HHS Public Access

Author manuscript

J Thorac Oncol. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

J Thorac Oncol. 2015 November ; 10(11): 1576–1589. doi:10.1097/JTO.0000000000000652.

Refining prognosis in lung cancer: A report on the quality and relevance of clinical prognostic tools

Alyson L. Mahar, MSc[1], **Carolyn Compton, MD, PhD**[2], **Lisa M. McShane, PhD**[3], **Susan Halabi, PhD**[4], **Hisao Asamura, MD**[5], **Ramon Rami-Porta, MD, FETCS**[6], **Patti A. Groome, PhD**[1], and **on behalf of the Molecular Modellers Working Group of the American Joint Committee on Cancer**

[1]Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's University, ON; Canada

[2] Professor Life Sciences, Arizona State University and Professor of Laboratory Medicine and Pathology, Mayo Clinic School of Medicine, and Chair, Precision Medicine Core, American Joint Committee on Cancer 8th Edition Editorial Board

[3]Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD

[4]Department of Biostatistics and Bioinformatics, Duke University and Alliance Statistics and Data Center, Durham, NC

[5]Division of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan

[6]Department of Thoracic Surgery, Hospital Universitari Mutua Terrassa, CIBERES Lung Cancer Group, and Chair, International Association for the Study of Lung Cancer Staging and Prognostic Factors Committee, Terrassa, Barcelona, Spain Terrassa, Barcelona, Spain

Abstract

Introduction—Accurate, individualized prognostication for lung cancer patients requires the integration of standard patient and pathologic factors, biologic, genetic, and other molecular characteristics of the tumor. Clinical prognostic tools aim to aggregate information on an individual patient to predict disease outcomes such as overall survival, but little is known about their clinical utility and accuracy in lung cancer.

Methods—A systematic search of the scientific literature for clinical prognostic tools in lung cancer published Jan 1, 1996-Jan 27, 2015 was performed. In addition, web-based resources were searched. *A priori* criteria determined by the Molecular Modellers Working Group of the American Joint Committee on Cancer were used to investigate the quality and usefulness of tools. Criteria included clinical presentation, model development approaches, validation strategies, and performance metrics.

Results—Thirty-two prognostic tools were identified. Patients with metastases were the most frequently considered population in non-small cell lung cancer. All tools for small cell lung cancer covered that entire patient population. Included prognostic factors varied considerably across tools. Internal validity was not formally evaluated for most tools and only eleven were evaluated for external validity. Two key considerations were highlighted for tool development: identification

of an explicit purpose related to a relevant clinical population and clear decision-points, and prioritized inclusion of established prognostic factors over emerging factors.

Conclusions—Prognostic tools will contribute more meaningfully to the practice of personalized medicine if better study design and analysis approaches are used in their development and validation.

Keywords

lung cancer; prognosis; clinical prediction tools; prediction models; prognostic model

Introduction

Anatomic stage as classified by the Tumor Node Metastasis (TNM) system is considered the predominant prognostic factor in lung cancer.[1-3] However, the purpose of a staging system is to classify anatomical extent of disease, and in isolation is not sufficient for accurate survival probability prediction. [1,2,4-6] A wide variety of other prognostic information exists, including biologic, genetic, and other molecular characteristics of the tumor and standard clinical and pathologic factors. These factors can be considered *alongside* TNM,[7-9] to refine prognosis. For example, age, gender, performance status and tumor histology are established prognostic factors in lung cancer.[2,6]

Prognostic information arising from clinical, pathologic and molecular data can be combined with (or without) the TNM classification to create prognostic risk scores or groups.[4] If developed and properly validated, these tools can help clinicians provide a more accurate estimate of prognosis for the individual patient, as well to facilitate clinical decision making including primary and adjuvant disease management.[10,11]

Little is known about the accuracy or clinical usefulness of available prognostic tools in lung cancer. The Molecular Modellers Working Group (MMWG) of the American Joint Committee on Cancer (AJCC) was charged with understanding how to use information beyond stage to more accurately predict prognosis and thereby better guide personalized patient management. The MMWG identified the need to review currently available clinical prognostic tools in lung and four other cancers as their first task. The initial findings were presented at the American Society for Clinical Oncology in 2013.[12] This paper reports on the MMWGs findings in lung cancer.

Materials & Methods

The MMWG was a collaboration of surgeons, medical oncologists, pathologists, computational scientists, epidemiologists and biostatisticians with expertise in clinical and molecular model development working within the AJCC. It has since become two core groups (Precision Medicine Core and Evidence Based Medicine and Statistics Core) preparing for the 8th edition of the TNM staging classification system.[13] As a first step, the MMWG called for the investigation of current clinical prognostic tools for their potential to reliably predict survival outcome based on aggregate prognostic information.[12] A focus on survival prognostication was chosen because of its overarching importance and because it

has traditionally been used in the assessment of the prognostic value of TNM stage. The quality and clinical relevance of clinical prognostic tools were studied across five cancer sites (breast, colorectal, lung, melanoma and prostate). The results of the lung cancer study are reported here.

Systematic Literature Review & Search of the Web-Based Scientific Community

The search for prognostic tools and information on their development and validation was performed via three mechanisms: a search of the peer-reviewed published literature (which included a systematic literature review and cited reference search); a search of the web-based scientific community; and contacting tool developers for further information about development of publicly available web-based tools. Prognostic tools were defined as any nomogram, risk classification system, equation, risk score, electronic calculator, or other statistical regression model-based tool developed with the purpose of predicting time to death for use in clinical practice.[10] Prognostic tools in this paper include those developed to estimate the probability of survival at a particular point along the disease trajectory (e.g. at diagnosis, following treatment) or for the purpose of using a survival probability to inform treatment decision-making. Loosely speaking, there is some form of statistical model underlying most prognostic tools and we will use the terms prognostic tool and prognostic model interchangeably in many of the discussions here. The two main types of lung cancer, non-small cell and small cell histology, were considered separately.

The search strategy was executed in Medline, Embase and HealthStar to cover the period Jan 1, 1996-Jan 27, 2015. MESH headings do not exist for prognostic tools and so a combination of alternate MESH headings and key words were used following consultation with a health sciences librarian. An example of the search strategy used for the OVID Medline database is provided in Figure 1. Similar searches were conducted for the other databases using the appropriate syntax. Tools which may have been originally developed outside the literature search timeframe but that were identified in validation articles were considered clinically relevant and included. Seemingly eligible studies were excluded if they met any of the following *a priori* exclusion criteria: 1) assessment of the prognostic impact of a single factor (unless it was updating the accuracy of an existing prognostic tool); 2) inappropriate analytic purpose (e.g. multivariate modeling not aimed at prognostication, development of novel statistical methods); 3) not specific to lung cancer patients; 4) not original data/research (e.g. editorial, review) or 5) the outcome was not survival. Eligible survival end-points included all time-to-death analyses (e.g. overall survival, cause-specific survival), as well as vital status analyses (e.g. probability of being dead 5-years following diagnosis). The search strategy was not developed to identify studies developing genomic classifiers built entirely on gene expression data. These studies were excluded.

Citations were assessed for inclusion by a single reviewer (AM), first through their titles and abstracts and then as full articles. Early on, a random sample of 20 citations was independently re-evaluated by a blinded second reviewer (PG) and the results compared. Percent agreement was calculated to estimate inter-rater reliability. Percent agreement was high (>95%) and any differences identified in this exercise were discussed and resolved through consensus. Based on these findings, it was judged that the rules for inclusion and

exclusion were being applied consistently, and we proceeded to screen the larger group of eligible studies.

