**Supplement**

**Identification of candidates for longer lung cancer screening intervals following a negative CT result**

Hilary A. Robbins,1 Christine D. Berg,2 Li C. Cheung,2, Anil K. Chaturvedi,2 Hormuzd A. Katki2

1Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland. Current affiliation: International Agency for Research on Cancer, Lyon, France.

2Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland

Table of Contents

Data source and analysis cohort

Pre-screening risk model: The Lung Cancer Risk Assessment Tool (LCRAT)

Format of LCRAT+CT models for individual risk during screening

CT feature selection for models of risk during screening

Four properties of NLST screening

Results for individual risk of interval lung-cancer

Detailed results for individual risk of next-screen lung cancer

Supplementary Table 1: National Lung Screening Trial CT-arm participants included in analysis of lung-cancer risk after a negative CT screen, by characteristics included in the Lung Cancer Risk Assessment Tool

Supplementary Table 2: Selection of features identified on a negative CT for inclusion in LCRAT+CT models of interval and next-screen lung cancer risk among participants in the National Lung Screening Trial

Supplementary Figure 1: Effect of features noted on a negative CT screen on 1-year risk of interval lung-cancer among participants in the National Lung Screening Trial

References

*Data source and analysis cohort*

The National Lung Screening Trial (NLST) randomized 53,452 ever-smokers to 3 annual screens (denoted T0, T1, and T2) with either low-dose CT or radiography in a 1:1 ratio.1 The trial took place at 33 United States medical centers with enrollment during 2002-2004 and initial follow-up through 2009. Recruitment utilized a variety of strategies2 and eligibility required age 55-74 years, 30 or more pack-years of smoking, and 15 or fewer quit-years. A positive CT had at least one non-calcified nodule with longest diameter ≥4 mm or other suspicious abnormalities.1,3 We analyzed lung-cancer risk among 23,328 CT-arm participants (described in **Supplementary Table 1**) who completed the baseline questionnaire and had at least one negative CT result.

Screening compliance in the NLST was nearly complete (95%), allowing us to model lung-cancer as a binary outcome within one-year periods.1 We defined interval-cancers as cases diagnosed within 1 year of a negative CT and before the next screen.4 Cancers occurring within 1 year of the final (T2) screen were included as interval-cancers. We defined “next-screen” (screen-detected) cancers as those detected due to a positive result at the next annual screen after the negative CT. A “linked-year method” identified next-screen cancers as those occurring within 1 year of a diagnostic follow-up initiated within 1 year after a positive screen.5

We note that our LCRAT+CT model was developed and thus should only be applied among individuals who fit the NLST definition of screen-negative, i.e. those without nodules or with nodules <4mm in longest diameter, but not larger nodules. The NLST used diameter-based assessment of nodules only, and did not employ volumetric measurements, which have been frequently used in more recent studies. It is possible that a different model would be obtained if the definition of screen-negative was based on volumetric measurements.

When the Lung-RADS nodule classification system is applied, LCRAT+CT could not be applied among some individuals in category-2, even though they are recommended to continue usual screening.6 Nodules with diameter 4-6mm are generally regarded as low risk, but do indicate higher risk than nodules <4mm or no nodules.6–9 Thus, most individuals with 4-6mm nodules are unlikely to be candidates for longer screening intervals. We also note that the NLST recorded detailed information only on nodules ≥4mm; thus, we could not predict risk based on individual smaller (<4mm) nodules. Prior studies that followed screened individuals with <4mm nodules support the consideration of biennial screening in this group.9

*Pre-screening risk model: The Lung Cancer Risk Assessment Tool (LCRAT)*  
 The LCRAT was fit to data on 39,180 ever-smokers in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) community care arm. It has been externally validated in 4 cohorts with expected/observed calibration statistics ranging from 0.94 to 1.06, and area under the ROC curve (AUC) discrimination statistics ranging from 0.70 to 0.80.10,11 LCRAT risk factors include age, education, sex, race, smoking intensity/duration/quit-years, body-mass index, family history of lung-cancer, and self-reported emphysema. Using LCRAT, we predicted each individual’s 1-year pre-screening risk, updating age and smoking intensity/duration/quit-years annually. Although we refer to risk predicted by LCRAT as “pre-screening risk,” it can and should be updated annually during screening as individuals accumulate years of age and smoking/quitting.

