Published in final edited form as:

Lung Cancer. 2015 July ; 89(1): 31–37. doi:10.1016/j.lungcan.2015.03.021.

A simple model for predicting lung cancer occurrence in a lung cancer screening program: The Pittsburgh Predictor

David O. Wilson^{1,*} and Joel Weissfeld^{2,*}

¹Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine

²Department of Epidemiology, University of Pittsburgh

Abstract

Background—A user-friendly method for assessing lung cancer risk may help standardize selection of current and former smokers for screening. We evaluated a simple 4-factor model, the Pittsburgh Predictor, against two well-known, but more complicated models for predicting lung cancer risk.

Methods—Trained against outcomes observed in the National Lung Screening Trial (NLST), the Pittsburgh Predictor used four risk factors, duration of smoking, smoking status, smoking intensity, and age, to predict 6-year lung cancer incidence. After calibrating the Bach and PLCO_{M2012} models to outcomes observed in the low-dose computed tomography arm of the NLST, we compared model calibration, discrimination, and clinical usefulness (net benefit) in the NLST and Pittsburgh Lung Screening Study (PLuSS) populations.

Results—The Pittsburgh Predictor, Bach, and PLCO_{M2012} represented risk equally well, except for the tendency of PLCO_{M2012} to overestimate risk in subjects at highest risk. Relative to the Pittsburgh Predictor, Bach and PLCO_{M2012} increased the area under the receiver operator characteristic curve by 0.007 to 0.009 and 0.012 to 0.021 units, respectively, depending on study population. Across a clinically relevant span of 6-year lung cancer risk thresholds (0.01 to 0.05), Bach and PLCO_{M2012} increased net benefit by less than 0.1% in NLST and 0.3% in PLuSS.

Conclusion—In exchange for a small reduction in prediction accuracy, a simpler lung cancer risk prediction model may facilitate standardized procedures for advising and selecting patients with respect to lung cancer screening.

Keywords

lung cancer screening; LDCT; Pittsburgh predictor; PLuSS

© 2015 Published by Elsevier Ltd.

Correspondence and requests for reprints should be addressed to: David O. Wilson, M.D., M.P.H., UPMC-Cancer Pavilion, 5150 Centre Ave., Suite 4C, Pittsburgh, PA 15232., wilsondo@upmc.edu.

*Drs. Wilson and Weissfeld contributed equally to this manuscript.

Drs. Wilson and Weissfeld have no conflicts or financial disclosures.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Lung cancer remains the leading cause of cancer mortality in the U. S. [1]. Enrolling 55 to 74 year-old 30 pack-year current or former smokers (quit <15 years), the National Lung Screening Trial (NLST) recently reported 20% lower lung cancer mortality after three annual screenings with low-dose computed tomography (LDCT) vs. chest radiography [2]. Guided by the success of the NLST and a comparative modeling study which validated the selection criteria used by NLST [3], the United States Preventive Services Task Force (USPSTF) endorsed LDCT in 55 to 80 year-old 30 pack-year smokers, with annual screening to stop after a 15-year smoke-free interval [4].

As a matter of policy, the USPSTF guideline may be used to define eligibility for LDCT. However, several lung cancer risk prediction models have been published [4–11]. Unlike the straightforward USPSTF guideline, these models place smokers on a continuum of risk, determined by a panel of risk factors. Some models define complex non-linear relationships between the smoking-related variables and lung cancer risk [5,11]. According to some analysts, the use of a model to select smokers for screening may improve the efficiency of LDCT, including the number of persons needed to screen to detect a fixed number with lung cancer [11].

Against this backdrop, we conceived a simple 4-factor lung cancer risk prediction model, less complicated than currently available models, perhaps more amenable to widespread application, yet devised to place individuals on a lung-cancer risk continuum. It is important to point out that we derived and validated our model in populations preselected to be high risk for lung cancer, that are consistent with the USPSTF recommendations for screening, but not representative of all smokers. We systematically evaluated our model, the Pittsburgh Predictor, against two other models, one validated by Bach *et al.* [5] and a second by investigators from the Prostate, Lung Colorectal, and Ovarian (PLCO) Cancer Screening Trial [11].

