Research Article

Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort

Kuanrong Li[1], Anika Hüsing[1], Disorn Sookthai[1], Manuela Bergmann[2], Heiner Boeing[2], Nikolaus Becker[1], and Rudolf Kaaks[1]

Abstract

Lung cancer risk prediction models are considered more accurate than the eligibility criteria based on age and smoking in identification of high-risk individuals for screening. We externally validated four lung cancer risk prediction models (Bach, Spitz, LLP, and PLCO$_{M2012}$) among 20,700 ever smokers in the EPIC-Germany cohort. High-risk subjects were identified using the eligibility criteria applied in clinical trials (NELSON/LUSI, DLCST, ITALUNG, DANTE, and NLST) and the four risk prediction models. Sensitivity, specificity, and positive predictive value (PPV) were calculated based on the lung cancers diagnosed in the first 5 years of follow-up. Decision curve analysis was performed to compare net benefits. The number of high-risk subjects identified by the eligibility criteria ranged from 3,409 (NELSON/LUSI) to 1,458 (NLST). Among the eligibility criteria, the DLCST produced the highest sensitivity (64.13%), whereas the NLST produced the highest specificity (93.13%) and PPV (2.88%). The PLCO$_{M2012}$ model showed the best performance in external validation (C-index: 0.81; 95% CI, 0.76–0.86; E/O: 1.03; 95% CI, 0.87–1.23) and the highest sensitivity, specificity, and PPV, but the superiority over the Bach model and the LLP model was modest. All the models but the Spitz model showed greater net benefit over the full range of risk estimates than the eligibility criteria. We concluded that all of the lung cancer risk prediction models apart from the Spitz model have a similar accuracy to identify high-risk individuals for screening, but in general outperform the eligibility criteria used in the screening trials. *Cancer Prev Res; 8(9); 777–85. ©2015 AACR.*

Introduction

Although the avoidance of smoking is the obvious primary strategy to prevent lung cancer, for those who already have accumulated a long-term smoking history, or who are at increased risk due to occupational (e.g. asbestos) exposures, early detection may be a promising secondary strategy for reducing lung cancer-related mortality.

In 2011, the U.S. National Lung Screening Trial (NLST)—the first completed and large-scale randomized trial to examine the efficacy of lung cancer screening by low-dose computed tomography (LDCT)—showed a 20% reduction in lung cancer mortality among long-term heavy smokers screened with LDCT, compared with screening with standard radiography (1). However, screening by LDCT also led to aberrant but mostly false-positive findings suggestive of cancer in around 25% of participants in each screening round, necessitating follow-up examinations associated with additional radiation exposure and psychologic stress. Furthermore, a smaller proportion of screening participants will eventually be classified with a false-positive disease diagnosis after invasive follow-up by endoscopic and/or surgical examinations.

In Europe, a number of smaller lung cancer screening trials with LDCT are still ongoing, including the Netherlands Leuven Screening ONderzoek (NELSON) trial (2), the Danish Lung Cancer Screening Trial (DLCST; ref. 3), the Italian ITALUNG trial (4), the Italian DANTE (Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Essays) trial (5), the German Lung Cancer Screening intervention (LUSI) trial (6), and the UK Lung Screening (UKLS) trial (7). Although an overall analysis of the screening effect on lung cancer mortality in the European screening trials is still pending, preliminary results from the European studies confirm the high rates of aberrant findings after screening and of false-positive findings after subsequent diagnostic work-up (3–6, 8).

In the NLST and the European lung cancer screening trials, high-risk individuals have been identified uniquely on the basis of age and a simplified index for smoking history. Their exact participant eligibility criteria, however, varied considerably in terms of age range, minimum lifetime smoking duration, cumulative smoking exposure (pack years), and maximum time since smoking cessation (Table 1). In parallel to the trials, concerns about the high

[1]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Nuthetal, Germany.

Corresponding Authors: Kuanrong Li, Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, D-69120 Heidelberg, Germany. Phone: 49-6221-42-2219; Fax: 49-6221-42-2203; E-mail: k.li@dkfz.de; and Rudolf Kaaks, E-mail: r.kaaks@dkfz.de.

Table 1. Eligibility criteria used in lung cancer screening trials

		Eligibility criteria
Trial	Age, y	Smoking history
NELSON	50–74	≥15 cigarettes/d for 25 y; or ≥10 cigarettes/d for at least 30 y; if former smokers, quitting time ≤10 y.
LUSI	50–69	≥15 cigarettes/d for at least 25 y; or ≥10 cigarettes/d for at least 30 y; if former smokers, quitting time ≤10 y.
DLCST	50–69	Pack-years ≥20; if former smokers, age at quitting >50 y and quitting time <10 y
ITALUNG	55–69	Pack-years ≥20 since the last 10 years; if former smokers, quitting time <10 y.
DANTE	60–74	Pack-years ≥20.
NLST	55–74	Pack-years ≥30; if former smokers, quitting time ≤15 y.

percentage of false-positive findings in LDCT screening have driven the interest in application of quantitative and more informative lung cancer risk prediction models to classify individuals into strata of high versus low risk (9–12). Several lung cancer risk prediction models have been developed from Western populations, including models by Bach and colleagues (13) and Spitz and colleagues (14), and from the Liverpool Lung Project (LLP; ref. 15) and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening trial (PLCO; refs. 16, 17). Besides a more detailed modeling of risks in relation to age and the lifetime duration plus average intensity of smoking, these models also incorporated additional risk predictors such as asbestos exposure, pre-existing lung diseases (COPD, pneumonia) or family history of cancer (Table 2). So far, however, only few studies have externally validated these models with prospective data and compared their performance with the simpler eligibility criteria used in terms of risk stratification (16–19). Furthermore, no study has addressed the optimal threshold of lung cancer risk above which individuals should be eligible for screening, and whether around a given threshold, the eligibility criteria or prediction models perform equally well.

