JOURNAL OF CLINICAL ONCOLOGY

EDITORIAL

Is Bootstrapping Sufficient for Validating a Risk Model for Selection of Participants for a Lung Cancer Screening Program?

Michael W. Marcus and John K. Field, *The University of Liverpool, Liverpool, United Kingdom*

Risk prediction models are powerful tools that use multivariable regression to combine predictors or predisposing factors to estimate the probability or risk of the presence or future occurrence of clinical outcomes such as lung cancer.[1-3] Several lung cancer risk prediction models have been developed.[4-12] Such models are usually constructed in data sets with information from a well-defined population with similar characteristics.[13] Discrimination is a measure of how well a model can separate diseased from nondiseased individuals and is most often measured using the area under the receiver operating characteristic curve or concordance c-statistic,[14] although other methods and metrics of performance of prediction models have been published.[15]

The discriminative performance of a risk model depends not only on the identification of individual risk factors, but also on how these risk variables interact with other variables, how accurately these factors can be measured, and the appropriateness of the population and statistical techniques used for modeling.[16] Generally, the discriminative performance of risk models in the initial development data set is better than the performance in other data sets.[1,13,17] This self-fulfilling prophecy, otherwise known as optimism, is a statistical phenomenon that is well described in the literature.[1,2,18]

Many risk prediction models fail in clinical settings because of overfitting, ie, spurious association as the result of noise in the data set, which may lead to overestimation or underestimation of predictive performance.[19,20] A valuable risk model will not only predict outcome in the initial development data but also show good discriminative performance in independent data sets.[2] Poor prediction performance of risk models can be prevented by conducting an unbiased internal validation.[21] The principal methods for conducting internal validation of risk models are data-splitting, cross-validation, and bootstrapping.[1] Of the three methods described for internal validation, bootstrapping has been applauded as the most efficient method.[1,13] This resampling technique invokes an iteration process by drawing samples with replacement from the original sample, and the original data set is used for model development.[1,13]

One of the major issues hampering effective lung cancer treatment is the presentation of patients at advanced stages of the disease, when current therapeutic regimens have poor outcomes.[22] The way forward for improved management and prognosis for individuals at high risk of lung cancer is early detection of the disease, which may be achieved through low-dose computed tomography screening.[23] The high percentage of false positives reported in various lung cancer screening trials may be attributed to variability in age and smoking history, which are solely used as eligibility criteria. Concerns about the high percentage of false positives and associated health hazards as the result of radiation exposure have awakened interest in the application of quantitative and more informative risk prediction models in identifying individuals at high risk.[24]

Different validated risk models using different risk criteria have been proposed; the Liverpool Lung Project risk model was used to select individuals with a ≥ 5% risk of developing lung cancer in a 5-year period[7,25]; the Prostate, Lung, Colorectal and Ovarian (PLCO) $PLCO_{m2012}$ risk ≥ 1.51% of lung cancer death over 6 years[26]; and the recently proposed model on the basis of the use of quintile of the risk of lung cancer death at 5 years, and more recently, risk modeling on the basis of US Preventive Services Task Force lung cancer screening recommendations.[27,28] Li et al[24] externally validated four lung cancer risk prediction models (Bach, Spitz, Liverpool Lung Project, and $PLCO_{m2012}$) among 20,700 ever smokers in the German European Prospective Investigation of Cancer and Nutrition (EPIC) study and concluded that all models apart from the Spitz model have a similar accuracy to identify individuals at high risk for screening and outperform age and smoking eligibility criteria used in screening trials.