A cited reference search of eligible articles was conducted using Web of Science to identify other articles not found using the original search strategy. We also performed an online search for web-based clinical prognostic tools; both those identified through the primary literature search and those that were purely web-based. The search was performed using Google and search terms included: “clinical prediction tool cancer”, “online calculator cancer”, and “nomogram cancer”. Tool developers and/or the developer's institution were contacted if there were no peer-reviewed publications or technical documents available describing the tool's development process. A standard email and information query form was sent to these contacts through the auspices of the AJCC.

Data Abstraction

We developed a list of critical criteria for the adequate development and validation of clinical prognostic tools. The list was based on the work of Harrell et al.[14,15], guidelines provided by Bouwmeester et al.,[16] a textbook on clinical prediction model development and validation,[10] and on the REMARK reporting guidelines.[17] Successive drafts of the list were vetted by members of the MMWG and informed by discussion at the MMWG face to face meetings in 2009, 2010, and 2012. The final criteria are provided in the online appendix. At the time that the list was developed, TRIPOD, a reporting guideline for clinical prediction tools[18,19] and the CHARM checklist[20], a reporting guideline for the systematic review of clinical prediction tools had not been published. The list created by the MMWG includes all key criteria identified in both guidelines. Some of the criteria were common to both development and validation studies including study descriptors (e.g. authors, location, purpose), design, population characteristics, outcome measurement, standard clinical and pathology variables, and laboratory assay-based measurements. Information specific to tool development included candidate variables for the prognostic tool, selection of candidate variables, statistical modeling methodology, number of events, how missing data were handled, and a list of the final variables in the model. Information on internal validity and external validity assessments included the type of internal validation (e.g. apparent, cross-validation, bootstrapping), type of external validation (e.g. geographic, temporal, independent), and measures of internal and external validity (e.g. overall measures of model fit (e.g. Brier score), survival curves, calibration (e.g. calibration plot), discrimination (e.g. Harrell's c-index). Definitions for the key terms evaluated are provided in Box 1. The assessment of clinical usefulness included consideration of the clinical population targeted by the clinical prediction tool, face validity, the purpose of the tool, and its practicality. Clinical relevance was defined as those additional tool attributes outside the scope of development and validation methodology that were important for consideration. The criteria in this category informed the practicality and appropriateness of using the tool in a clinical setting. Clinical relevance was informally assessed by evaluating the choice of eligible and final prognostic factors, the clinical population addressed, the purpose and clinical question or decision-point targeted, and the format of the prognostic tool (e.g. was it available online, was the equation available for use in the clinic).

Summary of Data Quality

We report descriptive statistics on the development and validation of the eligible prognostic tools. We defined formal statistical evaluation of internal or external validity as the assessment of the tool's calibration and/or discriminative ability, which have been established as the best means of evaluating a prognostic model.[10] We also tracked whether tools were assessed through a comparison of survival time distributions across prognostic groups. This approach provides evidence that the prognostic tool creates monotonically increasing risk groups and is the same approach that is often used to evaluate the prognostic ability of TNM stage,[21] but it is not a replacement for measures of calibration and/or discrimination.

Results

Literature Search Results

Figure 2 outlines the published literature and web search. Overall, we identified thirty-three articles or technical documents[22-54] that supported the development of thirty-two clinical prognostic tools[22-49]; with three additional articles reporting only external validations,[50-52] another focused on assessing the incremental value of updating an existing prognostic tool with a new piece of information,[53] and another reported on comparing the prognostic accuracy of an existing, validated tool to the prognostic accuracy of a radiation oncologist.[54] Twenty-five tools were directed at prognosis in non-small cell lung cancer (NSCLC) and 7 in small cell lung cancer (SCLC). **[FIGURE 2]**

Tables 1a & 1b document key information abstracted on each tool. Sixteen tools were developed to predict overall survival, two for lung cancer-specific survival, two for vital status (alive at 2 years), and one for cumulative survival. The end-point for the survival analysis (e.g. death from any cause, death from cancer) was not specified in eleven tools. An index date to measure survival time (e.g. from the time of diagnosis) was not provided for 17/32 tools. Four tools predicted survival from the time of diagnosis, eight predicted survival from the start of a particular treatment, and one from the time of recurrence. Nineteen of the tools were developed to define risk categories, providing the least precise estimate of prognosis for an individual patient but potentially informing key decision points based on risk assessment. Four tools were available for use on the Internet, and of those, two were not associated with a peer-reviewed article describing its development. **[FIGURE 3]** Tool development occurred primarily in the United States (8/32), China (4/32), South Korea (3/32), the United Kingdom (3/32), the Netherlands (3/32), and Spain (3/32).

Tool Development Methods [TABLE 2]

None of the tools were developed from data prospectively collected specifically for the purpose of creating a prognostic tool. Nine tools used prospective data gathered for other purposes: seven used data collected for one or more randomized controlled trials and two used data aimed at investigating individual prognostic factors. Time period of data collection ranged from 1969 through 2013, with twenty-three tools (72%) developed using data collected on patients diagnosed ten or more years ago.

Table 2 documents that the rationale for prognostic variable selection was not provided for 18 of the 32 tools. Eight reported that literature-based reasoning and/or clinical relevance influenced variable choice and two reported choosing variables from those that were conveniently available. Four chose variables based on statistical associations with the outcome based on univariate analysis. Variable measurement methods were rarely described and operational definitions were rarely provided.

Table 2 also provides details on model development, including the choice of statistical approach for tool development, and how continuous variables and missing data were handled. Twenty-six of the 32 tools were built using the Cox proportional hazards model for time- to-event outcomes, two used recursive partitioning, two used support vector machines and one was built using regression tree methods. In 19 tools, continuous prognostic factors were categorized prior to inclusion in the tool (e.g. abnormal vs normal lab values) for the purpose of creating risk groups. Information on the handling of missing data was not provided for 20/32 tools (63%). For 7 tools (22%) patients with any missing data were excluded completely from the analysis. This approach is known as a 'complete case analysis' and can lead to inaccurate predictions of the outcome.[10,18,19]

Populations and Prognostic Factors NSCLC [FIGURE 4, TABLE 3]

In NSCLC, the most commonly addressed population was advanced or metastatic disease (16/25). The remaining NSCLC tools targeted patients with all stages of disease (n=1), patients treated with curative intent (n=2), or patients with stage I, II, or IIIa disease (n=6). Key time points in lung cancer where prognostic tools could be beneficial include decision-making to undergo definitive, inductive or adjuvant treatment or palliative management and at the time of recurrence or disease progression. Baseline prognosis was estimated across all stage categories. Prognostic tools designed for patients with operable tumours often aimed to identify high risk patients who would benefit from adjuvant therapy. The purpose of most tools in metastatic populations was to help physicians and patients make palliative management decisions by refining prognosis. The probability of survival estimated by many tools was used to inform particular treatment decisions, including whether or not to offer second line chemotherapy, and to identify high risk patients who may be considered for treatment with erlotinib.

There was considerable heterogeneity in the selection of prognostic factor in the tools that were reviewed, they contained many emerging factors and there was incomplete coverage of some established factors. In stage I-III NSCLC, many tools included less established prognostic factors that are expensive or difficult to measure, while, at the same time basic pathologic features with proven prognostic impact that can be easily determined in the resected specimens were missing. For example, the TNM optional factors such as vascular invasion, lymphovascular permeability and perineural invasion are considered established pieces of prognostic information for resected NSCLC,[3,55] but none of the tools included this information. Figure 4 describes the sixteen metastatic tools, showing that 22 prognostic factors were common to at least two of the sixteen tools. None of these prognostic factors were included across all sixteen tools; however, when the target population definition ensures no heterogeneity of a specific prognostic factor (e.g. Stage IV M1 population and no

heterogeneity on the prognostic factor TNM stage), then it could be appropriate for this variable to be absent from the prognostic tool. Performance status was the most commonly included prognostic factor, incorporated in 10 tools. Thirty-three additional prognostic factors were included in only one of each of the sixteen tools (Table 3).

