Editorial

Statistical modeling to unravel multiple predictors of the choice of chemotherapy for non-small cell lung cancer

Zhongheng Zhang

Department of Emergency Medicine, Sir Run-Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, China
Correspondence to: Zhongheng Zhang, MMed. No. 3, East Qinchun Road, Hangzhou 310016, China. Email: zh_zhang1984@hotmail.com.

Submitted Aug 20, 2016. Accepted for publication Sep 18, 2016.
doi: 10.21037/jtd.2016.09.57
View this article at: http://dx.doi.org/10.21037/jtd.2016.09.57

Non-small cell lung cancer (NSCLC) is one of the leading causes of death worldwide (1). In north America, it is estimated that the standardized incidence rate (SIR) and the standardized mortality rate (SMR) of lung cancer equal to 23.1 and 19.7 (in 100,000 people), respectively (2). There were 1.82 and 1.59 million new lung cancer cases and deaths worldwide, respectively. Another study showed that there were 1.82 and 1.59 million new lung cancer cases and deaths worldwide in the year 2012, respectively (3). Given that lung cancer has become a global and public health problem, strenuous effort has been made to combat this potentially life threatening disease (4-6). There are a variety of guidelines being published to provide state-of-the-art evidence for the management of lung cancer. However, a substantial number of patients with lung cancer are in advanced stage at the initial diagnosis, leading to a short median survival time. Chemotherapy is an option for the advanced lung cancer, and studies have shown prolonged survival time with chemotherapy. However, many patients choose not to adopt chemotherapy due to a variety of reasons, including social factors, patients' expectations and economic status. In the study by Lazarus and colleagues, they reported that "Chemotherapy utilization for advanced NSCLC is increasing over time. Chemotherapy administration is associated with socioeconomic status, performance status, and access to care, relationships that likely reflect evolving clinical practice patterns" (7). However, statistical methods used in the study were simple, failing to reveal complex relationships between these interesting factors. They only included univariate analysis, which was not enough in retrospective observational studies. Except for the situation of well performed randomized controlled trial that confounding factors can be well controlled by randomization, observational studies usually requires multivariable analysis to exclude confounding effects of measured covariates. Utilization of regression models is common in clinical cancer researches (8-10).

Structural equation modeling is a method to explore complex relationships between many factors, allowing for including latent variables (11). A latent variable is also called a construct that captures common features of a group of variables. For example, inflammatory response in itself does not exist but it is measured by a variety of biomarkers such as C-reactive protein, procalcitonin and interleukins. These biomarkers can be summarized by an inflammatory variable named "inflammation". The structural equation model is more commonly used in sociology, education and psychology. Clinical investigators and practitioners may not be familiar with this method. Furthermore, it allows exploration of causal associations between factors. In this example, the use of chemotherapy could be employed as the outcome variable. The aim is to explore factors influencing the choice of chemotherapy. In univariate analysis, it appeared that the place of diagnosis was associated with receipt of chemotherapy. However, there must be one or more mediating factors explaining this linkage. For instance, the Eastern Cooperative Oncology Group (ECOG) status was significantly different between the two places. Thus, the mediation effect of ECOG status may be interesting, and the proportion of mediated effect could be quantified.

Alternatively, multivariable regression analysis could be employed to explore independent risk factors for the choice of chemotherapy. For the survival analysis, patients who received chemotherapy had a median survival of 9.8 months, compared to 1.9 months for those that did not receive chemotherapy (P<0.001). It was still unknown based

on current analysis whether this was due to the effect of chemotherapy or due to confounding factors (e.g., there is imbalance between groups with and without chemotherapy). In the meantime, the ECOG score was also associated with survival. Thus, it is best to regress survival time on ECOG score and chemotherapy, with Cox proportional hazard model or other parametric survival models. With these advanced methods, interactions between social status and medical conditions can also be explored. Probably, ECOG is the most important determinant of the survival time (12), and chemotherapy may some effect but not so large as the ECOG score. It is also possible that the effect of chemotherapy can be different depending on stages of lung cancer, ECOG score and EGFR mutations. This is called interaction effect in statistical modeling. An interaction means the effect of one factor on dependent variable is not constant and is dependent on another covariate. This phenomenon is universal in clinical researches. Nonetheless, investigation of these complex relationships requires a large number of degrees of freedom. In other words, the sample size is also should be large. Otherwise, the problem of overfitting may arise, leading to limited external validation of the regression model (13).

Acknowledgements

None.

Footnote

Conflict of Interest: The author has no conflicts of interest to declare.

References

1. Mao Y, Yang D, He J, et al. Epidemiology of Lung Cancer. Surg Oncol Clin N Am 2016;25:439-45.
2. Rafiemanesh H, Mehtarpour M, Khani F, et al. Epidemiology, incidence and mortality of lung cancer and their relationship with the development index in the world. J Thorac Dis 2016;8:1094-102.
3. Cheng TY, Cramb SM, Baade PD, et al. The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics. J Thorac Oncol 2016;11:1653-71.
4. Davis SW, Rahn DA III, Sandhu AP. Stereotactic body radiation therapy (SBRT) for non-small cell lung cancer (NSCLC): current concepts and future directions. Transl Cancer Res 2014;3:303-12.
5. Paravati AJ, Johnstone DW, Seltzer MA, et al. Negative predictive value (NPV) of FDG PET-CT for nodal disease in clinically node-negative early stage lung cancer (AJCC 7th ed T1-2aN0) and identification of risk factors for occult nodal (pN1-N2) metastasis: implications for SBRT. Transl Cancer Res 2014;3:313-9.
6. Gillespie EF, Atwood TF, Sandhu AP. Lung stereotactic body radiotherapy (SBRT): a single institution's outcomes and methodology in the context of a literature review. Transl Cancer Res 2015;4:372-80.
7. Lazarus MA, Schachter L, Xavier M. Social factors, treatment, and survival in patients with advanced-stage non-small cell lung cancer. Transl Cancer Res 2014;3:146-51.
8. Lin YC, Huang YL, Lee H. Lysophosphatidic acid in prostate cancer progression. Transl Cancer Res 2015;4:527-36.
9. Wang LB, Chuang EY, Lu TP. Identification of predictive biomarkers for ZD-6474 in lung cancer. Transl Cancer Res 2015;4:324-31.
10. Lutz RJ. Targeting the folate receptor for the treatment of ovarian cancer. Transl Cancer Res 2015;4:118-26.
11. Zhang Z, Wáng YX. English language usage pattern in China mainland doctors: AME survey-001 initial analysis results. Quant Imaging Med Surg 2015;5:174-81.
12. Bernard ME, Clump DA, LaLonde R, et al. Radiation therapy for locally advanced lung cancer. Transl Cancer Res 2015;4:356-71.
13. Zhang Z. Too much covariates in a multivariable model may cause the problem of overfitting. J Thorac Dis 2014;6:E196-7.

Cite this article as: Zhang Z. Statistical modeling to unravel multiple predictors of the choice of chemotherapy for non-small cell lung cancer. J Thorac Dis 2016;8(10):2672-2673. doi: 10.21037/jtd.2016.09.57

