (Table 1).

3.2. Univariable analysis

The mean follow-up time was 11.9 ± 0.8 months with a median of 8.2 months. On the study termination date, 23 patients (8.8%) were alive, 144 patients (55.0%) were deceased, and 95 patients (36.3%) were lost to follow-up. The median survival time in the whole study population was 13.8 ± 1.3 months (95%CI: 11.2–16.4 months) from presentation to death or last follow-up with a 5-year overall survival of 19.4%.

On univariable analysis, a longer survival was noted in patients aged < 65 years at diagnosis ($P = .019$), those smoking less than 45 pack-years ($P = .019$), and those having a BMI > 19 kg/m^2 ($P = .004$). The higher the PS, the shorter the survival. However, there was no clear association between chest symptoms/general signs and survival.

The absence of anaemia at presentation was associated with an improvement in survival for a haemoglobin cut-off value of 12 g/dL (17.1 versus 8.7 months, $P = .001$). A white cell count > 12 000/mm^3 and a CRP level > 20 mg/L were associated with a shorter survival, while platelet count did not have any impact. Additional biological factors – such as calcium, LDH and ALP levels – indicated a worse prognosis when their levels exceeded a certain cut-off value.

Endoscopy findings regarding the sites of lesions did not influence survival ($P = .873$), but the histological type had a significant effect