Populations and Prognostic Factors SCLC [FIGURE 5, TABLE 3]

In SCLC, all seven clinical prognostic tools were developed for use in the general SCLC population to refine prognosis at the time of diagnosis. None targeted a particular sub-population or time-point that may have benefitted from a tool due to medical decision-making uncertainty. These seven tools were developed using 14 different prognostic factors. Figure 5 shows that performance status and stage were common to seven and five tools respectively, and four other factors were common to at least two tools. Table 3 lists eight other factors that were included in only one tool.

Internal Validity [TABLE 4]

Twenty-eight tool development articles included evaluations of internal validity, but 18 of these used 100% of the data that were used to develop the model (which is defined as “apparent” internal validation) and 4 randomly split their sample. The use of “apparent” internal validation techniques lead to overly optimistic performance estimates.[10] Three internal validation analyses used cross-validation, and three used boot-strapping, both of which are more appropriate established methods.[18,19] Cross-validation iteratively splits the original sample into training (for model development) and testing (for model validation) sets to estimate performance of the tool. Bootstrapping follows a similar process but defines the training set by drawing data with replacement from the full data set (same data can be represented more than once in the training set). Sixteen of these 28 internal validations did not measure or report calibration or discrimination of the model, but instead purported to assess validity by establishing that there was a statistically significant difference in survival time distribution between risk groups defined by the prognostic tool. A Brier score was used to measure model performance in one tool, ranging from 0.119-0.162 across subgroups (smaller is better). Performance of four prognostic tools was assessed using an R^2 statistic, a measure of variation in the outcome predicted by the tool, with all values being less than 0.31. Calibration was evaluated in the development of three tools, two with results provided as graphs of predicted versus observed survival (calibration plots). Discrimination was evaluated in twelve tools with concordance statistic that ranged from 0.64-0.83. No temporal trends in the inclusion of an internal validation assessment were noted.

External Validity [TABLE 4]

The majority of prognostic tools (21/32) were not evaluated in an independent sample from that used for tool development (external validation). Ten articles performed 17 assessments of the external validity of eleven tools. Of the eleven tools with evidence of external validation, three were evaluated by reporting the statistical significance of the separation of survival curves by risk strata with no formal measures of calibration, discrimination or other valid ways of assessing predictive ability reported. Four tools were evaluated using Brier Score (0.071-0.163) or R^2 (0.08-0.343). Calibration was assessed in two tools, one via a calibration plot and one through the informal comparison of predicted versus observed

survival probabilities. Five tools were evaluated for their discrimination ability; concordance statistic ranged from 0.687-0.76. de Jong et al. (2007) evaluated the ability of three different tools to distinguish between two patients with better or worse prognosis, but did not formally calculate a concordance statistic. No trends in the performance of external validations were noted over time.

Discussion

This study described the clinical prognostic tool landscape in lung cancer. We identified 32 clinical prognostic tools from the peer-reviewed literature and web-based resources. Metastatic disease was the most commonly considered clinical population in NSCLC, and all tools in SCLC were intended for the entire population of patients with small cell lung cancer. There was significant heterogeneity in the prognostic factors included. Most tool developers did not conduct a formal evaluation of the internal validity of the underlying statistical model. Nine tools were evaluated for external validity, with varying degrees of rigor.

This study supports conclusions of previous investigations regarding the methodological quality of clinical prediction tools in other clinical settings: the methodology used to develop and validate many tools is poor, with little reliable external generalizability or attention to the impact these tools have on clinical decision-making and patient-outcomes.[11,16,56-63] The accuracy of the tools' predictions was generally insufficient to use them to justify deviating from standard clinical decisions. These findings provide further emphasis to the widespread recognition that methodological improvements are required to optimize clinical prediction tools for patient care.[11,16,19,58-63] Excellent guidance on best practices for the conduct of prognostic research, tool development, and prognostic study reporting has been published.[16,19,56-61] More leadership in the promotion of these methodological requirements to tool developers, tool resources and scientific journals is badly needed.

Previous reviews of clinical prediction tool methodology have not specifically evaluated lung cancer prognostic tools to gain a clinically relevant understanding of their particular strengths and weaknesses. This review identified that improvement in the development, validation and use of prognostic tools for survival in patient with lung cancer will require addressing the choice of relevant clinical populations and clinical management decision-points, and the significant heterogeneity in the consideration and inclusion of prognostic factors. This targeted information is necessary to build on the methodological findings of this review and guide future directions in clinical prognostic tool development and implementation in lung cancer patients.

In addition to prognostication at diagnosis, multiple prognostic judgements within sub-populations are also necessary as the disease progresses (or regresses), both to re-estimate prognosis and help to inform treatment decisions. This review identified that none of the tools developed for use in SCLC were targeted to a particular clinical management decision-point. Similarly, many in NSCLC were targeted to refining prognosis at the time of diagnosis. Stage-specific prognostication at diagnosis for small cell and non-small cell lung cancer could be improved by the combination of clinical, pathological and biological

factors. For example, for a T1aN0M0 NSCLC patient, the addition of pathological factors (perineural invasion[64], vascular invasion[65] or lymphatic permeation[66]), and established clinical factors (age, performance status) may alter their prognosis at diagnosis from relatively good to much worse.

We also propose that successive prognostic tools are needed along the disease trajectory. None of the prognostic tools in SCLC targeted any particular clinical management decision; while tools in NSCLC were designed for multiple purposes. Therefore, gaps in the coverage of key clinical decision points exist. Such individualized prognostication across the trajectory could provide information that a patient and their family may want for planning and has the potential to better inform management decisions. For example, before any surgical treatment has been done, post-operative 5-year survival based on TNM stage alone for an individual patient may be 90%. In the same patient, a postoperative prognostic tool that included the pathological TNM, the definitive histopathological type and EGFR mutation status, for example, would modify the prognostic assessment of the disease. The tool may drop that individual's personal estimate to 40% and the decision to move forward with surgery may change based on this more personalized prognosis.

Consistency and a balance between the practical and the ideal is needed when identifying relevant prognostic factors for inclusion in prognostic tools for lung cancer, regardless of the clinical population addressed or the decision-point targeted. There was substantial heterogeneity in prognostic factor choices in the tools we reviewed, even within similar clinical presentations. Tools also contained many emerging factors, the coverage of some established factors was not in evidence, and many of the tools included expensive, difficult to measure factors that are not reliably or routinely collected. Although the number of established prognostic factors in lung cancer is small,[3,55] consideration of fundamental prognostic knowledge during tool development is vital. Large scale, standardized data sharing agreements, such as that led by the IASLC[21,67,68] could better supply physicians and scientists with the appropriate information needed to develop high quality prognostic tools.

This study may underestimate the number of prognostic tools developed in lung cancer during the study time frame. The lack of standardized MESH headings for this type of study limited the ability to find all relevant prognostic tools. For example, the study by Blanchon et al (2006) was not identified through the search terms used in this systematic review. However, we applied many methods to optimize the capture of prognostic tools in lung cancer, including consultation with a health sciences librarian for the systematic review, cited reference searches and a search of online resources.

The existing clinical prognostic tool literature in lung cancer has both methodological flaws and clinical challenges. The future of prognostic tool development, validation and use in patients with lung cancer must begin with the identification of a clear, clinical objective, targeting a precise decision-point within the disease trajectory. High quality development and validation methods for prognostic tools that build upon established prognostic information will improve the accuracy of individualized estimates of prognosis, and provide necessary credibility for their implementation into clinical practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The American Joint Committee on Cancer (AJCC) provided a contract to Patti Groome and Alyson Mahar to support the work of identifying and evaluating existing prognostic tools for lung cancer as a preparatory step in the AJCC's development of prognostic tools for major cancers.