In the present study, we externally validated four lung cancer risk prediction models, that is, the Bach model, the Spitz model, the LLP model, and the 2012 version of the PLCO model (PLCO$_{M2012}$) within a population-based prospective cohort in Germany, and we compared the performance of these models with the various eligibility criteria used in screening trials in terms of sensitivity, specificity, and predictive values. In addition, we performed a decision curve analysis (20) to examine and compare the clinical utility of the various risk models and selection criteria in terms of expected net benefits. It should be noted that this study did not consider the lung cancer risk prediction model developed by Hoggart and colleagues (21) because it was originally developed on data including our cohort.

Materials and Methods

Study population

Our study population consisted of ever smokers in the German part of the European Prospective Investigation of Cancer and Nutrition (EPIC; refs. 22, 23). Participant recruitment and data collection for the European EPIC study and for the EPIC-Germany part have been described previously (22–24). In brief, the EPIC-Germany cohort was established from 1994 to 1998 in the cities of Heidelberg and Potsdam, respectively, and eventually recruited a total of 53,088 participants (30,255 women and 22,833 men), ages 35 to 69 years at study entry. The study was approved by the local ethics committees and an informed consent was provided by all study participants. In the present study, we limited our analysis to ever smokers who were 40 years or older and had complete information on basic risk factors to be studied ($N = 20,700$).

Assessment of risk factors and ascertainment of the disease outcome

Baseline information on lifestyle factors, including cigarette smoking, was collected using a self-administered questionnaire

Table 2. Predictors included in the Bach, Spitz, LLP, and PLCO$_{M2012}$ model and the availability in the EPIC-Germany cohort

Predictors	Bach	Spitz	LLP	PLCO$_{M2012}$	Availability in EPIC-Germany
Age	X	X[a]	X[a]	X	X
Sex	X	X[a]	X[a]		X
Smoking status				X	X
Cigarettes smoked/d	X			X	X
Smoking duration	X		X	X	X
Duration of cessation	X			X	X
Age at cessation		X			X
Pack-years		X			X
Asbestos exposure	X	X	X		X
Dust exposure		X			
Pneumonia			X		
COPD				X	
Emphysema		X			
Hay fever		X			X
Prior diagnosis of malignant tumor			X	X	X[b]
Family history of cancer		X			X
Family history of lung cancer			X	X	
BMI				X	X
Education				X	X

[a]Age and sex were not included the model as predictors. However, the effects of age and sex were incorporated into the risk prediction by using the age- and sex-specific lung cancer incidence rates to estimate the baseline hazards.
[b]This information was available but subjects with prior diagnosis of malignant tumor were excluded from the analysis.

survey combined with a computer-guided in-person interview. Detailed information was collected on smoking status and cigarettes smoked per day at 20, 30, 40, and 50 years of age, and at recruitment. Information on age at initiation and age at cessation of smoking was also collected when applicable. Occupational exposure to asbestos was assessed by asking participants whether they had worked in asbestos cement industry or asbestos insulation industry. Anthropometric measurements were performed in a baseline physical examination. Other risk factors that were included in some of the existing risk prediction models—emphysema and dust exposure (Spitz model), history of pneumonia (LLP model), chronic obstructive pulmonary disease (COPD; the PLCO$_{M2012}$ model), and family history of lung cancer (LLP model and PLCO$_{M2012}$ model)—were not assessed in the EPIC-Germany cohort and for analyses; we therefore assumed that these risk factors were absent in our study population.

Incident lung cancers occurring during the follow-up were ascertained by a combination of follow-up questionnaires, linkages to health insurance records and cancer registries, combined with full clinical verification against hospital and pathology records. In the present study, lung cancers are defined as invasive primary cancers in the lung, coded as C34 under the 10th Revision of the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD-10). Information on vital status was also collected through a combination of active follow-up and regular linkages to municipal population registers.

Statistical Analysis

We estimated the risk of developing lung cancer over the next 5 years for all the subjects in our study population. The original models by Bach and Spitz provide estimates on the risk of developing lung cancer in the next one year in a competing risk context. For Spitz' model, which included separate estimations for ex- and current smokers, we applied the model components for past or current smokers separately to the corresponding subsets of the EPIC-Germany cohort. To calculate 5-year lung cancer risks, we cycled the 1-year risk algorithm five times. In each cycle, we increased the relevant predictors, namely duration of smoking for current smokers and time since quitting for former smokers, by 1 year. The LLP model can be used to estimate 5-year risk directly. The PLCO$_{M2012}$ model was designed to estimate 6-year lung cancer risk. In order to derive 5-year risk estimates from this model, we assumed a linear increase in lung cancer risk with time over a restricted period of 6 years, and calculated 5-year risk by multiplying the 6-year risk by a factor of 5/6.

The discriminatory power of the risk prediction models was assessed using the C-index, which is equivalent to the area under the receiver operating characteristic (ROC) curve, and a C-index of 0.5 represents no discriminatory power (25). We examined model calibration by calculating the ratio of the expected number (E) to the observed number (O) of lung cancers diagnosed in the first 5 years of follow-up (first 6 years for PLCO$_{M2012}$). The 95% confidence interval (CI) of the ratio was calculated as $(E/O) * \exp(\pm 1.96 * \sqrt{\frac{1}{O}})$.