METHODS

Data sources

We used data from the National Lung Screening Trial (NLST) [2,12,13] (Cancer Data Access System, September 9, 2012¹) and the Pittsburgh Lung Screening Study (PLuSS) [14]. Between August 2002 and September 2007, NLST enrolled 53,454 55 to 74 year-old current and former (quit <15 years) cigarette smokers (30 pack years) and randomized 26,722 and 26,732 to three annual screenings with low-dose computed tomography (LDCT) or chest radiography (CXR), respectively. Between January 2002 and April 2005, the Pittsburgh Lung Screening Study (PLuSS) recruited 3654 50–79 year-old current and former (quit <10 years) cigarette smokers (>10 cigarettes per day for 25 years) and completed baseline LDCT in 3642 and repeat LDCT one year later in 3423, with selected subjects receiving additional protocol directed LDCT in later years.

¹<https://biometry.nci.nih.gov/cdas/studies/nlst/>

Questionnaire data

With the following exceptions, NLST and PLuSS collected comparable risk factor information from self-completed questionnaires administered at the beginning of follow-up. PLuSS and the non-ACRIN portion of NLST determined duration of smoking as the difference between current age (active smoker) or age quit (former smoker) and age smoking started, whereas the ACRIN portion of NLST recorded the number of years total a subject had smoked. NLST recorded smoking intensity as integer cigarettes per day, whereas PLuSS recorded smoking intensity in five discrete categories. Questionnaires used in NLST, but not PLuSS, included two items about asbestos. All data analyses represented age as age at study enrollment, history of chronic obstructive pulmonary disease (COPD) as history of a doctor diagnosis of emphysema, family history of lung cancer and personal cancer history.²

Predictive models

We considered two published lung cancer prediction models [5,11], the Bach and PLCO_{M2012} (Table 1). Bach [5] modeled data for 14,254 smokers and 3918 asbestos-exposed persons who participated in the Carotene and Retinol Efficacy Trial (CARET). Between June 1985 and September 1994, CARET enrolled 50–69 year-old current or former cigarette smokers. In addition to cigarette smoking (current or quit <15 years, 30 pack years), eligibility by virtue of asbestos exposure required radiologic evidence or relevant job history. With factors for sex, age, asbestos, and cigarette smoking (years smoking, years quit, and cigarettes per day), Bach predicts cumulative lung cancer risk from two proportional hazards regression equations, one equation for lung cancer risk and a second equation for mortality from causes other than lung cancer. Based on 1070 lung cancers in CARET over a mean 9.3 year follow-up, the Bach equations predict risks in 55 year-old persons followed to age 85 years.

PLCO_{M2012} [11] modeled data for the 39,928 current or former smokers (709 lung cancers over six years of follow-up) in the control arm of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Between November 1993 and July 2001, PLCO enrolled 55–74 year-old persons without a history of prostate, lung, colorectal, or ovarian cancer. With factors for age, cigarette smoking (years smoking, smoking status, years quit, and cigarettes per day), race or ethnic group, education, personal history of cancer, family history of lung cancer, history of chronic obstructive pulmonary disease, and BMI, PLCO_{M2012} uses a logistic equation to predict 6-year lung cancer risk.

Our simple Pittsburgh Predictor model included four risk factors, duration of smoking (four categories, <30, 30–39, 40–49, and 50 years), smoking status (two categories, currently smoking or quit <1 year and quit 1 year), smoking intensity (four categories, <20, 20–29, 30–39, and 40 cigarettes per day) and age. To moderate its correlation with duration of smoking, younger and older age groups were determined in relation to the smoking duration-specific median ages observed in NLST (Table 1). Model training excluded 793 (3.0%) and 1082 (4.0%) participants in NLST LDCT and CXR, respectively. Reasons for