References

1. Goldstraw P, Ball D, Jett JR, et al. Non-small-cell lung cancer. Lancet. 2011; 378:1727. [PubMed: 21565398]
2. van Meerbeeck JP, Fennell DA, De Ruysscher DKM. Small-cell lung cancer. Lancet. 2011; 378:1741. [PubMed: 21565397]
3. Edge, SB.; American Joint Committee on Cancer, Teton Data Systems (Firm). et al. AJCC Cancer Staging Manual. Springer; New York: 2010.
4. Greene FL, Sobin LH. A worldwide approach to the TNM staging system: collaborative efforts of the AJCC and UICC. J Surg Oncol. 2009; 99:269–272. [PubMed: 19170124]
5. Sculier J, Chansky K, Crowley JJ, et al. The impact of additional prognostic factors on survival and their relationship with the anatomical extent of disease expressed by the 6th Edition of the TNM Classification of Malignant Tumors and the proposals for the 7th Edition. J Thorac Oncol. 2008; 3:457–466. [PubMed: 18448996]
6. Chansky K, Sculier J, Crowley JJ, et al. The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM stage in surgically managed non- small cell lung cancer. J Thorac Oncol. 2009; 4:792–801. [PubMed: 19458556]
7. Gospodarowicz MK, Miller D, Groome PA, et al. The process for continuous improvement of the TNM classification. Cancer. 2004; 100:1–5. [PubMed: 14692017]
8. Sobin LH. TNM: principles, history, and relation to other prognostic factors. Cancer. 2001; 91:1589–1592. [PubMed: 11309755]
9. Hermanek P, Sobin LH, Fleming ID. What do we need beyond TNM? Cancer. 1996; 77:815–817. [PubMed: 8608468]
10. Steyerberg, EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer New York; New York, NY: 2009.
11. Vickers AJ. Prediction models in cancer care. CA Cancer J Clin. 2011; 61:315–326. [PubMed: 21732332]
12. Mahar AL, Halabi S, McShane L, et al. A survey of clinical prediction tools in colorectal and lung cancers and melanoma. J Clin Oncol. 2013; 31:1592. [PubMed: 23509311]
13. American Joint Committee on Cancer. [Sept 16, 2014] AJCC- 8th Edition Updates. Apr 23. 2014 2014. Available at: https://cancerstaging.org/About/Pages/8th-Edition.aspx. 2014
14. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996; 15:361–387. [PubMed: 8668867]
15. Harrell, FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer; New York: 2001.
16. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012; 9:1. [PubMed: 22629234]
17. McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst. 2005; 97:1180–1184. [PubMed: 16106022]
18. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and ElaborationThe

TRIPOD Statement: Explanation and Elaboration. Ann Intern Med. 2015; 162:W1–W73. [PubMed: 25560730]

19. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med. 2015; 162:55–63. [PubMed: 25560714]
20. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014; 11:e1001744. [PubMed: 25314315]
21. Goldstraw P, Crowley J, Chansky K, et al. The IASLC Lung Cancer Staging Project: Proposals for the Revision of the TNM Stage Groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours. J Thorac Oncol. 2007; 2:706–714. [PubMed: 17762336]
22. Adjuvant! for Lung Cancer (NSCLC). 2011. Available at: www.adjuvantonline.com. 2012
23. Cerny T, Blair V, Anderson H, et al. Pretreatment prognostic factors and scoring system in 407 small-cell lung cancer patients. Int J Cancer. 1987; 39:146–149. [PubMed: 3026969]
24. de Jong WK, Fidler V, Groen HJM. Prognostic Classification with Laboratory Parameters or Imaging Techniques in Small-Cell Lung Cancer. Clin Lung Cancer. 2007; 8:376–381. [PubMed: 17562238]
25. Dehing-Oberije C, Aerts H, Yu S, et al. Development and Validation of a Prognostic Model Using Blood Biomarker Information for Prediction of Survival of Non–Small-Cell Lung Cancer Patients Treated With Combined Chemotherapy and Radiation or Radiotherapy Alone (NCT00181519, NCT00573040, and NCT00572325). Int J Radiat Oncol Biol Phys. 2011; 81:360–368. [PubMed: 20888135]
26. Dehing-Oberije C, Yu S, De Ruysscher D, et al. Development and External Validation of Prognostic Model for 2-Year Survival of Non-Small-Cell Lung Cancer Patients Treated With Chemoradiotherapy. Int J Radiat Oncol Biol Phys. 2009; 74:355–362. [PubMed: 19095367]
27. Di Maio M, Lama N, Morabito A, et al. Clinical assessment of patients with advanced non- small-cell lung cancer eligible for second-line chemotherapy: a prognostic score from individual data of nine randomised trials. Eur J Cancer. 2010; 46:735–743. [PubMed: 20045311]
28. Florescu M, Hasan B, Seymour L, et al. A clinical prognostic index for patients treated with erlotinib in National Cancer Institute of Canada Clinical Trials Group study BR.21. J Thorac Oncol. 2008; 3:590–598. [PubMed: 18520796]
29. Gagnon B, Agulnik JS, Gioulbasanis I, et al. Montreal prognostic score: estimating survival of patients with non-small cell lung cancer using clinical biomarkers. Br J Cancer. 2013; 109:2066. [PubMed: 24064979]
30. Hoang T, Xu R, Schiller JH, et al. Clinical model to predict survival in chemonaive patients with advanced non-small-cell lung cancer treated with third-generation chemotherapy regimens based on eastern cooperative oncology group data. J Clin Oncol. 2005; 23:175–183. [PubMed: 15625371]
31. Hoang T, Dahlberg SE, Sandler AB, et al. Prognostic Models to Predict Survival in Non–Small-Cell Lung Cancer Patients Treated with First-Line Paclitaxel and Carboplatin with or without Bevacizumab. J Thorac Oncol. 2012; 7:1361–1368. [PubMed: 22843087]
32. Hong S, Cho BC, Choi HJ, et al. Prognostic factors in small cell lung cancer: a new prognostic index in Korean patients. Oncology. 2010; 79:293–300. [PubMed: 21412014]
33. Kawahara M, Fukuoka M, Saijo N, et al. Prognostic factors and prognostic staging system for small cell lung cancer. Jpn J Clin Oncol. 1997; 27:158–165. [PubMed: 9255270]
34. Kim ST, Lee J, Sun J, et al. Prognostic model to predict outcomes in non-small cell lung cancer patients with erlotinib as salvage treatment. Oncology. 2010; 79:78–84. [PubMed: 21071994]
35. Lopez-Encuentra A, Lopez-Rios F, Conde E, et al. Composite anatomical-clinical-molecular prognostic model in non-small cell lung cancer. Eur Respir J. 2011; 37:136–142. [PubMed: 20817702]
36. Maestu I, Pastor M, Gómez-Codina J, et al. Pretreatment prognostic factors for survival in small-cell lung cancer: a new prognostic index and validation of three known prognostic indices on 341 patients. Ann Oncol. 1997; 8:547–553. [PubMed: 9261523]