The discrimination capacity for the eligibility criteria used in the clinical trials (LUSI, DLCST, ITALUNG, DANTE, and NLST) was examined by calculating sensitivity, specificity, and positive predictive values (PPV). As a contrast with PPV, we also calculated the complement of negative predictive values (NPV), namely 1 – NPV

(denoted as cNPV), to indicate the risk stratification by the eligibility criteria and lung cancer risk prediction models. The NELSON criterion and the LUSI criterion yielded an identical selection in our cohort data and therefore they were hereafter referred to as the NELSON/LUSI criteria. The degree of concordance between high-risk groups selected in EPIC-Germany through the various eligibility criteria was evaluated by calculating pairwise kappa statistics, and by drawing intersection diagrams using the "SPAN" software (Marshall R, University of Auckland, New Zealand). To compare the discrimination capacity between the trial eligibility criteria and the risk prediction models, for each model, we determined a risk threshold so that the number of subjects with lung cancer risk estimate above that threshold was equal to the number of high-risk subjects identified with specific selection criteria. Then we calculated the sensitivity, specificity, PPV, and cNPV of the risk stratifications defined with the selection criteria and the corresponding model-based risk thresholds.

For each of the eligibility criteria and risk prediction models, we also performed decision curve analyses to estimate the net benefits if it was used for high-risk subject identification. Decision curve analysis exploits a dualistic interpretation of a risk threshold. It is considered as (i) determining a risk threshold to select persons for screening/treatment, which entails specific harm and benefits and as (ii) representing a loss-to-profit ratio for such screening/treatment. While the first is straightforward, the second meaning is described in detail by Vickers and Rousson (9, 26). In short, a risk threshold p_t, would only be considered for screening selection, if the expected profit (P) gained by the screening for a true positive subject was assumed to counterbalance the potential loss (L) if that subject was false positive, that is, $p_t*P = (1-p_t)*L$. Resolving this to $L/P = p_t/(1-p_t)$ allows to derive the net benefit u, as the difference between the true-positive rate (a) and the corresponding false-positive rate (b) weighted by L/P (i.e., the odds of p_t), namely, $u = a – b*p_t/(1-p_t)$. The net benefits and the corresponding risk thresholds are combined in a decision curve. A selection criterion or risk model showing higher net benefit for thresholds within a potentially relevant range of L/P is considered superior over others.

Statistical analyses were performed with SAS (version 9.3, SAS Institute) and R (version 3.0.2, R Foundation for Statistical Computing). The decision curve analyses were performed using the R program from Vickers and colleagues (27).

Results

The sex-specific distributions of baseline characteristics associated with lung cancer risk among 20,700 ever smokers in the EPIC-Germany cohort are shown in Table 3. Heavy smoking in terms of lifetime smoking duration and number of cigarettes smoked per day was more common among men than women, although among the women, a larger proportion of current smokers was observed. The prevalence of occupational history related to possible asbestos exposures was low in both sexes (1.2% among men, 0.1% among women). Men overall had a higher body mass index (BMI) and higher education level than women. The frequency of self-reported family history of cancer was higher in women than in men. A total of 92 study participants had a clinical diagnosis of lung cancer within the first 5 years of prospective follow-up (and 126 cases in the first 6 years).

Applying the various eligibility criteria of the NLST and the ongoing European screening trials to the 20,700 ever smokers in

Li et al.

Table 3. Baseline characteristics of the study population by sex among the ever smokers in the EPIC-Germany cohort ($N = 20,700$)

	Men ($n = $ **12,565**)	Women ($n = $ **8,135**)
Age (range)	53.2 (40.0–69.4)	49.6 (40.0–69.0)
LC in the first 5-years of follow-up	76	16
Age at diagnosis of lung cancer	63.5 (42.8–76.6)	62.0 (44.7–75.6)
Current smokers (%)	4,648 (37.0)	3,649 (44.8)
Duration of smoking (y)	24.0 (0.5–53.5)	23.0 (0.5–52.0)
No. of cigarettes/d	15.0 (0.04–85.0)	8.8 (0.02–72.6)
Pack-years	12.0 (0.05–127.5)	5.6 (0.05–66.0)
Quitting time (former smokers)	16.0 (0.1–43.8)	14.1 (0.1–43.2)
Asbestos exposure	151 (1.2)	8 (0.1)
Hay fever	1,253 (10.0)	1,253 (15.4)
BMI (kg/m^2)	26.9 (14.5–55.2)	25.0 (15.1–58.7)
Family history of cancer	3,882 (30.9)	3,129 (38.5)
College education	4,609 (36.7)	2,051 (25.2)

Abbreviation: LC, lung cancer.

EPIC-Germany led to the inclusion of quite variable subsets of individuals considered at risk, with numbers ranging from 3,409 (NELSON/LUSI) to 1,458 (NLST; Supplementary Fig. S1). As shown in Supplementary Table S1, one major reason for obtaining such variable subsets was different age ranges covered by the eligibility criteria. When pairwise comparisons between the eligibility criteria were restricted to the overlapping age ranges, discordances between the identified high-risk subsets largely disappeared, although some degree of discordance remained as a result of different ways of using smoking history as discriminator.

Table 4 shows the validation performance of the four lung cancer risk prediction models in terms of discriminatory power and model calibration. With a C-index of 0.81 (95% CI, 0.76–0.86), the Bach model and the PLCO$_{M2012}$ model had a marginally higher discriminatory power than the Spitz model and the LLP model, which showed C-indices of 0.78 and 0.79, respectively. Compared with 92 lung cancer cases actually diagnosed in the first 5 years of follow-up, the Bach model and the LLP model predicted 81 and 103 cases, respectively, whereas the Spitz model overestimated the predicted case number by nearly four times (345). The PLCO$_{M2012}$ model predicted 130 cases in contrast with 126 cases actually diagnosed in the first six years. The PLCO$_{M2012}$ showed the best overall calibration ($E/O = 1.03$, 95% CI, 0.87), whereas the models by Bach and LLP also showed E/O ratios not significantly deviating from 1.0. The Spitz model yielded an approximately 4-fold overestimation of the absolute lung cancer risk. Supplementary Fig. S2 shows the ROC curves for each of the lung cancer risk prediction models.

