Refining Prognosis in Lung Cancer

A Report on the Quality and Relevance of Clinical Prognostic Tools

Alyson L. Mahar, MSc, Carolyn Compton, MD, PhD,†‡ Lisa M. McShane, PhD,§ Susan Halabi, PhD,‖ Hisao Asamura, MD,¶ Ramon Rami-Porta, MD, FETCS,# and Patti A. Groome, PhD,* On behalf of the Molecular Modellers Working Group of the American Joint Committee on Cancer*

Introduction: Accurate, individualized prognostication for lung cancer patients requires the integration of standard patient and pathologic factors, biological, 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 from January 1, 1996 to January 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 11 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.

*Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's University, ON; Canada; †Arizona State University, Phoenix, Arizona; ‡Laboratory Medicine and Pathology, Mayo Clinic School of Medicine, Rochester, Minnesota; §Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland; ‖Department of Biostatistics and Bioinformatics, Duke University and Alliance Statistics and Data Center, Durham, North Carolina; ¶Division of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan; and #Department of Thoracic Surgery, Hospital Universitari Mutua Terrassa, Barcelona, Spain.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: Patti A. Groome, PhD, Division of Cancer Care and Epidemiology, Cancer Research Institute, Queen's Cancer Research Institute, Level 2, 10 Stuart Street, Kingston, ON K7L 3N6, Canada. E-mail: groomep@queensu.ca

DOI: 10.1097/JTO.0000000000000652

ISSN: 1556-0864/15/1011-1576

Key Words: Lung cancer, Prognosis, Clinical prediction tools, Prediction models, Prognostic model.

(*J Thorac Oncol.* 2015;10: 1576–1589)

Anatomical 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, it is not sufficient for accurate survival probability prediction.[1,2,4–6] A wide variety of other prognostic information exists, including biological, 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 as 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 article reports on the MMWGs' findings in lung cancer.

MATERIALS AND 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 and Search of the Web-Based Scientific Community

The search for prognostic tools and information on their development and validation was performed through 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 article include those developed to estimate the probability of survival at a particular point along the disease trajectory (e.g., at diagnosis, after 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 from January 1, 1996 to January 27, 2015. Medical Subject Headings (MeSH) do not exist for prognostic tools, and so a combination of alternate MESH headings and key words were used after 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 that 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 endpoints included all time to death analyses (e.g., overall survival and cause-specific survival), and vital status analyses (e.g., probability of being dead 5 years after diagnosis). The search strategy was not developed to identify studies developing genomic classifiers built entirely on gene expression data. These studies were excluded.

```
1. models, statistical/
2. exp prognosis/
3. 1 and 2
4. predict* model*.mp.
5. exp nomogram/
6. exp "Neural Networks (Computer)"/
7. prognos* model*.mp.
8. predict* tool.mp.
9. Lung Neoplasms/
10. 9 and (3 or 4 or 5 or 6 or 7 or 8)
11. limit 10 to english language
12. limit 11 to yr="1996-2011"
```

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

Citations were assessed for inclusion by a single reviewer (A.M.), first through their titles and abstracts and then as full articles. Early on, a random sample of 20 citations was independently reevaluated by a blinded second reviewer (P.G.), and the results were compared. Percent agreement was calculated to estimate interrater reliability. Percent agreement was high (>95%), and any differences identified in this exercise were discussed and resolved through consensus. On the basics of 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 on-line 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 e-mail 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 Supplementary