*Format of LCRAT+CT models for individual risk during screening*

Separate models were fit for interval-cancer and for detection of next-screen cancer. In our LCRAT+CT models, future risk of lung cancer, denoted *r(x)*,is calculated as a function of pre-screening risk, *r*0(*x*), where *x* denotes the risk factors in LCRAT. Specifically, pre-screening risk *r*0(*x*) is raised to an exponent determined by CT features. The exponent for a given individual is calculated by summing the regression coefficients corresponding to each feature on the CT image. For example, a model including only CT-detected emphysema would calculate risk as follows (*CTemp*=1 if emphysema was detected and *CTemp*=0 if not):

Thus, if an individual’s recent negative CT did not detect emphysema, then their future risk is calculated as their pre-screening risk raised to power *β*0: . If emphysema was detected on the CT, then their pre-screening risk is raised to power *β*0+*β*1: This model enforces that, for two people with the same features on their recent negative CT, the person with higher pre-screening risk will also have higher future risk after accounting for CT results. Note that exponents greater than 1 imply decreased risk, and *vice versa*.

We assessed calibration of LCRAT+CT using 10-fold cross-validation and discrimination using the optimism-corrected AUC statistic.12 We used generalized estimating equations (GEE) to confirm that residual intra-individual correlation did not affect estimates (data not shown). We defined outliers as data points falling 1.5\*interquartile-range below the first quartile or above the third quartile.

*CT feature selection for models of risk during screening*

Separately for interval and next-screen lung-cancer, we selected from among 11 CT image features to include in LCRAT+CT models. In the National Lung Screening Trial, CT features were recorded only if they were present, and they were assumed to be absent where no record existed. Thus, there were no missing data for CT features.

For each of interval and next-screen lung-cancer, we examined a list of 11 CT features identified on the recent negative CT (listed in **Supplementary** **Table 2**). First, to identify a list of features to consider, we separately applied a) backwards-stepwise selection to minimize the AIC (Akaike Information Criterion) and b) the lasso (least absolute shrinkage and selection operator)13. Among the features identified by either, we chose features with a large effect and/or high statistical significance, strong biological plausibility, and potential for consistent identification in clinical practice. We did not adjust p-values for multiple testing.

For interval lung-cancer, adenopathy and consolidation were identified as risk factors by both backwards stepwise-selection and lasso. Because they are each rare (~1%) and had similar effect sizes, we combined them as “adenopathy or consolidation” in LCRAT+CT. Stepwise-selection also selected two other features that we rejected as being smaller effects and of unclear measurability in clinical practice.

For next-screen cancer, CT-detected emphysema (30%) and consolidation (0.6%) on the negative CT were identified as risk factors by stepwise-selection, although not lasso. Note that CT-detected emphysema was much more common (30% at T0) than the self-reported baseline emphysema included in LCRAT (7%). Stepwise-selection also selected two other features that we rejected as being of borderline statistical significance and lower plausibility (**Supplementary Table 2**).

*Four properties of NLST screening*

Our LCRAT+CT models account for 4 properties of NLST screening that we observed during model development. First, models including individual pre-screening risk dramatically fit the data better *versus* models that assigned each individual the average risk, both for interval-cancers (likelihood-ratio p<0.0001 and AIC 686 vs. 706) and for next-screen cancers (likelihood-ratio p<0.0001 and AIC 1721 vs. 1810). Thus, accounting for individual pre-screening risk improves risk prediction within a screening program.

Second, model fit did not improve by including any pre-screening risk factor (listed in **Supplementary Table 1**) as an exponent (all p>0.05 for both interval and next-screen cancer models). Thus, pre-screening risk sufficed to encapsulate the effect of pre-screening risk factors on risk during screening.

Third, risks of next-screen cancer and interval-cancer, without considering CT features, were similar across the different NLST screens and inter-screen intervals, respectively (i.e., the Markov model is homogeneous). For risk of interval-cancer in each of the 3 intervals, the exponents were 1.29, 1.42, and 1.35 respectively for T0-T1 (N=19,074), T1-T2 (N=17,802), and 1-year-post-T2 (N=20,045) (p=0.23 for homogeneity). For next-screen cancers, exponents were 1.03 and 1.01 respectively for the T1 (N=18,245) and T2 (N=17,285) screens (p=0.38 for homogeneity).