As shown in Table 5, the NELSON/LUSI and DLCST criteria, which covered the broadest age range (50–69 years) led to the inclusion of the largest numbers of individuals classified at risk (3,409 and 2,931, respectively), and showed the highest sensitivities (63.04% and 64.13%) but relatively lower specificities

Table 4. Validation performance of the lung cancer risk models applied to the ever smokers in the EPIC-Germany cohort ($N = 20,700$)

	Discrimination (C-index, 95% CI)	Calibration[a] (E/O, 95% CI)
Bach model	0.81 (0.76–0.86)	0.88 (0.72–1.08)
Spitz model	0.78 (0.73–0.83)	3.75 (3.06–4.60)
LLP model	0.79 (0.73–0.83)	1.12 (0.92–1.37)
PLCO$_{M2012}$	0.81 (0.76–0.86)	1.03 (0.87–1.23)

[a]The expected and observed numbers of lung cancer are 81/92 (Bach), 345/92 (Spitz), 103/92 (LLP), and 130/126 (6-year risk, PLCOM$_{2012}$).

(83.74% and 86.06%) and PPV (1.70% and 2.01%) for the prediction of future lung cancer diagnoses. The DANTE and NLST criteria, in contrast, resulted in the inclusion of fewer subjects classified at risk (1,500 and 1,458, respectively) and were associated with higher specificities (92.90% and 93.13%) but lower sensitivities (39.13% and 45.65%). The PPV was highest for the NLST criterion (age range above 55 years), due to the best combination of specificity and sensitivity, and was also clearly higher than for the DANTE criterion, which covered an older age range (above 60 years) but used a somewhat less stringent definition of smoking history. The cNPV was lowest (best value) for the DLCST (0.18%) and highest for DANTE (0.29%). For the ITALUNG criterion, the number of subjects classified at risk ($n = 2,170$), sensitivity, specificity, PPV, and cNPV all had intermediate values.

Table 5 also compares the performances of the risk prediction models to the inclusion criteria used in screening trials (NELSON/LUSI, DLCST, ITALUNG, and NLST). In particular, the PLCO$_{M2012}$ model performed systematically better than, or at least as well as, the various eligibility criteria used in trials thus far, whereas the performance of the Spitz model was the worst in all comparisons. For example, among 3,409 high-risk subjects as identified with the NELSON/LUSI criterion, 58 of the 92 lung cancer cases were diagnosed in the first 5 years of follow-up, whereas given the same size, the high-risk group identified with the PLCO$_{M2012}$ model contained 12 more lung cancer cases than the one identified with the Spitz model (61 vs. 49). Likewise, in comparison with the ITALUNG criterion (2,170 classified high-risk; 51 cases identified), PLCO$_{M2012}$ identified 55 cases and the model by Spitz only 41. At any fixed size for the high-risk subset, PLCO$_{M2012}$ was the model showing the highest sensitivity and PPV and lowest cNPV, whereas the Spitz model always led to the lowest sensitivity and PPV and highest cNPV, and the other models had intermediate performances.

Figure 1A compares the net benefits obtained from screening high-risk subsets identified via different selection approaches, including two extreme strategies: "screen-all" (indicated by the gray line) and "screen-none" (represented by the x-axis). The line for the screen-all strategy crosses the x-axis at the risk threshold (p_t) that corresponds to the disease incidence of the cohort (20). In the present study, this threshold was 0.46%, which indicated a failure of the screen-all strategy when the expected benefit from selecting a single true positive for screening is deemed to weigh less than 216 times the potential harms due to selecting a false positive for screening [$L/P = 0.0046/(1-0.0046) = 1/216$]. For risk thresholds greater than 0.18% (DLCST, corresponding to an $L/P > 1/554$) to 0.26% (NLST, corresponding to an $L/P > 1/384$), the binary eligibility criteria were superior to the screen-all strategy. However, as compared with the screen-none strategy, in Fig. 1A represented by the x-axis, the trial eligibility criteria showed greater net benefit only for risk thresholds up to 1.76% (NELSON/LUSI; $L/P = 1/56$) to 3.04% (NLST; $L/P = 1/32$). Regarding the prediction models, compared with both the screen-all or screen-none strategies, all models except that of Spitz showed greater net benefit up over the full range of risk thresholds, and they performed generally as well as or better than each of the binary eligibility criteria. For the Spitz model, the net benefit became negative when the threshold was above 0.65% ($L/P = 1/153$). After dividing the risk estimates by a factor of 3.75; however, the Spitz model also showed positive net benefits for all risk thresholds, although its performance in decision curve

Table 5. Comparison of the lung cancer risk prediction models and the eligibility criteria among ever smokers in the EPIC-Germany cohort ($N = 20,700$)

