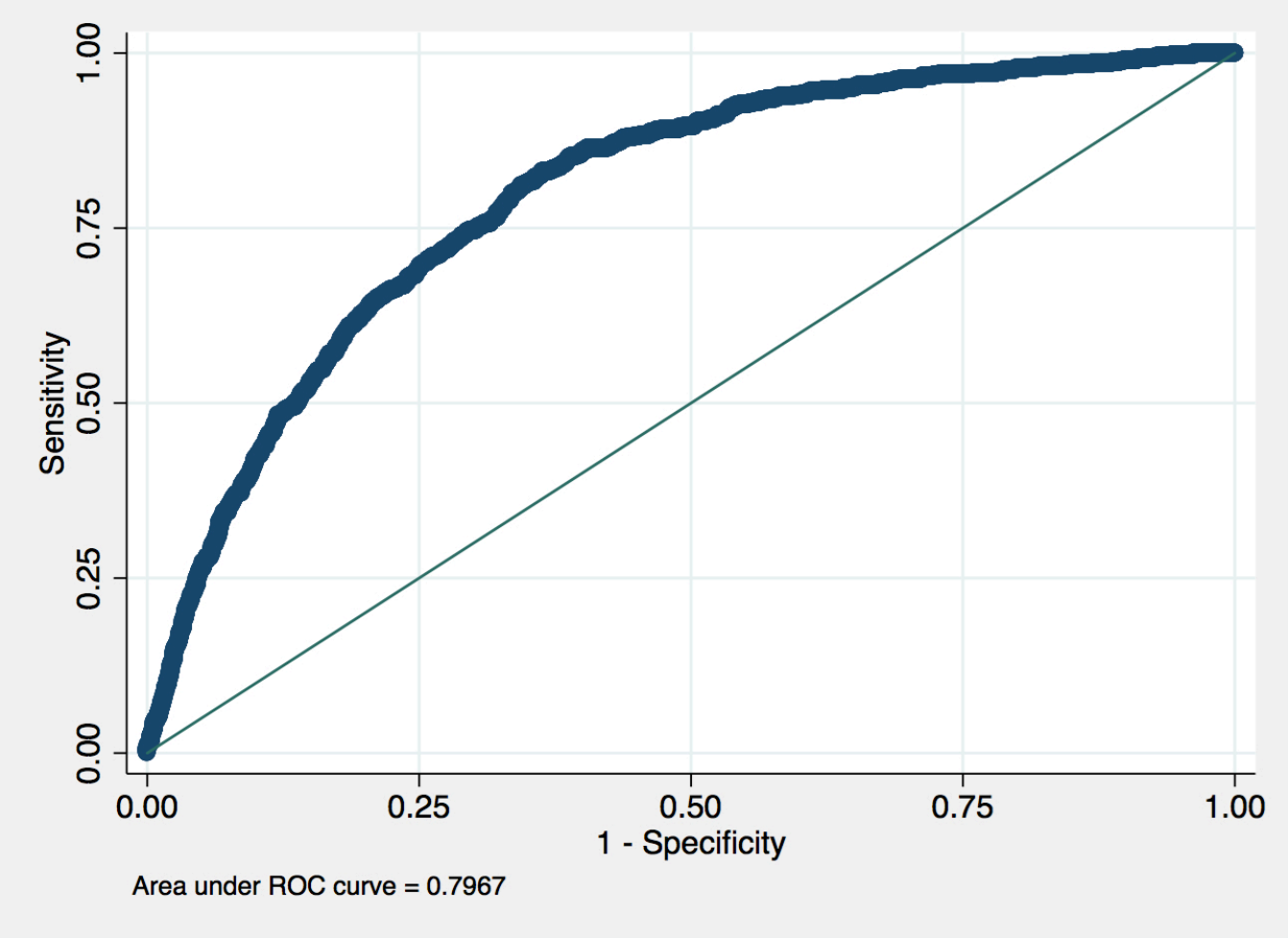
Supplementary Appendix

Supplement to: Tammemägi MC, Katki HA, Hocking WG, et al. *Selection Criteria for Lung-Cancer Screening*

Contents

Figure S1. Receiver operator characteristic curve for $PLCO_{M2012}$ in the external validation data (PLCO intervention arm)



Abbreviations: $PLCO_{M2012}$, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial updated 2012 Model; ROC, receiver operator characteristic.