²<https://biometry.nci.nih.gov/shared/cdas-static/forms/nlst/lss/MHQ.pdf>

exclusion included absent baseline LDCT or CXR and post randomization evidence indicating a possible violation of NLST entry criteria [12]. Retaining 25,929 LDCT and 25,648 CXR NLST-eligible and screened participants, we used our four risk factors and lung cancer occurrence over six years to fit two logistic regression models. One model used data from NLST LDCT (1000 lung cancers; area under the receiver operator characteristic curve – AuROC, 0.679) and the second model used data from NLST CXR (854 lung cancers; AuROC, 0.687). Obtaining roughly equal parameter estimates with LDCT and CXR data, we averaged the risk factor-specific parameter estimates from LDCT and CXR models, multiplied by 10, and rounded to the nearest integer (Table 2). Relative to a reference category, the resulting integer point value associated with each risk factor category represents 10 times the increase in the logit of the expected 6-year lung cancer risk. A summation of the integer point values associated with each of our four risk factors produces a total risk score with possible values ranging between -17 points (<57 year-old former smoker of <20 cigarettes/day for <30 years) and 23 points (≥68 year-old current smoker of ≥40 cigarettes/day for ≥50 years). A total score of 0 points captures the risk associated with the reference categories, <59 year-old current smoker of 20–29 cigarettes/day for 30–39 years (Figure 1).

As a measure of clinical usefulness, we calculated net benefit (for details see the Supplement). Using net benefit, our analysis aimed, in part, to develop a quantitative understanding of the relative clinical usefulness of our simple model, the Pittsburgh Predictor.

Data analysis

NLST LDCT and NLST CXR data were used to evaluate model calibration (Bach and PLCO_{M2012}) and discrimination (Pittsburgh Predictor, Bach, and PLCO_{M2012}). Approximately 0.2% and 4.7% of Bach and PLCO_{M2012} predictions in NLST replaced missing or extreme risk factor data with imputed values (data not shown). After recalibrating Bach and PLCO_{M2012} against NLST LDCT, we evaluated model calibration and discrimination in P_{LuSS}, restricted to the 3642 subjects with baseline LDCT. Lacking data about asbestos, the application of Bach to P_{LuSS} assumed an 8.1% and 0.8% frequency of asbestos exposure in men and women, respectively, the frequency of asbestos work or asbestosis reported by men and women in NLST. Limited to smoking intensity data in discrete categories, the application of Bach and PLCO_{M2012} to P_{LuSS} assigned values of 10.0, 15.8, 20.6, 30.4, and 45.3 cigarettes per day to smokers of <10, 10–19, 20–29, 30–39, and ≥40 cigarettes per day, the average cigarettes per day reported by NLST participants in corresponding categories.

To evaluate calibration, we estimated the intercept term in logistic regressions with the logit of model predictions entered as an offset, examined discretized and local regression (LOESS) plots of observed vs. expected lung cancer risk, and calculated the Hosmer–Lemeshow χ^2 test statistic [15]. Comparisons of the discrimination achieved by different models used integrated plots of predictiveness and classification performance and estimates of the AuROC, net reclassification improvement (NRI), continuous NRI (cNRI), and integrated discrimination improvement (IDI) [16]. Finally, measures of net benefit were

used to compare clinical usefulness [17]. Data analyses used SAS for Windows (version 9.3).

RESULTS

Description of the study cohorts

PLuSS had 3642 subjects, 51% men, median age 58 years (interquartile range, IQR, 54–64 years), and 60% current smokers with median 39 years duration (IQR 34–45 years) and 9% 2 packs per day. NLST had 51,577 subjects (25,929 LDCT and 25,648 CXR), 59% men, median age 60 years (IQR 57–65 years), and 49% current smokers with median duration 40 years (IQR 35–45 years) and 23% 2 packs per day. The mean numbers of protocol-based lung cancer screenings in PLuSS and NLST were 2.5 and 2.8 per subject, respectively. Six-year lung cancer risks in PLuSS, NLST LDCT, and NLST CXR were 3.9% (95% confidence interval, CI, 3.3–4.6%, 143 cases), 3.9% (95% CI 3.6–4.1%, 1000 cases), and 3.3% (95% CI 3.1–3.6%, 854 cases), respectively. Overall, 0.7%, 2.2%, and 1.6% of lung cancer-free survivors in PLuSS, NLST LDCT, and NLST CXR, respectively, were lost to follow-up before the fifth anniversary of study enrollment.