37. Mandrekar SJ, Schild SE, Hillman SL, et al. A prognostic model for advanced stage nonsmall cell lung cancer. Cancer. 2006; 107:781–792. [PubMed: 16847887]
38. Mou W, Liu Z, Luo Y, et al. Development and cross-validation of prognostic models to assess the treatment effect of cisplatin/pemetrexed chemotherapy in lung adenocarcinoma patients. Med Oncol. 2014; 31:1–9.
39. Will your patient benefit from post-operative radiotherapy?. 2008. Available at: http://skynet.ohsu.edu/nomograms/postrt/nsc-lung.html. 2012
40. Padilla J, Calvo V, Peñalver JC, et al. Survival and risk model for stage IB non-small cell lung cancer. Lung Cancer. 2002; 36:43–48. [PubMed: 11891032]
41. Park MJ, Jae Park M, Lee J, et al. Prognostic model to predict outcomes in nonsmall cell lung cancer patients treated with gefitinib as a salvage treatment. Cancer. 2009; 115:1518–1530. [PubMed: 19177485]
42. Putila J, Remick SC, Guo NL. Combining clinical, pathological, and demographic factors refines prognosis of lung cancer: a population-based study. PloS One. 2011; 6:e17493. [PubMed: 21364765]
43. Souhami RL, Bradbury I, Geddes DM, et al. Prognostic significance of laboratory parameters measured at diagnosis in small cell carcinoma of the lung. Cancer Res. 1985; 45:2878. [PubMed: 2985256]
44. Tibaldi C, Vasile E, Bernardini I, et al. Baseline elevated leukocyte count in peripheral blood is associated with poor survival in patients with advanced non-small cell lung cancer: a prognostic model. J Cancer Res Clin Oncol. 2008; 134:1143–1149. [PubMed: 18347812]
45. Ulas A, Turkoz FP, Silay K, et al. A laboratory prognostic index model for patients with advanced non-small cell lung cancer. PloS One. 2014; 9:e114471. [PubMed: 25474743]
46. Vincent MD, Ashley SE, Smith IE. Prognostic factors in small cell lung cancer: a simple prognostic index is better than conventional staging. Eur J Cancer Clin Oncol. 1987; 23:1589. [PubMed: 2828070]
47. Wigren T, Oksanen H, Kellokumpu-Lehtinen P. A practical prognostic index for inoperable non-small-cell lung cancer. J Cancer Res Clin Oncol. 1997; 123:259–266. [PubMed: 9201248]
48. Wojas-Krawczyk K, Krawczyk P, Mlak R, et al. The applicability of a predictive index for second- and third-line treatment of unselected non-small-cell lung cancer patients. Respiration. 2011; 82:341–350. [PubMed: 21242665]
49. Zhang K, Bar Ad V, Li B, et al. Modeling the overall survival of patients with advanced-stage non-small cell lung cancer using data of routine laboratory tests. Int J Cancer. 2015; 136:382–391. [PubMed: 24866905]
50. Di Maio M, Krzakowski M, Fougeray R, et al. Prognostic score for second-line chemotherapy of advanced non-small-cell lung cancer: external validation in a phase III trial comparing vinflunine with docetaxel. Lung Cancer. 2012; 77:116. [PubMed: 22361218]
51. Wang F, Zhang Y, Zhao H, et al. Validation of a clinical prognostic model in Chinese patients with metastatic and advanced pretreated non-small cell lung cancer treated with gefitinib. Med Oncol. 2011; 28:331–335. [PubMed: 20204544]
52. Wigren T. Confirmation of a prognostic index for patients with inoperable non-small cell lung cancer. Radiother Oncol. 1997; 44:9–15. [PubMed: 9288851]
53. Putila J, Guo NL. Combining COPD with clinical, pathological and demographic information refines prognosis and treatment response prediction of non-small cell lung cancer. PloS One. 2014; 9:e100994. [PubMed: 24967586]
54. Oberije C, Steyerberg E, Dingemans A, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. Radiother Oncol. 2014; 112:37–43. [PubMed: 24846083]
55. Goldstraw, P. IASLC Staging Manual in Thoracic Oncology. 1st edn.. Editorial RX Press; Florida: 2009.
56. Mallett S, Royston P, Dutton S, et al. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med. 2010; 8 20-7015-8-20.
57. Mallett S, Royston P, Waters R, et al. Reporting performance of prognostic models in cancer: a review. BMC Med. 2010; 8 21-7015-8-21.

58. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013; 346:e5595. [PubMed: 23386360]
59. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ. 2013; 346:e5793. [PubMed: 23386361]
60. Riley RD, Altman DG, Hemingway H, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013; 10:e1001380. [PubMed: 23393429]
61. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013; 10:e1001381. [PubMed: 23393430]
62. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012; 98:691–698. [PubMed: 22397946]
63. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012; 98:683–690. [PubMed: 22397945]
64. Yilmaz A, Duyar SS, Cakir E, et al. Clinical impact of visceral pleural, lymphovascular and perineural invasion in completely resected non-small cell lung cancer. Eur J Cardiothorac Surg. 2011; 40:664–670. [PubMed: 21334917]
65. Shimada Y, Ikeda N, Saji H, et al. Pathological Vascular Invasion and Tumor Differentiation Predict Cancer Recurrence in Stage ia Non–Small-Cell Lung Cancer After Complete Surgical Resection. J Thorac Oncol. 2012; 7:1263–1270. [PubMed: 22673056]
66. Matsumura Y, Hishida T, Shimada Y, et al. Impact of Extratumoral Lymphatic Permeation on Postoperative Survival of Non–Small-Cell Lung Cancer Patients. J Thorac Oncol. 2014; 9:337–344. [PubMed: 24496002]
67. Giroux DJ, Rami-Porta R, Chansky K, et al. The IASLC Lung Cancer Staging Project: data elements for the prospective project. J Thorac Oncol. 2009; 4:679–683. [PubMed: 19461401]
68. Rami-Porta R, Bolejack V, Giroux DJ, et al. The IASLC lung cancer staging project: the new database to inform the eighth edition of the TNM classification of lung cancer. J Thorac Oncol. 2014; 9:1618–1624. [PubMed: 25436796]
69. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology. 2010; 21:128–138. [PubMed: 20010215]
70. Williams BA, Sugimura H, Endo C, et al. Predicting Postrecurrence Survival Among Completely Resected Nonsmall-Cell Lung Cancer Patients. Ann Thorac Surg. 2006; 81:1021–1027. [PubMed: 16488713]

Box 1

Definitions of key terms used in developing and validating prognostic models[10,69]

Term	Definition
Model Performance	Model performance refers to the ability of the underlying statistical model or algorithm to predict the outcome of interest (e.g. overall survival). Two important aspects of performance are calibration and discrimination. Performance should be evaluated both during the model development process to assess *internal validity* as well as in an external population to assess *external validity*. A prognostic tool that does not have evidence of both internal and external validity cannot be relied upon for accurate estimation of individualized prognosis.
Internal Validity	In prognostic modeling, internal validity is assessed by testing model performance during tool development, using the same sample of patients. This quantifies the consistency of the model within the study sample. Types of internal validation measurement methods: • *Apparent Validation*: using the same sample that the model was developed in. • *Split Sample*: randomly splitting the sample in half at the beginning of the study, using half of the population to develop the model, and measuring the model performance in the other half. • *Cross-Validation*: consecutively in a random part of the sample, with model development in other parts (e.g. tenfold: sample split into parts, model is developed in 9/10 and validated in the remaining 1/10; this process is repeated until all subsets have validated the model). • *Bootstrap*: uses multiple samples drawn with replacement from the original full sample. Model is developed on the selected subsample and evaluated on the corresponding non-selected subsample. Performance estimates from the multiple iterations of the subsampling process are averaged to give an overall measure of performance for the model.
External Validation	In prognostic modeling, external validity is assessed by testing the model performance in plausibly similar samples of patients that did not contribute to the development data. This may be measured in a population from a different geographic area (*geographic external validation*); in a different time period from model development (*temporal validation*) or both. This may also be done independently by a research group with no affiliation to the tool developers (*fully independent validation*).
Calibration	Calibration is a measure of agreement between the outcomes observed in the data for individual patients, with the outcome values predicted for individual patients by the statistical model. For prognostic tools of time to event outcomes, this is typically measured at a particular time point. A number of different methods may be used to evaluate calibration of a prognostic tool; however, the most common is a *calibration plot* (apparent calibration). A calibration graph plots the predictions of the model on the x-axis and the observed outcomes on the y-axis. Graphical presentation of calibration, assessed visually is a common form of evaluation for prognostic tools. For other methods of evaluating calibration, see references [10,69].
Discrimination	Discrimination is a measure of how well a model can discern between individuals with and without the outcome or event. In prognostic tools, it is more often a measure of how well the model can rank a pair of individuals such that the individual predicted to survive longer is the individual who actually survived longer. Typically, this is measured using *Harrell's overall c statistic*. For other methods of evaluating discrimination, see references [10,69].