In the article accompanying this editorial, Muller et al[29] reported a lung cancer risk prediction model incorporating lung function in the UK Biobank prospective cohort study. In their study, they used flexible parametric survival models to estimate the 2-year probability of lung cancer, accounting for the competitive risk of death in 502,321 participants. During accumulated follow-up of 1,469,518 person years, 738 developed lung cancer. Their model incorporating all predictors had excellent discrimination. The c-statistic of their risk model and the bias-corrected bootstrap resampling were similar: 0.85 (95% CI, 0.82 to 0.87) and 0.84 (95% CI, 0.82 to 0.86), respectively. In addition, the full model had better discrimination than standard lung cancer screening eligibility criteria c-statistics: 0.66 (95% CI, 0.64 to 0.69). Internal validation suggested that the model will perform well in discriminating between patients with lung cancer and population control subjects. A model with such high discrimination could improve eligibility criteria for lung cancer screening programs after validation in external data sets.

Internal validation of studies by bootstrapping is a well-established statistical technique that does not examine the generalizability of risk models. As mentioned above, bootstrapping uses the original entire data set to estimate predictive performance of a model. This method, albeit, is a self-fulfilling prophecy that is often optimistic. Steyerberg et al[30] argued that internal validation is statistically inefficient and methodologically weak because no difference in time or place exist other than by chance. Many published prediction models have never been validated as the result of uncollected predictor variables in otherwise suitable validation cohorts.[31] However, several imputation techniques have been developed to overcome this barrier. Clinical application of a model to predict an individual's risk of disease (ie, lung cancer) is dependent on its successful validation in independent populations. Although the model developed by Muller et al has good discrimination, this model must pass the litmus test of external validation (such as the aforementioned lung cancer risk prediction models) before its clinical utility can be considered for lung cancer screening programs.

Lung cancer risk prediction models will continue to play an important role in this era of personalized medicine, particularly in the selection of individuals for prevention and surveillance interventions. Risk estimates from a predictive model may help identify and counsel individuals at elevated risk of lung cancer, raising awareness that can lead to risk-minimizing behaviors.[32] Alternatively, the model's prediction may be useful in defining a high-risk population to include in prevention trials or to target for screening and prevention efforts. In all of these contexts, accurate, internally and externally validated risk models will be most useful in making clinical decisions regarding patient stratification for prevention and early detection interventions of lung cancer.[33]

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at jco.org.

AUTHOR CONTRIBUTIONS

Financial support: Michael W. Marcus
Administrative support: Michael W. Marcus
Provision of study materials or patients: Michael W. Marcus
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. 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 15:361-387, 1996

2. Altman DG, Vergouwe Y, Royston P, et al: Prognosis and prognostic research: Validating a prognostic model. BMJ 338:b605, 2009

3. Pavlou M, Ambler G, Seaman SR, et al: How to develop a more accurate risk prediction model when there are few events. BMJ 351:h3868, 2015

4. Bach PB, Kattan MW, Thornquist MD, et al: Variations in lung cancer risk among smokers. J Natl Cancer Inst 95:470-478, 2003

5. Spitz MR, Etzel CJ, Dong Q, et al: An expanded risk prediction model for lung cancer. Cancer Prev Res (Phila) 1:250-254, 2008

6. Spitz MR, Hong WK, Amos CI, et al: A risk model for prediction of lung cancer. J Natl Cancer Inst 99:715-726, 2007

7. Cassidy A, Myles JP, van Tongeren M, et al: The LLP risk model: An individual risk prediction model for lung cancer. Br J Cancer 98:270-276, 2008

8. Etzel CJ, Kachroo S, Liu M, et al: Development and validation of a lung cancer risk prediction model for African-Americans. Cancer Prev Res (Phila) 1:255-265, 2008

9. Marcus MW, Chen Y, Raji OY, et al: LLPi: Liverpool Lung Project risk prediction model for lung cancer incidence. Cancer Prev Res (Phila) 8:570-575, 2015

10. Hoggart C, Brennan P, Tjonneland A, et al: A risk model for lung cancer incidence. Cancer Prev Res (Phila) 5:834-846, 2012

11. Park S, Nam BH, Yang HR, et al: Individualized risk prediction model for lung cancer in Korean men. PLoS One 8:e54823, 2013