Criteria/model	Threshold (%)	Cases included[a]	Sensitivity (%)	Specificity (%)	PPV (%)	cNPV (%)
$n_{positive} = 3,409$						
NELSON/LUSI	—	58	63.04	83.74	1.70	0.20
Bach	0.764932	57	61.96	83.73	1.67	0.20
Spitz	3.511813	49	53.26	83.69	1.44	0.25
LLP	0.852661	54	58.70	83.72	1.58	0.20
PLCO$_{M2010}$	1.006991	61	66.30	83.75	1.79	0.18
$n_{positive} = 2,931$						
DLCST	—	59	64.13	86.06	2.01	0.18
Bach	0.898835	57	61.96	86.05	1.94	0.20
Spitz	3.742099	45	48.91	86.0	1.54	0.26
LLP	0.940001	51	55.43	86.02	1.74	0.23
PLCO$_{M2010}$	1.127746	59	64.13	86.06	2.01	0.18
$n_{positive} = 2,170$						
ITALUNG	—	51	55.43	89.72	2.35	0.22
Bach	1.178258	51	55.43	89.72	2.35	0.22
Spitz	4.180923	41	44.56	89.67	1.89	0.28
LLP	1.179983	46	50.0	89.69	2.12	0.25
PLCO$_{M2010}$	1.726509	55	59.80	89.74	2.53	0.20
$n_{positive} = 1,500$						
DANTE	—	36	39.13	92.90	2.40	0.29
$n_{positive} = 1,458$						
NLST	—	42	45.65	93.13	2.88	0.26
Bach	1.553146	44	47.83	93.14	3.02	0.25
Spitz	4.723909	34	36.96	93.09	2.33	0.30
LLP	1.528621	40	43.48	93.12	2.74	0.27
PLCO$_{M2010}$	2.110365	46	50.0	93.15	3.16	0.24

[a]Cases that were diagnosed the first 5 years of follow-up.

analysis remained inferior to that of the other models. Figure 1B shows the cumulative probability plots for the absolute 5-year risk estimates from each of the four risk prediction models. It is worth noting that for the well-calibrated models (all but the Spitz model), the maximum 5-year absolute risk estimate in our cohort was about 7%.

Discussion

In this German study population, we found that the eligibility criteria used in various lung cancer screening trials and lung cancer risk prediction models overall had broadly similar accuracy in identifying high-risk populations for screening, although some performed marginally better than the others. Nonetheless, decision curve analyses indicated that well-validated lung cancer risk prediction models may have broader clinical utility than the binary classifiers, given the generally more comprehensive ranges of risk thresholds for which these models showed a positive net benefit.

Applying the eligibility criteria of different lung cancer screening trials to our study population resulted in substantial variability in the number of high-risk subjects identified. One major reason for this variability is that the criteria covered different age ranges. Our data document that extending the age range to include younger persons (as with the NELSON/LUSI and DLCST criteria, in comparison with DANTE and NLST) increases sensitivity at the cost of a substantial reduction in specificity and PPV. Apart from age, a smaller part of the variability between the high-risk subsets was due to differences in the way smoking history was accounted for. Although the criteria used in DANTE and NLST, in particular, identified two largely overlapping high-risk groups of similar size (Supplementary Fig. S1), the former contained six fewer lung cancers than the latter (36 vs. 42, out of 92). Thus, compared with

NLST, the DANTE criterion appeared to be less accurate for selecting high-risk subjects, and most likely this is explained by the inclusion of long-term quitters. Of all eligibility criteria tested, the NLST criterion resulted in the best risk stratification, and it presently is also the eligibility criterion most referred to in screening recommendations, motivated by the positive findings of the NLST trial.

Contrary to the binary eligibility criteria, the risk prediction models allow the identification of individuals reaching a minimum risk threshold in a multivariable fashion without setting predefined age limits. Apart from smoking, each of the models included additional risk variables, such as previous or existing lung disease (COPD, pneumonia), exposures to dust or asbestos, and family history of cancer (general, or for lung cancer specifically). As in our data, a large part of these latter variables were missing; however, differences in discrimination performance observed in our study were essentially determined by the way in which smoking history was modeled as a multidimensional exposure. The PLCO$_{M2012}$ and Bach models showed modestly higher discriminatory power compared with the Spitz and LLP models. This may be explained by the fact that the PLCO$_{M2012}$ and Bach models include separate variables for cigarettes smoked per day, smoking duration, and duration of cessation, whereas the Spitz model only take into account pack-years and age at cessation and the LLP model only included smoking duration without considering intensity. Furthermore, the PLCO$_{M2012}$ and Bach models, both account for the nonlinear effect of smoking intensity (cigarettes per day). Several independent studies have shown a significantly better model fit for lung cancer risk models that consider smoking intensity and duration separately, and that account for nonlinearity with smoking intensity, as compared with models based on intensity, duration, or their product (pack-years) alone (28–30).
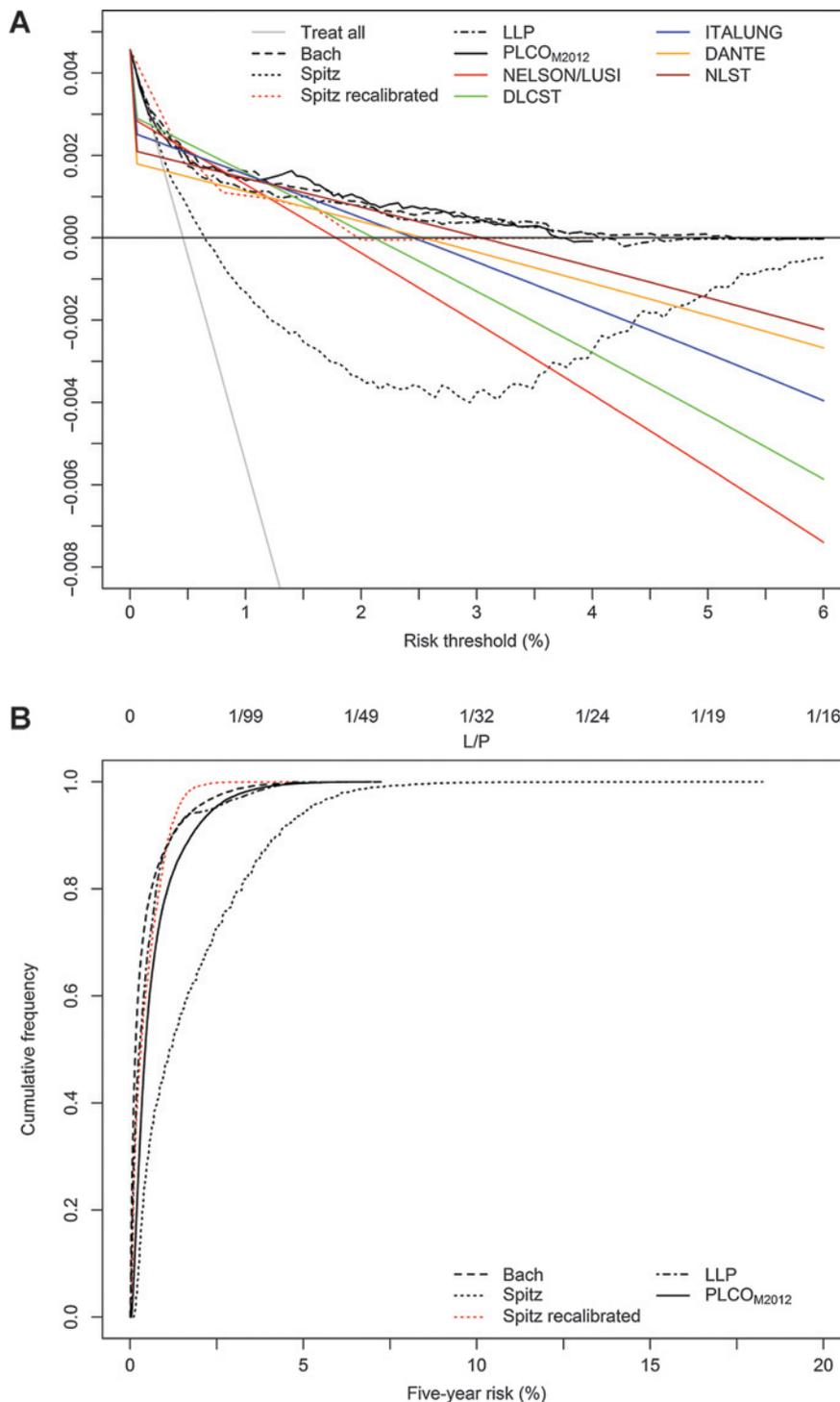
Figure 1.
Net benefits obtained from different selection approaches (A) and the cumulative distribution of 5-year absolute risk estimates (B).