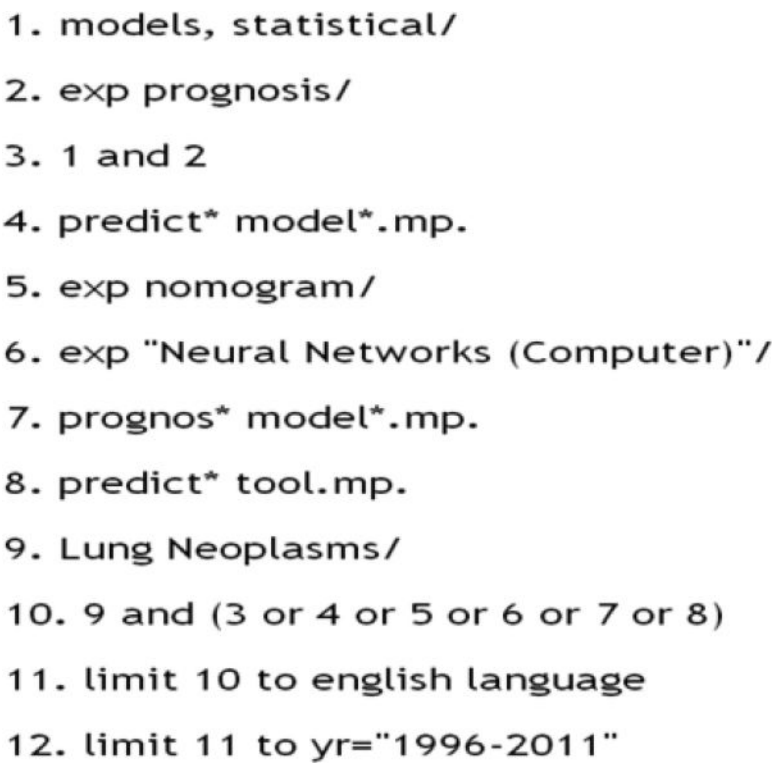

Figure 1.
Example of the systematic literature search strategies used to identify clinical prognostic tools and articles evaluating their validation in lung cancer

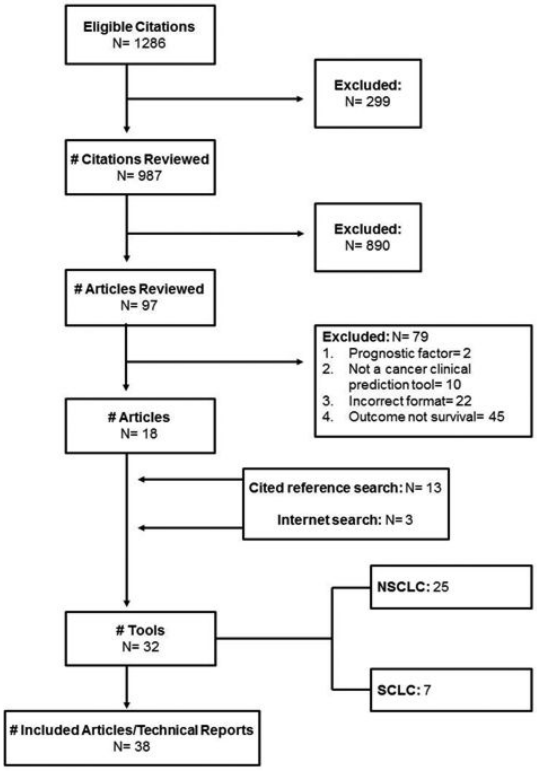

Figure 2.
Results of the search for clinical prognostic tools and their validation articles in lung cancer.

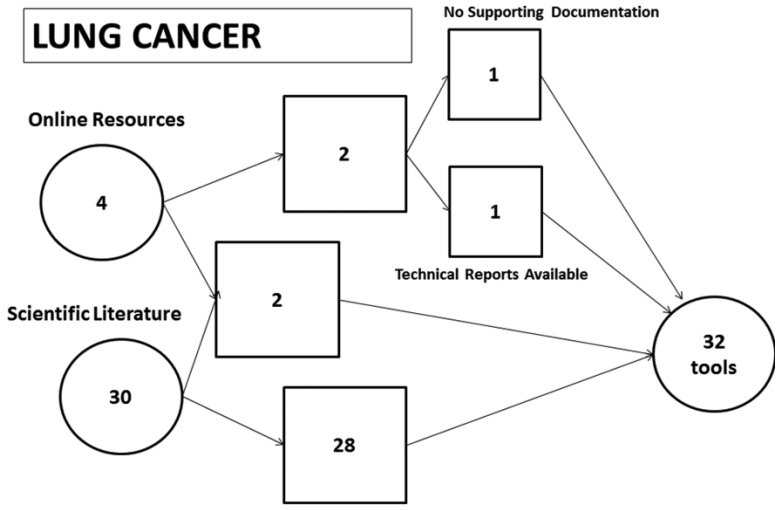

Figure 3.

The identification of tools related to survival outcomes accessing both the scientific literature and web-based resources

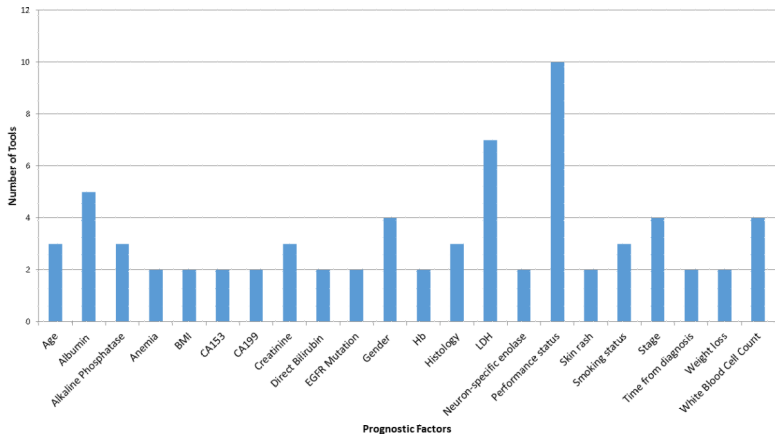

Figure 4.

Predictors used in more than one clinical prognostic tool for survival in advanced/incurable NSCLC (n=12 tools)*

* See Table 3 for predictors used in only one tool

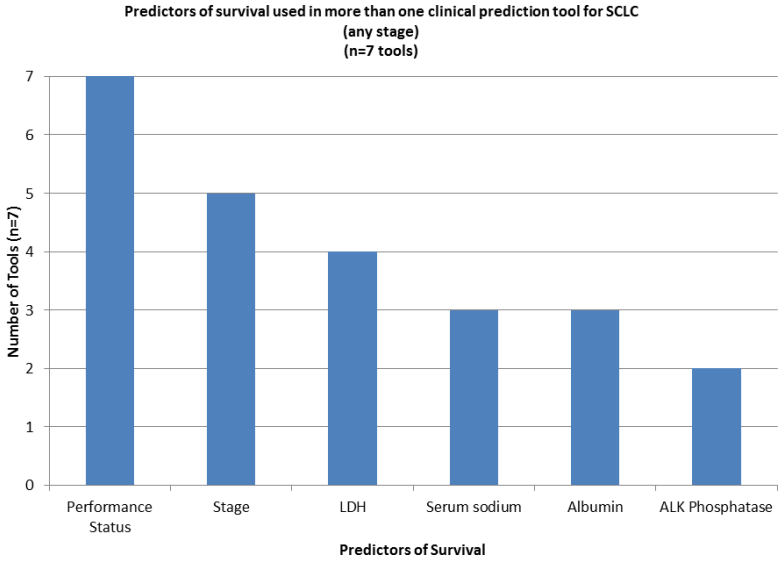

Figure 5.