The fourth property of our LCRAT+CT Markov model is that risk calculations were similar among individuals with a recent negative CT (i.e. at T1), even if their prior result (T0) differs (i.e. a second-order Markov model is unnecessary). For next-screen cancer, we compared the T2 risk calculation between individuals who had a negative T1 result but differed in their T0 result: either T0-negative (N=15,781) or T0-false-positive (N=1,439). Exponents for pre-screening risk were 1.01 for the T0-negative/T1-negative group (95%CI=0.97-1.06) *versus* 0.93 for the T0-false-positive/T1-negative group (95%CI=0.82-1.07). This small difference implies that if calculations were performed separately for screen-negatives with a prior negative vs. false-positive result, next-screen risk would be only marginally higher for those with a prior false-positive. We did not separate these groups because this difference was small and not statistically significant (p=0.26). For interval-cancers, using the post-T1 and post-T2 intervals, exponents were almost exactly the same: 1.38 (95%CI=1.30-1.48) for the negative/negative group (N=32,226 including 16,241 post-T1 and 15,985 post-T2) and 1.37 (95%CI=1.20-1.60) for the false-positive/negative group (N=5,204 including 1,487 post-T1 and 3,717 post-T2) (p=0.99). Thus, given the current negative CT result, we did not include prior screen results in our primary analysis, as they did not contribute substantial risk information.

This fourth property, i.e. that prior results provided little risk stratification beyond the most recent negative screen, was the most surprising. It is not inconsistent with previous analyses of NLST data showing decreasing risk as negative screens accumulated.14 Our analysis compared individuals with the same recent CT result but a different prior result, whereas the previous analysis started with the oldest screen (the least informative screen) and considered the additional informativeness of more recent screens (more informative screens). We note that the point estimates for exponents did suggest a small (though statistically insignificant) increase in risk with a prior false-positive, and thus this issue merits further study using data describing longer-term screening. Regardless, our results suggest that pre-screening risk and features of the recent negative CT are the most important information to consider when considering future risk, and that prior results (whether negative or false-positive) should be given only secondary consideration.

*Results for individual risk of interval lung-cancer*

The final LCRAT+CT model for interval-cancer risk included a single term for the presence of adenopathy or consolidation. This model had good cross-validated internal calibration (43 cases observed vs. 44.9 predicted, p=0.76) and discrimination (optimism-corrected AUC=0.75).

The risk of interval lung-cancer after a negative CT is shown in **Supplementary Figure 1**, which shows risk within the T0-T1 interval for illustration. A total of 43 interval-cancers were diagnosed following 56,921 negative screens, yielding an average risk of 0.08%, a substantial decrease from 0.31% median pre-screening risk.

However, interval-cancer risk was heterogeneous. Median interval-cancer risk was 10-fold lower among the 98% of participants who had neither adenopathy nor consolidation noted on their negative screen (0.03% median risk, IQR=0.02%-0.07%, range=0.002%-0.14%, excluding outliers). Their risk was calculated as *r*0(*x*)1.39 (95%CI=1.32-1.45; *n.b.* exponents greater than 1 imply decreased risk). In marked contrast, the 2% of screen-negative participants with adenopathy or consolidation had 2-fold higher median risk of 0.61% (IQR=0.40%-0.99%). Their risk was calculated as *r*0(*x*)0.91 (95%CI=0.78-1.03). Therefore, after a negative CT in the NLST, risk of interval lung-cancer was reduced 10‑fold for everyone, except the 2% with adenopathy or consolidation for whom risk increased.

When considering potential risk thresholds for longer intervals, we considered LCRAT+CT next‑screen risk only and did not incorporate interval-cancer risk. Some recent studies, for example the Manchester Lung Health Checks,15 have not observed the phenomenon of interval cancers. A recent review of NLST interval cancers concluded that 91% of the negative CTs preceding them were mis-classified and should have been read as positive, which is consistent with our finding that some of these “negative” CTs showed adenopathy.4 If these 91% had been correctly classified, then the true risk of interval cancer would have been essentially negligible. Further, any individual prediction of interval cancer risk would become a function of pre-screening risk only, which is captured in the risk calculation for next-screen risk.

*Detailed results for individual risk of next-screen lung cancer*

The final LCRAT+CT model for next-screen cancer risk included terms for CT-detected emphysema and consolidation. It had good cross-validated internal calibration (138/138.5 cases observed/predicted, p=0.93) and reasonable discrimination (optimism-corrected-AUC=0.73).

The risk of detecting lung-cancer at the screen one year after a negative CT is shown in **Figure 1**, which shows detection at the T1 screen for illustration. A total of 138 next-screen cancers were detected following 35,530 negative screens, giving an average risk of 0.39%, similar to 0.31% median pre-screening risk.