Model calibration

The published Bach and PLCO_{M2012} models predicted average lung cancer risk in NLST CXR after slight (<10%) recalibration (OR_{recal} 0.99 and 1.06, respectively; Figure 2) and in NLST LDCT after moderate (10%) recalibration (OR_{recal} 1.16 and 1.24, respectively; Figure 2). Apart from over estimation of risk by PLCO_{M2012} for NLST LDCT subjects in the highest risk category, the Pittsburgh Predictor, Bach, and PLCO_{M2012} represented risk well across levels of risk (Figure 2). Lung cancer occurred more frequently in PLuSS than predicted by the Pittsburgh Predictor (OR_{recal} 1.24, 95% CI 1.04–1.46), Bach (OR_{recal} 1.21, 95% CI 1.02–1.43), or PLCO_{M2012} (OR_{recal} 1.16, 95% CI 0.97–1.37), with expected lung cancer occurrence determined by models calibrated to NLST LDCT. After recalibration to PLuSS, the Pittsburgh Predictor, Bach, and PLCO_{M2012} represented risk well across levels of risk (Figure 3), except again for the tendency of PLCO_{M2012} to overestimate risk for subjects at higher risk.

Scatter plots showed a linear association in PLuSS between the risk predictions offered by the Pittsburgh Predictor and Bach, but curvilinear associations for comparisons that involved PLCO_{M2012}, with the PLCO_{M2012} producing greater estimates of risk for persons with 6-year risks >6% according to either the Pittsburgh Predictor or Bach (Figure 4).

Model discrimination

Depending on the cohort, models discriminated with AuROC values between 0.678 and 0.721 (Figures 2 and 3). Some improvements in AuROC were statistically significant (*e.g.*, PLCO_{M2012} relative to the Pittsburgh Predictor in NLST CXR, AuROC difference 0.015, *p*-value 0.016. Across a clinically relevant span of 6-year lung cancer risk thresholds (0.01 to 0.05), the Bach and PLCO_{M2012}, relative to the Pittsburgh Predictor, increased net benefit by less than 0.1% in NLST and by less than 0.3% in PLuSS (see Supplement for details).

Lung cancer risk calculator

For illustration, we programmed Excel to calculate the Pittsburgh Predictor (Figure 1). The calculator defines four risk levels (0, 1–8, 9–14, and 15–23 total points), corresponding to 1.4%, 2.6%, 5.4%, and 9.5% average 6-year lung cancer risks in NLST LDCT.

DISCUSSION

We describe a simple 4 factor model, the Pittsburgh Predictor, for assessing lung cancer risk in the high risk population potentially eligible for LDCT lung cancer screening. LDCT will likely enter routine clinical practice and potentially affect as many as nine million current or former U.S. smokers [18]. Appropriately, the imprecision of the information available regarding the benefits and harms of LDCT provokes discussion about risk thresholds for screening, with higher thresholds expected to maximize benefit in relation to harm. In a clinical context, the optimal threshold might vary from smoker to smoker in accordance with personal valuations of the many potential benefits and risks of screening.

In this context, we formulated and evaluated a simple lung cancer risk prediction model, suitable for use in populations similar to NLST and PLoSS, and most importantly, the population for which the USPSTF has advocated lung cancer screening. This model, the Pittsburgh Predictor, calculates 6-year lung cancer risk based on four factors, duration of smoking, smoking status, smoking intensity, and age (Table 2). With two previously published models (Bach [5] and PLCO_{M2012} [11]) as benchmarks, the Pittsburgh Predictor calibrated well (Figures 2–4), with qualifications noted in the following paragraph. As expected, the Pittsburgh Predictor discriminated less well overall, as expressed by reductions in the AuROC of 0.007 to 0.021 units (Figures 2–3), depending on study population and choice of model for comparison. The simplicity of the PP (4 factors vs. 6 factors for Bach vs. 12 factors for PLCO_{M2012}) explains the lower AUC. Yet, across a clinically relevant span of decision thresholds, a span that includes the average risks observed in NLST and PLoSS, this weakening in prediction accuracy corresponded to reductions of net benefit no greater than 0.1% in NLST and 0.3% in PLoSS (see Supplement for details).