Figure S2. $PLCO_{M2012}$ calibration – The mean absolute difference between observed and predicted lung cancer probabilities by model risk deciles

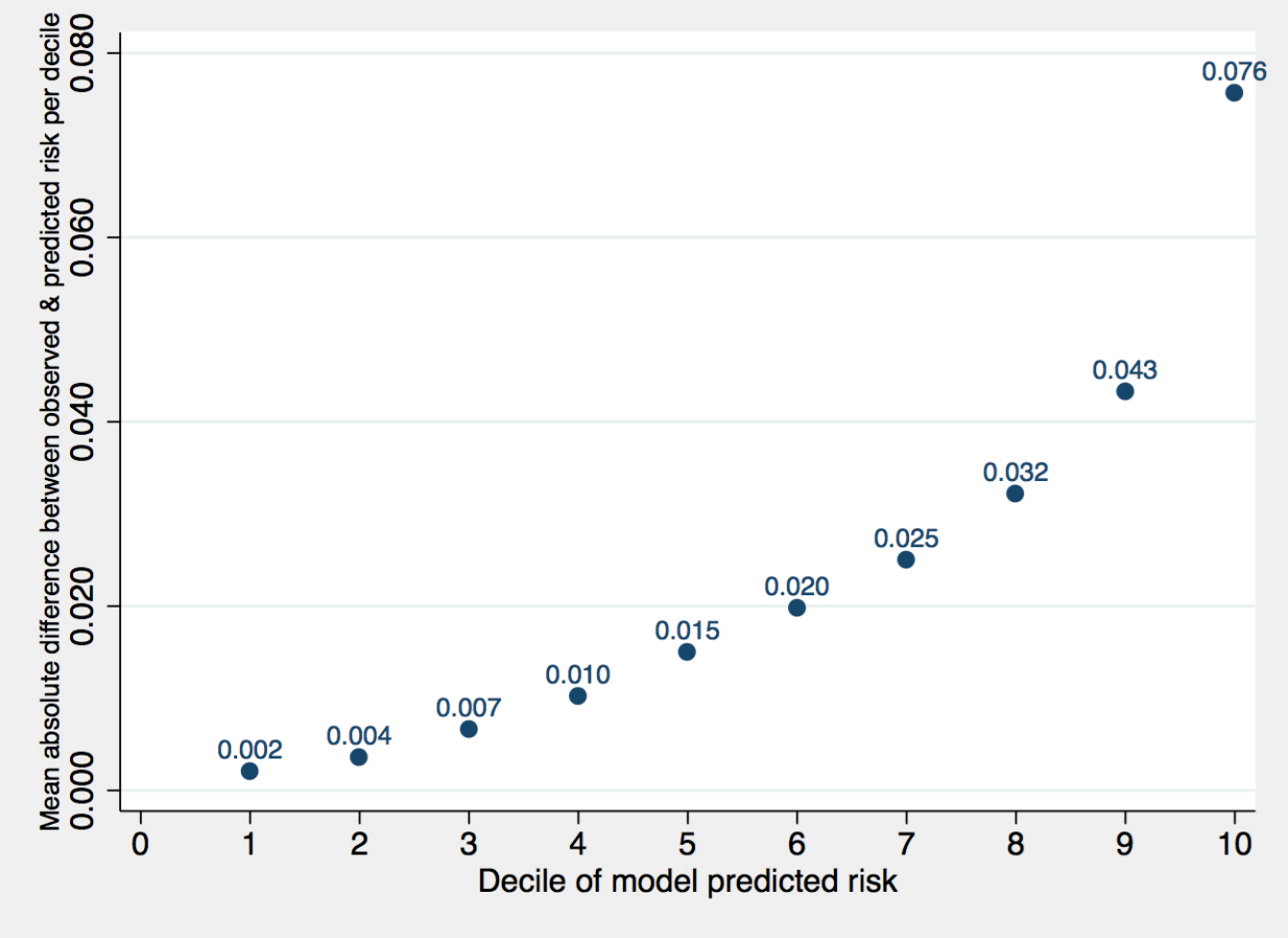


Table S1. Distribution of predictor variables in smokers in the PLCO control and intervention (CXR) arms, in the NLST (CXR & LDCT arms pooled) and in the PLCO intervention arm participants who met the NLST smoking criteria for study entry*, by lung cancer status†‡