However, next-screen risk was heterogeneous. Median next-screen detection was reduced nearly 2-fold among the 70% of participants who had neither emphysema nor consolidation noted on their negative CT (median risk=0.17%, IQR=0.11%-0.29%). Their next-screen risk was calculated as *r*0(*x*)1.08 (95%CI=1.03-1.12). In contrast, for the 30% of participants with emphysema noted on their negative-CT, median risk increased 1.6-fold (median risk=0.51%, IQR 0.32%-0.81%). Their risk was calculated as *r*0(*x*)0.96 (95%CI=0.91-1.00). Furthermore, for the 0.6% of screen-negative participants with consolidation on their negative CT, median risk increased 5-fold (median risk=1.58%, IQR 1.03%-2.48%). Their risk was calculated as *r*0(*x*)0.77 (95%CI=0.60-0.93).

**Supplementary Table 1: National Lung Screening Trial CT-arm participants included in analysis of lung-cancer risk after a negative CT screen, by characteristics included in the Lung Cancer Risk Assessment Tool**

|  |  |
| --- | --- |
| **Characteristic (at first CT screen)** | **N (%) or median (IQR)** |
| Total number of unique individuals | 23,328 |
| 1-year pre-screening risk at T0 | 0.31% (0.19-0.52%) |
| Included in analysis of interval-cancer risk |  |
| At least once (number of unique individuals) | 23,328 |
| Once (after either T0, T1, or T2) | 4,444 (19.1) |
| Twice (after two of T0, T1, and T2) | 4,175 (17.9) |
| Three times (after each of T0, T1, and T2) | 14,709 (63.1) |
| Included in analysis of next-screen cancer risk |  |
| At least once (number of unique individuals) | 19,749 |
| Once (at either T1 or T2) | 3,968 (20.1) |
| Twice (at both T1 and T2) | 15,781 (79.9) |
| Sex |  |
| Male | 13,813 (59.2) |
| Female | 9,515 (40.8) |
| Age |  |
| 55-59 | 10,244 (43.9) |
| 60-64 | 7,127 (30.6) |
| 65-69 | 4,025 (17.3) |
| 70-74 | 1,932 (8.3) |
| Race/ethnicity |  |
| Non-Hispanic White | 20,910 (89.6) |
| Non-Hispanic Black | 1,054 (4.5) |
| Hispanic | 427 (1.8) |
| Asian/Other | 937 (4.0) |
| Education |  |
| No high school diploma | 1,379 (5.9) |
| High school graduate or equivalent | 5,782 (24.8) |
| Post-high school training other than college | 3,287 (14.1) |
| Associate degree or some college | 5,497 (23.6) |
| Bachelor’s degree | 4,021 (17.2) |
| Graduate school | 3,362 (14.4) |
| Body mass index |  |
| Underweight | 193 (0.8) |
| Normal weight | 6,515 (27.9) |
| Overweight | 9,925 (42.5) |
| Obese | 6,695 (28.7) |
| Family history of lung cancer |  |
| No first-degree relatives | 19,600 (84.0) |
| 1 first-degree relative | 3,468 (14.9) |
| 2 or more first-degree relatives | 260 (1.1) |
| Years since quitting smoking |  |
| Current smoker | 12,066 (51.7) |
| 1 to 5 | 4,212 (18.1) |
| 6 to 10 | 3,381 (14.5) |
| 11 or more | 3,669 (15.7) |
| Total pack-years |  |
| 30-39 | 6,093 (26.1) |
| 40-49 | 6,291 (27.0) |
| 50 or more | 10,944 (46.9) |
| Total years smoked |  |
| <30 | 1,755 (7.5) |
| 30-39 | 9,282 (39.8) |
| 40-49 | 10,119 (43.4) |
| 50 or more | 2,172 (9.3) |
| Cigarettes per day |  |
| <20 | 1,183 (5.1) |
| 20-29 | 11,179 (47.9) |
| 30-39 | 5,456 (23.4) |
| 40 or more | 5,510 (23.6) |
| Self-reported emphysema |  |
| No | 21,609 (92.6%) |
| Yes | 1,719 (7.4%) |

Missing values were imputed as previously described.10 Following these negative screens, 43 interval cancers occurred (18 after T0, 9 after T1, and 16 after T2) and 138 next-screen cancers were detected (64 at T1 and 74 at T2).