The results described here reflect model performance in a setting of active lung cancer screening. As a consequence, the accuracy of the absolute predictions delivered by the Pittsburgh Predictor and the recalibrated Bach and PLCO_{M2012} models depends on comparability between NLST and a target population with respect to the screening, diagnostic, and follow-up experiences of the two populations. In this context, 6-year lung cancer risks in PLoSS were ~20% greater than predicted by NLST LDCT-calibrated models, as demonstrated by values for $OR_{\text{recal}} > 1.00$ (Figure 3; Pittsburgh Predictor 1.24, Bach 1.21, PLCO_{M2012} 1.16). Other explanations for the higher than expected lung cancer risk observed in PLoSS include 1) risk determinants (*e.g.*, indoor radon), uniquely prevalent in PLoSS, but not captured by any model and 2) random chance (Type I error). For two reasons, we anticipated that models calibrated to NLST LDCT might underpredict in PLoSS; 1) PLoSS, in its early days, pursued over aggressively a surgical diagnosis for LDCT-detected abnormalities [14] and 2) PLoSS, continued for some participants, as noted above, periodic LDCT after the initial screening rounds, factors expected to contribute to over-diagnosis and lead time still unresolved at the end of follow-up, respectively. Similarly,

differences in clinical practice might explain some of the variability in risk-adjusted lung cancer incidence observed center to center in NLST. As a final note, the amplified risks calculated by PLCO_{M2012} for PLuSS participants in the highest decile of risk (Figure 4) explain completely (data not shown) the superior calibration achieved overall by PLCO_{M2012} relative to Bach (OR_{recal} 1.16 vs. 1.21). This apparent superiority of PLCO_{M2012} could reflect more accurate estimation of risk among person placed at very high risk by virtue of risk factors unique to PLCO_{M2012} (e.g., family history of lung cancer or personal history of chronic obstructive pulmonary disease).

Using a predictor of lung cancer death, Kovalchik *et al.* [19] estimated that 99% of the lung cancer death prevented by LDCT could have been achieved by restricting LDCT screening to the 80% of NLST at highest risk. For comparison, we determined that 93% of 6-year lung cancer incidence in NLST LDCT occurred among persons who belonged to the top 80% of risk, as determined by the Pittsburgh Predictor (*i.e.*, >0 total score, >1.7% predicted risk). These examples illustrate the manner in which stratification according to risk might be used to limit LDCT to those persons most likely to benefit.

To demonstrate its simplicity, we programmed the Pittsburgh Predictor as a Microsoft Excel® application, available for download. As shown in figure 1, this application uses a 5-item questionnaire to ascertain the four factors that determine the Pittsburgh Predictor total score (Figure 1, table 2). The simplicity of the Pittsburgh Predictor would not preclude use of other less common factors predictive of lung cancer for decision making purposes (e.g., family history of lung cancer, quit time, occupational or environmental exposures, and personal history of emphysema), particularly for persons with calculated risks close to a decision threshold, but our model does not take these factors into account.

There are several limitations to our model. First, it is derived in preselected high risk populations (PLuSS and NLST) and not necessarily applicable to the general population of smokers. It is applicable to the population eligible for LDCT based on USPSTF recommendations. Second, the USPSTF criteria for screening include some low risk individuals and exclude some high risk individuals. Our model may help identify a relatively low risk subpopulation within a preselected high risk population. Third, the PP has a ceiling effect of about 12% risk, although for decision making purposes our focus is on lower risk individuals. Fourth, our model was derived and tested in the United States and applicability to other populations will need to be tested.