An analysis within the U.S. PLCO cohort showed that the PLCO$_{M2012}$ model, compared with the NLST selection criterion, identified about 12% more lung cancers and improved the sensitivity from 71.1% to 83.0% and PPV from 3.4% to 4.0% without losing specificity (18). In our analyses, the PLCO$_{M2012}$ model also performed better than the NLST criterion, with four additional lung cancer cases (46 vs. 42) out of 92 identified at

equal size of the group at risk. Furthermore, in the PLCO cohort and in our data, the PLCO$_{M2012}$ model showed similarly good discrimination (a *C*-index of 0.81 for both cohort studies), although absolute risk distributions differed substantially between PLCO and our data (in the PLCO study 38% of subjects met the NLST criterion, whereas only 7% of the subjects in our study did).

Decision analysis is a still relatively novel way of evaluating the performance of a diagnostic procedure. It is a method to calculate a net benefit, namely the difference between the expected benefit of screening if a person actually has the disease minus the harms that would follow a false-positive test or diagnostic sequence (9, 31). Central to decision curve analysis is the concept of a risk threshold above which an individual would choose to be screened/treated. If a diagnostic chain and subsequent treatments have high efficacy and only minimal adverse effects, the risk threshold for screening will be low. In contrast, the risk threshold will be high when diagnostic accuracy is low and false-positive results incur major potential harms. In decision curve analysis, if one decision curve is highest over the full range of risk thresholds, then the associated (pre-) diagnostic approach will be the generally preferred way of testing regardless of an individual's risk threshold value. The latter is what we actually observed for each of the risk prediction models, except that of Spitz, compared with the various trial eligibility criteria. Furthermore, the $PLCO_{M2012}$ model appeared to perform uniformly (although only slightly) better than the other risk prediction models, and this fits the observation that $PLCO_{M2012}$ risk estimates appeared to provide slightly better discrimination and was better calibrated in comparison with the other models as well as to the various binary eligibility criteria. The poor performance of the Spitz model can be largely explained by the gross miscalibration of this model with respect to our study population. Calibration is a basic prerequisite for decision curve analysis to provide valid results, as documented by recent mathematical simulation analyses (32).

From an operational viewpoint, a key question is what absolute risk threshold should be used to select individuals for lung cancer screening. The answer to this question depends on the weights given to measurable clinical risks and additional harms such as psychologic stress inflicted upon individuals who eventually would have a false-positive finding upon screening and subsequent diagnostic work-up, and again, decision curve analyses may provide a valuable approach especially for this type of evaluation. In the NLST, three annual rounds of LDCT screening imply a 0.31% reduction in absolute risk of dying from lung cancer within a prolonged time period of 5 years, indicating that 320 individuals must participate in up to three rounds of screening to prevent one lung cancer death (1). From NLST, and other screening trials, it can also be approximated that in three screening rounds among 320 individuals, there will be around 125 positives including 5 true positives and 195 negatives including 2 false negatives, given the probability of having at least one positive screening result being 39%, an overall PPV of 4%, and an overall NPV of 99%. The majority of these screening participants will have to undergo noninvasive, radiologic follow-up investigations associated with psychologic stress and moderate radiation exposures, whereas a much smaller fraction of 4 to 5 false positives will undergo more invasive endoscopic examinations, with or without microsurgery. In terms of decision analysis, if overall it was considered acceptable to have 120 false positives for one lung cancer death prevented with LDCT within a screening period of 5 years, we would require a 5-year risk threshold that leads to a lung cancer incidence rate of 2.2% (7/320) in the identified high-risk group. This risk threshold is about 1.13% according to the $PLCO_{M2012}$ model and 1.18% according to the LLP model. A higher threshold should be used, however, if the greater severity of the more invasive follow-up examinations (4 to 5 individuals per lung cancer death pre-

vented) is given additional weight. A further potential harm, not yet accounted for in the above calculation, is overdiagnosis. A recent analysis of the NLST data indicates that more than 18% of all lung cancers detected by LDCT may be indolent, and that for one life saved (i.e., out of 320 individuals screened), the expected number of cases of overdiagnosis is 1.38 (33).