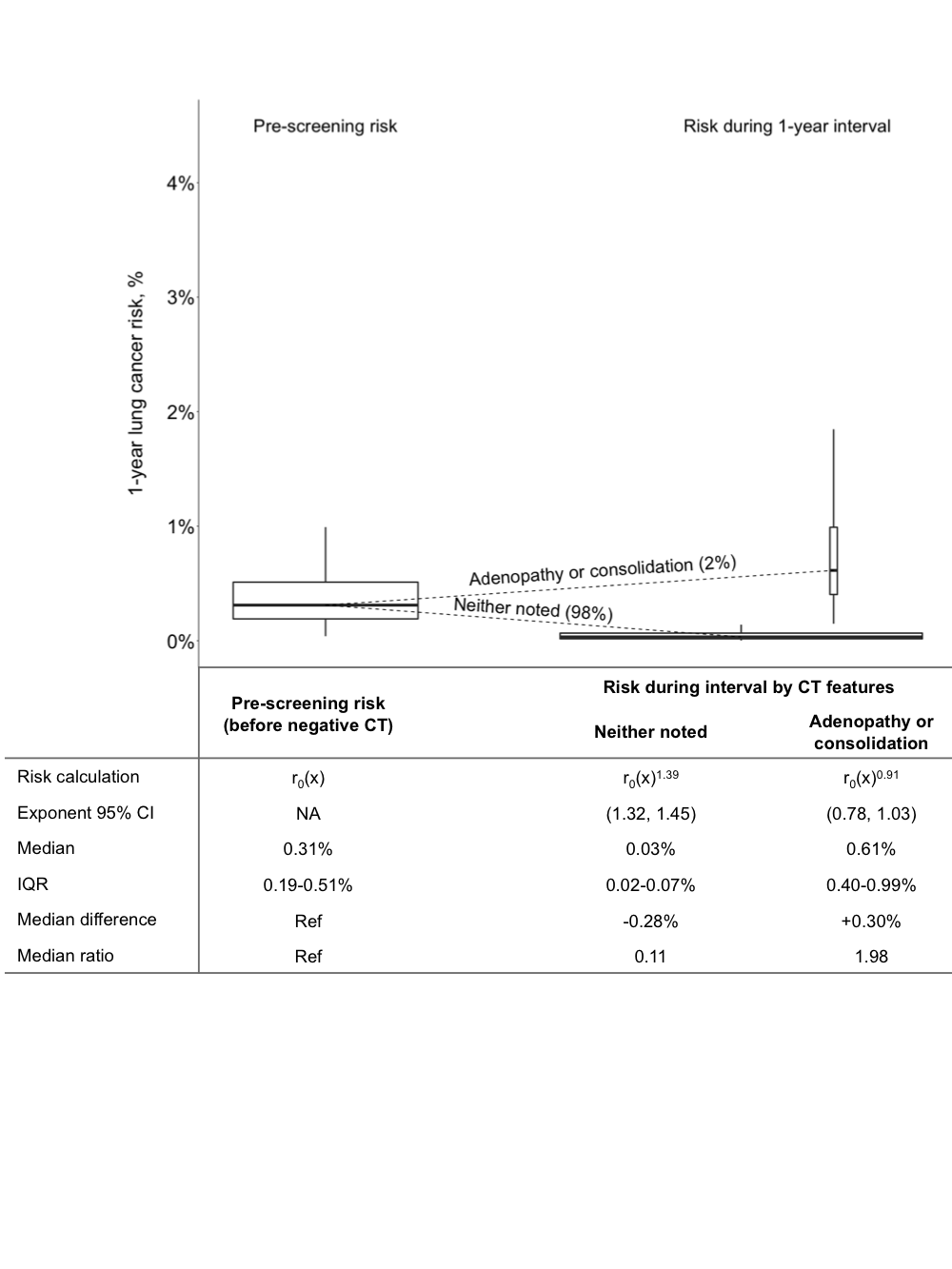
**Supplementary Table 2: Selection of features identified on a negative CT for inclusion in LCRAT+CT models of interval and next-screen lung cancer risk among participants in the National Lung Screening Trial**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Feature** | **Interval lung cancer** | | | **Next screen lung cancer** | | |
| **Stepwise** | **Lasso** | **Final model** | **Stepwise** | **Lasso** | **Final model** |
| Non-calcified hilar/mediastinal adenopathy or mass, ≥10 mm on short axis | Higher risk, 0.46, p<0.001 | Higher risk, 0.47, p<0.001 | Higher risk, 0.48, p<0.001\* | --- | --- | --- |
| Atelectasis, segmental or greater | --- | --- | --- | --- | --- | --- |
| Benign lung nodule(s) | Lower risk, 0.18, p=0.04 | --- | --- | --- | --- | --- |
| Consolidation | Higher risk, 0.48, p<0.001 | Higher risk, 0.48, p<0.001 | Higher risk, 0.48, p<0.001\* | Higher risk, 0.31, p<0.001 | --- | Higher risk, 0.31, p<0.001 |
| Emphysema | --- | --- | --- | Higher risk, 0.11, p=0.001 | --- | Higher risk, 0.12, p<0.001 |
| Non-calcified micronodule(s), opacity <4mm | --- | --- | --- | --- | --- | --- |
| Six or more nodules not suspicious for cancer (opacity ≥4 mm) | --- | --- | --- | Higher risk, 0.19, p=0.04 | --- | --- |
| Reticular/reticulonodular opacities, honeycombing, fibrosis, or scar | --- | --- | --- | Higher risk, 0.07, p=0.04 | --- | --- |
| Other potentially significant abnormality above the diaphragm | Higher risk, 0.21, p=0.02 | --- | --- | --- | --- | --- |
| Other potentially significant abnormality below the diaphragm | --- | --- | --- | --- | --- | --- |
| Pleural thickening or effusion | --- | --- | --- | --- | --- | --- |