12. Tammemagi CM, Pinsky PF, Caporaso NE, et al: Lung cancer risk prediction: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial models and validation. J Natl Cancer Inst 103:1058-1068, 2011

13. Steyerberg EW, Harrell FE Jr, Borsboom GJ, et al: Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 54:774-781, 2001

14. Cook NR: Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 115:928-935, 2007

15. Steyerberg EW, Vickers AJ, Cook NR, et al: Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 21:128-138, 2010

16. Field JK, Marcus MW: Risk prediction models in lung cancer: The methodology for identifying high risk individuals for future lung cancer computed tomography (CT) screening programs, in AACR Education Book. Philadelphia, PA, American Association for Cancer Research, 2015.

17. Van Houwelingen JC, Le Cessie S: Predictive value of statistical models. Stat Med 9:1303-1325, 1990

18. Bleeker SE, Moll HA, Steyerberg EW, et al: External validation is necessary in prediction research: A clinical example. J Clin Epidemiol 56:826-832, 2003

19. Altman DG, Royston P: What do we mean by validating a prognostic model? Stat Med 19:453-473, 2000

20. Castaldi PJ, Dahabreh IJ, Ioannidis JP: An empirical assessment of validation practices for molecular classifiers. Brief Bioinform 12:189-202, 2011

21. Collins GS, Ogundimu EO, Altman DG: Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. Stat Med 35:214-226, 2016

22. Dizon DS, Krilov L, Cohen E, et al: Clinical Cancer Advances 2016: Annual Report on Progress Against Cancer From the American Society of Clinical Oncology. J Clin Oncol 34:987-1011, 2016

23. Marcus MW, Raji OY, Field JK: Lung cancer screening: Identifying the high risk cohort. J Thorac Dis 7(Suppl 2):S156-S162, 2015

24. Li K, Hüsing A, Sookthai D, et al: Selecting high-risk individuals for lung cancer screening: A prospective evaluation of existing risk models and eligibility criteria in the German EPIC cohort. Cancer Prev Res (Phila) 8:777-785, 2015

25. Raji OY, Duffy SW, Agbaje OF, et al: Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: A case-control and cohort validation study. Ann Intern Med 157:242-250, 2012

26. Tammemägi MC, Church TR, Hocking WG, et al: Evaluation of the lung cancer risks at which to screen ever- and never-smokers: Screening rules applied to the PLCO and NLST cohorts. PLoS Med 11:e1001764, 2014

27. Kovalchik SA, Tammemagi M, Berg CD, et al: Targeting of low-dose CT screening according to the risk of lung-cancer death. N Engl J Med 369:245-254, 2013

28. Katki HA, Kovalchik SA, Berg CD, et al: Development and validation of risk models to select ever-smokers for CT lung cancer screening. JAMA 315:2300-2311, 2016

29. Muller DC, Johansson M, Brennan P: A lung cancer risk prediction model incorporating lung function: Development and validation in the UK Biobank prospective cohort study. J Clin Oncol 35:861-869, 2017

30. Steyerberg EW, Moons KG, van der Windt DA, et al: Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. PLoS Med 10:e1001381, 2013

31. Held U, Kessels A, Garcia Aymerich J, et al: Methods for handling missing variables in risk prediction models. Am J Epidemiol 184:545-551, 2016

32. Sherratt FC, Marcus MW, Robinson J, et al: Utilizing lung cancer risk prediction models to promote smoking cessation: Two randomized controlled trials. Am J Health Promot 0890117116673820, 2016

33. Field JK, Hansell DM, Duffy SW, et al: CT screening for lung cancer: Countdown to implementation. Lancet Oncol 14:e591-e600, 2013

© 2017 by American Society of Clinical Oncology **819**

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Is Bootstrapping Sufficient for Validating a Risk Model for Selection of Participants for a Lung Cancer Screening Program?

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc *or* ascopubs.org/jco/site/ifc.

Michael W. Marcus
No relationship to disclose

John K. Field
No relationship to disclose