Predictors used in more than one clinical prognostic tool for survival SCLC (any stage) (n=7 tools)*

* See Table 3 for predictors used in only one tool

Table 1a

Details on included small cell lung cancer studies (all tools developed for all stages)

Tool Citation	Year of Publication	Dates of Data Collection	Study Design	Sample Size	Sample Size/ Power Calculation Performed	Events	Duration of Follow-Up	Outcome	Stage (%)	Final Variables in Model	Internal Validation	External Validation
[23]	1987	1979-1985	Retrospective Cohort	407	NR	NR	NR	NOS	Limited: 60 Extensive: 40	Lactate dehydrogenase, extensive stage, serum sodium, pre-treatment Karnofsky score, Alkaline phosphatase, bicarbonate	None	None
[24]	2007	1996-2004	Retrospective Cohort	156	No	NR	NR	NOS	Limited: 40 Extensive: 60	Stage, performance status, WBC count, platelet count, serum Lactate dehydrogenase, age	Approach: Apparent Discrimination: **	None
[32]	2010	2002-2007	Retrospective Cohort	295	No	NR	Median 9.4 months	OS	Limited: 44.4 Extensive: 55.6	Stage, CFRYA 21-1, performance status	Approach: Apparent *	None
[33]	1997	1985-1988	RCT-PC	286	No	NR	NR	NOS	Limited: 51 Extensive: 49	Serum Lactate dehydrogenase, performance status, serum sodium	Approach: Apparent *	Overall: R^2=0.13 Discrimination: **
[36]	1997	1981-1993	Retrospective Cohort	341	No	NR	Minimum 1 year	NOS	Limited: 62 Extensive: 38	Lactate dehydrogenase, albumin, neutrophils, extended versus limited disease, ECOG Performance status	Approach: Apparent *	None
[43]	1985	1979-1982	RCT-PC	371	No	NR	NR	CS	NR	Performance status, alkaline phosphatase, disease extent, plasma albumin, plasma sodium	Approach: Apparent *	Overall: R^2=0.13 Discrimination: **
[46]	1987	1978-1985	Retrospective Cohort	177	No	NR	NR	NOS	NR	Plasma albumin, liver scans, alanine transaminase, performance status	None	Overall: R^2=0.08 Discrimination: **

NR= Not Reported RCT-PC= secondary use of randomized controlled trial data; NOS= survival not otherwise specified; OS= overall survival; CS= cumulative survival; ECOG= Eastern Cooperative Oncology Group

* Kaplan-Meier survival curves presented

** Provided a table reporting the ability of the model to rank individuals by their survival time and accurately distinguish them into risk groups, but did not calculate the concordance statistic

Table 1b

Details on included non-small cell lung cancer tools

Tool Citation	Population	Year of Publication	Dates of Data Collection	Study Design	Sample Size	Sample Size/Power Calculation Performed	Events	Duration of Follow-Up	Outcome	Stage (%)	Final Variables in Model	Internal Validation	External Validation
Stage I-IV													
[42]	Stage I-IV	2011	1988-2006	Retrospective Cohort	150158	NR	NR	NR	DSS		Tumor stage, grade, age, race and gender	Approach: Apparent Discrimination: 0.72-0.763	Discrimination: 0.687-0.721
Stage I-III													
[40]	Stage IB	2002	1969-1998	Retrospective Cohort	659	No	NR	NR	OS	IB: 100	Tumor size, cell type	Approach: Split Sample*	None
[22]	Stage I-IIIa	2004	1988-2004	Other	17,310	NR	NR	5 years minimum	DSS	NR	Age, sex, comorbidity, pathological T, pathological N, histologic grade, tumor diameter, adjuvant therapy	None	None
[70]	Curative intent	2006	1997-2001	Prospective Cohort	390	No	NR	NR	NOS	IA: 19 IB: 22 IIA: 6 IIB: 16 IIIA: 26 IIIB: 11	Performance status at recurrence, symptoms at recurrence, liver recurrence, stage, number of recurrences	Approach: Apparent Discrimination: 0.70	None
[39]	Curative intent	2008	NR	Retrospective Cohort	NR	NR	NR	NR	OS	NR	Age, sex, depth of tumor invasion, nodal status, histology	None	None
[26]	Stage I-IIIB	2009	2002-2006	Retrospective Cohort	322	NR	NR	Median 4.1 years	OS	I: 23 II: 9 IIIA: 25 IIIB: 42 Missing: 1	Gender, WHO-PS, FEV1, GTV, PLNS	Approach: Cross-Validation Discrimination: 0.74	Discrimination: 0.75-0.76
[35]	Stage I-II	2011	1993-1997	Prospective Cohort	512	Yes	NR	Mean 120 months	OS	pT1: 20.9 pT2: 71.3 pT3: 7.8 pN0: 84 pN1: 16	Nodule in the same lobe, tumor size, pTdi, proximal bronchus, atelectasis-pneumonitis, arterial hyptertension, age, performance status, smoking status, previous tumor, COPD, haemoglobin, phopho-	Approach: Bootstrap Discrimination: 0.66	None

Tool Citation	Population	Year of Publication	Dates of Data Collection	Study Design	Sample Size	Sample Size/Power Calculation Performed	Events	Duration of Follow-Up	Outcome	Stage (%)	Final Variables in Model	Internal Validation	External Validation
											ACCC, Ki67, P63, E-cadherin, phospho-mTOR, p27, NF-xB		
[25]	Stage I-IIIB	2011	2004-2007	Retrospective Cohort	106	NR	71	Median 38 months	OS	I: 17 II: 9 IIIA: 23 IIIB: 51	Gender, WHO-PS, FEV1, GTV, PLNS	Approach: Cross-Validation Discrimination: 0.76	None
[38]	Stage I-IIIB	2014	2008-2013	Retrospective Cohort	53	NR	NR	NR	NOS	NR	Serum total protein, age, total triglyceride, albumin, gender, uric acid, CYFRA21-1	Approach: Apparent*	None
Advanced/Incurable Disease													
[47]	Inoperable	1997	1974-1981	Retrospective Cohort	502	No	NR	NR	OS	NR	Feinstein symptom score, stage, Karonsofsky score, hemoglobin, tumor size	Approach: Apparent*	Calibration: Y
[37]	Stage IIIB/Stage IV	2006	1985-2001	RCT-PC	782	Yes	NR	NR	OS	IIIB: 11 IV: 89	Gender, age, performance status, stage, BMI, white blood cell count, hemoglobin level, creatinine level	Approach: Boostrap Calibration: Y Discrimination: 0.65	Calibration: Y
[28]	Stage IIIB/IV	2008	NR	RCT-PC	485	No	376	NR	OS	NR	ECOG PS, smoking history, weight loss, anemia, Lactate dehydrogenase, time from diagnosis, response to prior treatment, EGFR-FISH gene copy number, ethnicity, number of prior regimens	Approach: Apparent*	None
[30]	Advanced	2008	1993-1999	RCT-PC	1436	No	NR	NR	NOS	IIIB (with effusion) : 8 IV: 92	Subcutaneous metastasis, performance status, loss of appetite, liver metastasis, # metastatic sites, previous lung surgery	Approach: Split Sample Calibration: Y	None
[44]	Stage IIIb/IV	2008	1994-2005	Retrospective Cohort	320	No	280	NR	OS	IIIB: 26.6 IV: 73.4	Performance status, leukocyte count, histology, brain metastases	Approach: Apparent*	None