Numbers following “higher/lower risk” indicate the amount by which the pre-screening risk exponent decreases (if higher risk) or increases (if lower risk) if the feature is present. P-values are derived from a model including all the variables selected in the corresponding column. “---” indicates that a variable was not selected.

\*In the final model for interval-cancers, adenopathy and consolidation were combined as one effect, as both are uncommon and their effect estimates were nearly identical.

**Supplementary Figure 1: Effect of features noted on a negative CT screen on 1-year risk of interval lung-cancer among participants in the National Lung Screening Trial**



For illustration, this figure was constructed using data from individuals who screened negative at the first NLST screen (T0) and were subsequently at risk for lung cancer during the interval between the first and second screens (T0-T1) (N=19,074). Among these individuals, 338 (2%) had either adenopathy or consolidation, including 225 (1.2%) with adenopathy and 113 (0.6%) with consolidation (N=2 with both). Pre-screening risk r0(x) was calculated using the Lung Cancer Risk Assessment Tool.10 Outliers are not included in the figure, but are included in the calculations in the table. Within each group of boxplots, boxplot widths are scaled by the percentage of the population represented, boxplot heights represent the interquartile range, and the vertical lines (whiskers) represent the range of data excluding outliers. IQR, interquartile range.

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365(5):395–409.

2. Aberle DR, Adams AM, Berg CD, et al. Baseline characteristics of participants in the randomized National Lung Screening Trial. J Natl Cancer Inst 2010;102(23):1771–9.

3. MacMahon H, Austin JHM, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology 2005;237(2):395–400.

4. Gierada DS, Pinsky PF, Duan F, et al. Interval lung cancer after a negative CT screening examination: CT findings and outcomes in National Lung Screening Trial participants. Eur Radiol 2017;27(8):3249–56.

5. 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(3):245–54.

6. American College of Radiology. Lung CT Screening Reporting and Data System (Lung-RADS). Accessed 2017 May 18. Available at https://www.acr.org/Quality-Safety/Resources/LungRADS.

7. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 2013;369(10):910–9.

8. Horeweg N, van Rosmalen J, Heuvelmans MA, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. Lancet Oncol 2014;15(12):1332–41.

9. Tammemagi MC, Schmidt H, Martel S, et al. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. Lancet Oncol 2017;18(11):1523–31.

10. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and validation of risk models to select ever-smokers for CT lung cancer screening. JAMA 2016;315(21):2300–11.

11. Katki HA, Petito LC, Cheung LC, et al. Implications of 9 risk prediction models for selecting ever-smokers for CT lung-cancer screening. Ann Intern Med 2018;169(1):10–9.

12. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361–87.

13. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B 1996;58(1):267–88.

14. Patz EF, Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. Lancet Oncol 2016;17(5):590–9.

15. Crosbie PA, Balata H, Evison M, et al. Second round results from the Manchester “Lung Health Check” community-based targeted lung cancer screening pilot. Thorax 2018; in press.