In conclusion, we present a detailed and transparent description of the development and validation of the Pittsburgh Predictor. While results from NLST established the benefits of LDCT, the utility in clinical practice of risk-based strategies for advising at-risk persons about LDCT, while intuitive, has not been proven. Nonetheless, simplicity, a motivating principle behind the Pittsburgh Predictor, may facilitate this process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by the University of Pittsburgh Cancer Institute's Specialized Program of Research Excellence (SPORE) in Lung Cancer (NCI P50-CA90440) and the Cancer Center Core Grant (NCI 2P30 CA047904)

The authors thank Dr. Brenda Diergaarde for her critical review of this manuscript.

Abbreviations

ACRIN	American College of Radiology Imaging Network
AuROC	area under the receiver operator characteristic curve
CDAS	Cancer Data Access System
CXR	chest radiography
LDCT	low-dose computed tomography
NLST	National Lung Screening Trial
NRI	net reclassification improvement
PLCO	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
PLuSS	Pittsburgh Lung Screening Study

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin*. 2013; 63:11–30. [PubMed: 23335087]
2. Aberle DR, Adams AM, et al. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365:395–409. [PubMed: 21714641]
3. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: A comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014; 160:311–20. [PubMed: 24379002]
4. Humphrey LL, Deffenbach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med*. 2013; 159:411–20. [PubMed: 23897166]
5. Bach PB, Kattan MW, Thornquist MD, et al. Variations in lung cancer risk among smokers. *J Natl Cancer Inst*. 2003; 95:470–8. [PubMed: 12644540]
6. Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer*. 2008; 98:270–6. [PubMed: 18087271]
7. Maisonneuve P, Bagnardi V, Bellomi M, et al. Lung cancer risk prediction to select smokers for screening CT--a model based on the Italian COSMOS trial. *Cancer Prev Res (Phila)*. 2011; 4:1778–89. [PubMed: 21813406]
8. Spitz MR, Etzel CJ, Dong Q, et al. An expanded risk prediction model for lung cancer. *Cancer Prev Res (Phila)*. 2008; 1:250–4. [PubMed: 19138968]
9. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst*. 2007; 99:715–26. [PubMed: 17470739]
10. 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*. 2011; 103:1058–68. [PubMed: 21606442]
11. Tammemagi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med*. 2013; 368:728–36. [PubMed: 23425165]

12. Aberle DR, Berg CD, et al. National Lung Screening Trial Research Team. The National Lung Screening Trial: overview and study design. *Radiology*. 2011; 258:243–53. [PubMed: 21045183]
13. Church TR, Black WC, et al. National Lung Screening Trial Research Team. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med*. 2013; 368:1980–91. [PubMed: 23697514]
14. Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Screening Study (PLuSS): outcomes within 3 years of a first computed tomography scan. *Am J Respir Crit Care Med*. 2008; 178:956–61. [PubMed: 18635890]
15. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med*. 1997; 16:965–80. [PubMed: 9160492]
16. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27:157–72. discussion 207–12. [PubMed: 17569110]
17. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006; 26:565–74. [PubMed: 17099194]
18. Ma J, Ward EM, Smith R, Jemal A. Annual number of lung cancer deaths potentially avertable by screening in the United States. *Cancer*. 2013; 119:1381–5. [PubMed: 23440730]
19. 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*. 2013; 369:245–54. [PubMed: 23863051]

Highlights

The Pittsburgh predictor is a simple 4 factor prediction model for lung cancer occurrence.

For 1 – 5 % 6 year lung cancer risk prediction, the Pittsburgh predictor performs well.

The Pittsburgh predictor may facilitate standardized procedures for advising and selecting patients with respect to lung cancer screening.

To anticipate your experience in a lung cancer screening program, respond to each of the following questions.

55	How old are you?
18	At what age did you become a regular cigarette smoker?
Currently smoking	Are you currently smoking cigarettes or have you quit?
	At what age did you last quit smoking cigarettes?
20-29	During the entire time you smoked, how many cigarettes did you smoke daily, on average?