Although our analyses document the extent to which a higher risk threshold may further enrich a screening sample with actual lung cancer cases, they do not provide any direct information on how a higher-risk selection may simultaneously affect the clinical and molecular characteristics of tumors and patient survival, as well as numbers and subtypes of false-positive diagnoses or the likelihood of being overdiagnosed. Furthermore, there may be theoretical issues in matching the cumulative risk period (e.g., 5 years) for which an individual's lung cancer risk is estimated through a prediction model to the time period for which cancers may be diagnosed through a given screening program (e.g., annual rounds of screening over three years, plus the average sojourn time for lung cancer, which is estimated to be around 2 years; ref. 34). These and other considerations show that the one-to-one matching of estimated absolute lung cancer risks to the theoretical risk thresholds in decision curve analyses is not without complications. To overcome these methodologic issues, decision curve analyses should be ideally performed within datasets obtained from actual randomized screening trials, covering all steps from selection of screening participants till final outcomes of diagnosis, treatments, and proportions of possible over-diagnosis.

Besides the balance between expected clinical benefits and harms of lung cancer screening, cost-effectiveness may be another important dimension to be taken into account when establishing an appropriate risk threshold to select individuals for a screening program. A recently published work from Black and colleagues shows that confining screening with LDCT to higher risk subjects (fourth and fifth quintiles of the $PLCO_{M2012}$ risk score) led to significantly lower cost per quality-adjusted life year gained when compared with screening among the lower-risk subjects (35).

It should be noted that the well-calibrated risk prediction models, including the models of $PLCO_{M2012}$, LLP and Bach predicted a maximum 5-year risk of lung cancer around 4%, and all showed positive net benefit for absolute risk thresholds within the approximate range of 1% to 3.5%. These estimates differ strongly from those reported in other study populations, such as the Liverpool Lung Cancer Project cohort, where much higher absolute risks were observed for larger parts of the study population. This indicates that absolute risk estimates and the results of decision curve analyses may be very population dependent. Our analyses focused on ever smokers aged up to 69 years maximally at baseline, and recruited in the 1990s, whose overall lifetime duration of smoking and average smoking intensity may have been generally lower than observed in other studies, such as LLP or PLCO, which had a higher maximum age at recruitment and whose participants were drawn from a population in which there was an earlier post-war introduction of widespread smoking habits (36, 37). It is important to note, however, that the distributions of the 5-year absolute risks among ever smokers in the EPIC-Germany cohort, as reported 1994–1998, were very similar to those derived from the more recent LUSI trial data (2007–2009; Supplementary Fig. S3). The absence of several risk predictor variables that were missing in the EPIC cohort, but are part of

some of prediction models, may have further contributed to a narrower distribution of risk estimates; however, simulation analyses (Supplementary Fig. S4) showed that this may have introduced only very modest biases.

In the United States, the American Cancer Society and several other national advisory bodies have recommended lung cancer screening to individuals who meet the NLST eligibility criterion (38–40). The U.S. Preventive Services Task force recommended the same criterion for screening, but with a wider age range (55–79 years; ref. 41). In Europe, no equivalent guidelines have been formulated yet. Whether the eligibility criteria formulated thus far are optimal for population-based screening programs is still being debated (12, 13, 18). On one hand, these criteria exclude from screening younger heavy smokers who nonetheless may have a comparatively high risk. On the other hand, the higher number of false-positive findings justifies the need for more accurate risk stratification. In principle, the use of a well-calibrated and discriminating lung cancer risk prediction model, combined with a clinically well-motivated absolute risk threshold that indicates a balance between benefits and harms of screening, would provide a response to these issues. It is anticipated that the ongoing European lung cancer screening trials will soon provide additional data on the complex balance of benefits of lung cancer screening with LDCT in terms of lives saved against a variety of negative side effects.

Ethical Approval

The EPIC-Germany cohort study received approval from local ethnics committees. All cohort participants provided an informed consent.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: K. Li, A. Hüsing, M. Bergmann, H. Boeing, R. Kaaks
Development of methodology: K. Li,
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. Li, M. Bergmann, H. Boeing, N. Becker, R. Kaaks
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): K. Li, A. Hüsing, R. Kaaks
Writing, review, and/or revision of the manuscript: K. Li, A. Hüsing, D. Sookthai, H. Boeing, N. Becker, R. Kaaks
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D. Sookthai, M. Bergmann, H. Boeing, R. Kaaks
Study supervision:

References

1. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.
2. van Iersel CA, de Koning HJ, Draisma g, Mali WP, Scholten ET, Nackaerts K, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). Int J Cancer 2007; 120:868–74.
3. Pedersen JH, Ashraf H, Dirksen A, Bach K, Hansen H, Toennesen P, et al. The Danish randomized lung cancer CT screening trial–overall design and results of the prevalence round. J Thorac Oncol 2009;4:608–14.
4. Lopes PA, Picozzi G, Mascalchi M, Maria CF, Carrozzi L, Comin C, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. Lung Cancer 2009;64:34–40.
5. Infante M, Cavuto S, Lutman FR, Brambilla G, Chiesa G, Ceresoli G, et al. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. Am J Respir Crit Care Med 2009;180:445–53.
6. Becker N, Motsch E, Gross ML, Eigentopf A, Heussel CP, Dienemann H, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. J Cancer Res Clin Oncol 2012;138:1475–86.
7. Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM, Field JK. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. Thorax 2011;66:308–13.
8. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volum CT screening. N Engl J Med 2009;361:2221–9.
9. Cassidy A, Duffy SW, Myles JP, Liloglou T, Field JK. Lung cancer risk prediction: a tool for early detection. Int J Cancer 2007;120:1–6.
10. Duffy SW, Raji OY, Agbaje OF, Allgood PC, Cassidy A, Field JK. Use of lung cancer risk models in planning research and service programs in CT screening for lung cancer. Expert Rev Anticancer Ther 2009;9:1467–72.
11. Field JK, Raji OY. The potential for using risk models in future lung cancer screening trials. F1000 Med Rep 2010;2:pii:38.
12. Field JK, Chen Y, Marcus MW, Mcronald FE, Raji OY, Duffy SW. The contribution of risk prediction models to early detection of lung cancer. J Surg Oncol 2013;108:304–11.
13. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003;95: 470–8.
14. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007;99:715–26.
15. Cassidy A, Myles JP, van TM, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008;98:270–6.
16. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, 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.
17. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013; 368:728–36.
18. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, 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 2012;157:242–50.
19. Veronesi G, Maisonneuve P, Rampinelli C, Bertolotti R, Petrella F, Spaggiari L, et al. Computed tomography screening for lung cancer: results of ten years of annual screening and validation of cosmos prediction model. Lung Cancer 2013;82:426–30.

20. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74.

21. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Ostergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res 2012;5:834–46.

22. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European prospective investigation into cancer and nutrition. Int J Epidemiol 1997;26:S6–14.

23. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5:1113–24.

24. Boeing H, Wahrendorf J, Becker N. EPIC-Germany–A source for studies into diet and risk of chronic diseases. European investigation into cancer and nutrition. Ann Nutr Metab 1999;43:195–204.

25. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 2004;23:2109–23.

26. Rousson V, Zumbrunn T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. BMC Med Inform Decis Mak 2011;11:45.

27. Available from: www.mskcc.org/research/epidemiology-biostatistics/health-outcomes/decision-curve-analysis-0. 2014.

28. Leffondre K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. Am J Epidemiol 2002;156:813–23.

29. Lubin JH, Alavanja MC, Caporaso N, Brown LM, Brownson RC, Field RW, et al. Cigarette smoking and cancer risk: modeling total exposure and intensity. Am J Epidemiol 2007;166:479–89.

30. Vlaanderen J, Portengen L, Schuz J, Olsson A, Pesch B, Kendzia B, et al. Effect modification of the association of cumulative exposure and cancer risk by intensity of exposure and time since exposure cessation: a flexible method applied to cigarette smoking and lung cancer in the SYNERGY Study. Am J Epidemiol 2014;179:290–8.

31. Steyerberg EW, Vickers AJ. Decision curve analysis: a discussion. Med Decis Making 2008;28:146–9.

32. van Calster B, Vickers AJ. Calibration of Risk Prediction Models: Impact on Decision-Analytic Performance. Med Decis Making 2015;35:162–9.

33. Patz EF Jr, Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. JAMA Intern Med 2014;174:269–74.

34. Chien CR, Chen TH. Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. Int J Cancer 2008;122:2594–99.

35. Black WC, Gareen IF, Sneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. N Engl J Med 2014;371:1793–802.

36. Burns DM., Lee L, Shen LZ., Gilpin E, Tolly HD., Vaughn J, et al. Cigarette smoking behavior in the United States. In: Tobacco Control Monograph Series, No 8: *Changes in Cigarette-Related Disease Risks and Their Implication for Prevention and Control.* 2014. Available from: http://cancercontrol.cancer.gov/brp/tcrb/monographs/8/index.html.

37. Forey BA, Hamling J, Hamling J, Lee P. International smoking statistics (web edition): a collection of worldwide historical data, Australia. P N Lee Statistics and Computing Ltd. 2010. Available from: http://www.pnlee.co.uk/Downloads/ISS/ISS-Australia_061219.pdf

38. Wender R, Fontham ET, Barrera E Jr, Colditz GA, Church TR, Ettinger DS, et al. American Cancer Society lung cancer screening guidelines. CA Cancer J Clin 2013;63:107–17.

39. Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, MacMahon H, et al. Development of The American Association for Thoracic Surgery guidelines for low-dose computed tomography scans to screen for lung cancer in North America: recommendations of The American Association for Thoracic Surgery Task Force for lung cancer screening and surveillance. J Thorac Cardiovasc Surg 2012;144:25–32.

40. Wood DE, Eapen GA, Ettinger DS, Hou L, Jackman D, Kazerooni E, et al. Lung cancer screening. J Natl Compr Canc Netw 2012;10:240–65.

41. Moyer VA. U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2014;160:330–8.

Cancer Prevention Research

Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort

Kuanrong Li, Anika Hüsing, Disorn Sookthai, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1940-6207.CAPR-14-0424
Supplementary Material	Access the most recent supplemental material at: http://cancerpreventionresearch.aacrjournals.org/content/suppl/2015/06/16/1940-6207.CAPR-14-0424.DC1

Cited articles	This article cites 38 articles, 3 of which you can access for free at: http://cancerpreventionresearch.aacrjournals.org/content/8/9/777.full#ref-list-1
Citing articles	This article has been cited by 7 HighWire-hosted articles. Access the articles at: http://cancerpreventionresearch.aacrjournals.org/content/8/9/777.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerpreventionresearch.aacrjournals.org/content/8/9/777. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.