Tool Citation	Population	Year of Publication	Dates of Data Collection	Study Design	Sample Size	Sample Size/Power Calculation Performed	Events	Duration of Follow-Up	Outcome	Stage (%)	Final Variables in Model	Internal Validation	External Validation
[41]	Metastatic/ recurrent	2009	2002-2005	Retrospective Cohort	316	No	290	Median 47.4 months	OS	IIIB: 19 IV: 81	ECOG Performance status, presence of intra-abdominal metastasis, Alkaline phosphatase, time interval from diagnosis to geftinib therapy, serum albumin, progression-free time during previous chemotherapy, white blood cell count, smoking status	Approach: Apparent*	None
[27]	Advanced	2010	1999-2007	RCT-PC	1197	No	956	NR	OS	IIIB: 17.8 IV: 82.2	Gender, performance status, tumor stage, histologic type, type of first-line therapy, objective response to first-line	Approach: Apparent Discrimination: 0.643	None
[34]	Advanced/ metastatic	2010	2006-2008	Retrospective Cohort	257	No	NR	NR	OS	IIIB: 7 IV: 93	ECOG performance status, serum Lactate dehydrogenase, skin rash	Approach: Bootstrap*	None
[48]	Advanced IIIb/IV	2011	NR	Retrospective Cohort	73	No	NR	NR	OS	IIIB: 29 IV: 71	Performance status, rash, time from diagnosis, weight loss, gender, Lactate dehydrogenase level, time from first-line chemotherapy, smoking status, EGFR mutation, anemia	Approach: Apparent*	None
[31]	Stage IIIb/IV	2012	NR	RCT-PC	850	NR	NR	NR	OS	IIIB: 12 IV/Recurrent: 88	Skin metastasis, low BMI (<18.5), high serum Lactate dehydrogenase, adrenal metastasis, performance status, low serum albumin, sex, bone metastasis, histology, mediastinal nodes, bevacizumab	Approach: Split Sample*	None
[29]	Inoperable Stage III/IV	2013	2002-2008	Retrospective Cohort	258	NR	NR	NR	OS	III: 26.7 IV: 73.3	Stage, C-reactive protein, N/L, Lactate dehydrogenase, albumin	Approach: Cross-Validation Discrimination: 0.665	None

Tool Citation	Population	Year of Publication	Dates of Data Collection	Study Design	Sample Size	Sample Size/Power Calculation Performed	Events	Duration of Follow-Up	Outcome	Stage (%)	Final Variables in Model	Internal Validation	External Validation
[38]	Stage IV, more than two metastatic sites	2014	2008-2013	Retrospective Cohort	46	NR	NR	NR	NOS	IV: 100	Serum total bilirubin, direct bilirubin, creatinine kinase, neuron-specific enolase, lactate dehydrogenase, CA153, CA125, CA199	Approach: Apparent*	None
[38]	Stage IV, two metastatic sites	2014	2008-2013	Retrospective Cohort	55	NR	NR	NR	NOS	IV: 100	Creatinine kinase, total triglyceride, CA153	Approach: Apparent*	None
[38]	Stage IV, one metastatic site	2014	2008-2013	Retrospective Cohort	73	NR	NR	NR	NOS	IV: 100	Direct bilirubin, age, neuron-specific enolase, CA199	Approach: Apparent*	None
[45]	Stage IIIb/IV	2014	2000-2010	Retrospective Cohort	462	NR	391	Median 44 months	OS	IIIB: 45 IV: 55	WBC, Lactate dehydrogenase, Alkaline phosphatase, Calcium, Albumin	Approach: Apparent*	None
[49]	Stage IIIb/IV	2014	1998-2011	Retrospective Cohort	1161	NR	NR	Median 6.6 months	OS	IIIB: 18.4 IV: 81.6	Age, alcohol consumption, stage, chemotherapy, Surgery, Albumin, International normalized ratio, Protein, blood urea nitrogen, Alkaline phosphatase	Approach: Split Sample Discrimination 0.83	None

NR= Not Reported RCT-PC= secondary use of randomized controlled trial data; NOS= survival not otherwise specified; OS= overall survival; CS= cumulative survival; DSS: disease-specific survival; ECOG= Eastern Cooperative Oncology Group, BMI: body mass index; EGFR- epidermal growth factor receptor; ^= concordance index based on Harrell's C statistic for models using time to event data; Y= calibration was performed

*Kaplan-Meier survival curves presented

Table 2

A description of quality criteria of prognostic tools targeting prognosis for patients with lung cancer (n=32)

Quality Criteria	N (%)
Prognostic Factor Selection Method	
Literature-based/clinical reasoning	8 (25)
Screened using univariate analysis	4 (13)
Available in existing dataset	2 (6)
Not reported	18 (56)
Missing Data Methods	
Complete case analysis	7 (22)
Imputation	3 (9)
Missing value indicator/unknown category	1 (3)
Input favourable value for missing variables	1 (3)
Not reported	20(63)
Handling of Continuous Predictors	
Continuous	5 (16)
Dichotomized/categorized	19 (59)
Not Reported	8 (25)
Analytic Model Used	
Cox proportional hazards regression	26 (81)
Recursive partitioning	2 (6)
Support vector machines	2 (6)
Regression Tree	1 (3)
Method not specified	1 (3)
Statistical Model Assumptions Checked	5 (16)

Table 3

Predictors included in only one tool in predicting survival for patients with incurable/metastatic NSCLC (n=9 tools) and SCLC (n=7 tools).

NSCLC	SCLC
# Metastatic Sites	Bicarbonate
Adrenal metastases	White blood cell count
Alcohol consumption	Platelet count
Bevacizumab	Age
Bone metastasis	CRFYA 21-1
brain metastases	Neutrophils
BUN	Liver Scan Results
CA125	Alanine transaminase
Calcium	Feinstein Symptom Score
Chemotherapy	
CRP	
Ethnicity	
Feinstein Symptom Score	
INR	
Intra-abdominal Metastasis	
Liver Metastasis	
Loss of Appetite	
Mediastinal nodes	
N/L ratio	
Number of prior regimens	
Objective response to first-line	
Progression-free interval during previous chemotherapy	
Protein	
Response to prior treatment	
Skin metastasis	
Subcutaneous metastasis	
Surgery	
Time from diagnosis to geftinib therapy	
Time from first-line chemotherapy	
Total Serum bilirubin	
Total triglycerides	
Tumor size	
Type of first-line therapy	

Table 4

Details of tool performance evaluations

Performance Measure	Internal Validation (n= 28)	External Validation (n= 11)
Internal Validation Method		
Apparent	18 (64)	--
Cross-Validation	3 (11)	--
Split Sample	4 (14)	--
Bootstrapping	3 (11)	--
External Validation Method		
Independent^	--	5 (45)
Overall Model Performance**		
Brier Score	1 (4)	1 (9)
R-square	3 (11)	4 (36)
Calibration		
Graph (Plot/intercept/slope)	2 (7)	1 (9)
Other	1 (4)	1 (9)
Discrimination		
C-statistic*	9 (32)	2 (18)
Other	1 (4)	3 (27)
Survival Analysis Only with Significance Test	16 (57)	3 (27)

*Concordance index based on Harrell's C statistic for models using time to event data

**Brier Score and R^2 could have been calculated on the same model

^Other approaches to external validation such as geographic validation or temporal validation could have been used, but were not used for any of the tools reviewed (see Box 1 for further details)