Your Total Score is 0 (Risk Level 1). Refer to the Table to learn the experience of NLST participants who had a risk profile similar to yours. The Table shows lung cancer risk for each of four Risk Levels. The Table expresses risk as the percentage of NLST participants who received a lung cancer diagnosis in the six years after a low-dose computed tomography (LDCT) or conventional chest x-ray (CXR) screening. The Table shows the Range of Risk associated with the Scores covered by each Risk Level and the Average Risk observed in NLST.

Risk Level	Score	Range of Risk		Average Risk	
		LDCT	CXR	LDCT	CXR
1 (≈20%)	≤0	≤1.7	≤1.4	1.4	1.1
2 (≈40%)	1-8	1.9-3.7	1.6-3.2	2.6	2.1
3 (≈30%)	9-14	4.1-6.5	3.5-5.6	5.4	5.0
4 (≈10%)	15-23	≥7.1	≥6.8	9.5	8.0

Figure 1.
Screen shot from Pittsburgh Predictor

Figure 2. Pittsburgh Predictor, Bach, and PLCO_{M2012} calibration plots in NLST LDCT and NLST

CXR. The Pittsburgh Predictor calculates 6-year lung cancer risks as $\frac{1}{1+e^{-(\alpha+0.10 \cdot S)}}$, where $\alpha = -4.0641$ after LDCT or -4.2195 after CXR and S = point total calculated according to the method shown in Table 2. The Bach and PLCO_{M2012} predictions use equations in Bach *et al.* [5] and Tammemagi *et al.* [11], inflated in accordance with OR_{recal} , the ratio between the lung cancer odds with and without recalibration. OR_{recal} estimates and 95% confidence intervals were derived from the intercept terms of logistic regressions with the logit of Bach or PLCO_{M2012} predictions entered as offsets. Abbreviations: H-L p-value, statistical significance of the Hosmer–Lemeshow χ^2 goodness-of-fit statistic; AuROC, lung cancer discrimination represented as the area (95% confidence interval) under the receiver operator characteristic curve.

Figure 3. Pittsburgh Predictor, Bach, and PLCO_{M2012} calibration plots in PLSuSS. Predicted risks derive from NLST LDCT-calibrated equations, as defined for Figure 2, inflated further by amounts shown in each panel as OR_{recal} . See Figure 2 legend for definitions.

Figure 4.

Six-year lung cancer risks for individuals (• symbols in gray) participating in the Pittsburgh Lung Screening Study (PLuSS; N=3642), according to the Pittsburgh Predictor, Bach, and PLCO_{M2012} prediction models calibrated to NLST LDCT. The gray line marks the points of equivalence between the predictions offered by two models, one shown on the x-axis and the second on the y-axis. The red curves show results from local regressions of the y-axis prediction on the x-axis prediction.

Table 1

Risk factors included in the Pittsburgh Predictor (PP), Bach [5] and PLCO_{M2012} [11] lung cancer risk prediction models

Factor	PP	Bach	PLCO _{M2012}
Age	X	X	X
Duration of smoking	X	X	X
Smoking intensity	X	X	X
Smoking status	X		X
Smoking quit time		X	X
Sex		X	
Asbestos exposure		X	
Race or ethnic group			X
Education			X
Personal history of cancer			X
Family history of lung cancer			X
Chronic obstructive pulmonary disease			X
Body-mass index			X

Table 2

Pittsburgh Predictor, a 4-factor lung cancer risk prediction model

Risk factor		Points
1. Duration of smoking, years	<30	−10
	30–39	0
	40–49	8
	50	14
2. Age, years [1]	younger	0
	older	4
3. Smoking status [2]	current	0
	former	−3
4. Smoking intensity, cigarettes/day	<20	−4
	20–29	0
	30–39	2
	40	5

¹ Older age defined as age 57, 59, 61, and 68 years for persons smoking <30, 30–39, 40–49, and 50 years, respectively.

² The quit smoking category includes former smokers with current attained age at least one year greater than age quit.