	PLCO controls*			**PLCO CXR arm**			**NLST (LDCT & CXR arms)**			**PLCO CXR arm NLST criteria +ve)**		
Variables	No lung cancer N = 39,219 (98.2%)	Lung cancer N = 709 (1.8%)	Total N = 39,928 (100%)	No lung cancer N = 39,694 (98.1%)	Lung cancer N = 753 (1.9%)	Total N = 40,447 (100%)	No lung cancer N = 51,286 (96.4%)	Lung cancer N = 1,916 (3.6%)	Total N = 53,202 (100%)	No lung cancer N = 14,570 (96.5%)	Lung cancer N = 529 (3.5%)	Total N = 15,099 (100%)
Age (years)	62.4 {5.3}	64.6 {5.2}	62.5 {5.3} p < 0.0001	62.3 {5.3}	64.7 {5.2}	62.4 {5.3} p < 0.0001	61.3 {5.0}	63.7 {5.3}	61.4 {5.0} p < 0.0001	61.8 {5.1}	64.1 {5.1}	61.9 {5.1} p < 0.0001
Sex												
Female	16545 (98.4%)	271 (1.6%)		16544 (98.3%)	279 (1.7%)		21037 (96.5%)	774 (3.5%)		5712 (96.9%)	182 (3.1%)	
Male	22674 (98.1%)	438 (1.9%)	p = 0.035	23150 (98.0%)	474 (2.0%)	p = 0.01	30249 (96.4%)	1142 (3.6%)	p = 0.60	8858 (96.2%)	347 3.8%)	p = 0.03
Race/ethnicity												
White	34678 (98.2%)	630 (1.8%)		35066 (98.2%)	664 (1.8%)		46125 (96.4%)	1739 (3.6%)		13133 (96.6%)	457 (3.4%)	
Black	2181 (97.6%)	53 (2.4%)		2237 (96.8%)	73 (3.2%)		2233 (95.8%)	98 (4.2%)		734 (93.6%)	50 6.4%)	
Hispanic	806 (99.0%)	8 (1.0%)		799 (98.5%)	12 (1.5%)		650 (98.3%)	11 (1.7%)		231 (97.1%)	7 (2.9%)	
Asian	1185 (98.9%)	13 (1.1%)		1223 (98.8%)	15 (1.2%)		1063 97.1%)	32 (2.9%)		339 (97.1%)	10 (2.9%)	
Native American	241 (99.6%)	1 (0.4%)		218 (96.5%)	(3.5%)		180 (95.2%)	9 (4.8%)		75 (94.9%)	4 (5.1%)	
Pacific Islander	104 (96.3%)	4 (3.7%)	p = 0.01	127 99.2%)	1 (0.8%)	p < 0.001	188 (97.5%)	5 (2.6%)	p = 0.01	57 (98.3%)	1 (1.7%)	p = 0.003)
Education												
	3394 (97.3%)	96 (2.7%)		3410 (96.5%)	123 (3.5%)		3062 (94.7%)	172 (5.3%)		1448 (94.7%)	81 (5.3%)	
	8619 (97.8%)	197 (2.2%)		8751 (97.8%)	199 (2.2%)		12146 (95.7%)	545 (4.3%)		3531 (96.0%)	147 (4.0%)	
	5380 (98.2%)	98 (1.8%)		5317 (98.3%)	92 (1.7%)		7140 (96.2%)	285 (3.8%)		2078 (97.1%)	61 2.9%)	
	9111 (98.3%)	161 (1.7%)		9232 (98.2%)	165 (1.8%)		11837 (96.5%)	427 (3.5%)		3610 (96.7%)	123 (3.3%)	
	6363 (98.5%	94 (1.5%)		6603 (98.5%)	97 (1.5%)		8689 (97.2%)	247 (2.8%)		2093 (96.6%)	74 (3.4%)	
	6197 (99.1%)	59 (0.9%)	p < 0.001	6309 (98.8%)	76 (1.2%)	p < 0.001	7391 (97.3%)	203 (2.7%)	p < 0.001	1792 (977%)	42 (2.3%)	p < 0.001
Body mass index (kg/m^2)	27.4 {4.8}	26.5 {4.5}	27.4 {4.8} p < 0.0001	27.5 {4.9}	26.5 {4.3}	27.4 {4.9} p < 0.0001	28.0 {5.1}	26.9 {4.7}	28.0 {5.1} p < 0.0001	27.5 {5.0}	26.3 {4.3}	27.5 {4.9} p < 0.0001
Personal history of cancer												
No	37385 (98.3%)	653 (1.7%)	p < 0.001	37887 (98.2%)	708 (1.8%)	p = 0.06	49105 (96.5%)	1790 (3.5%)		13890 (96.5%)	502 (3.5%)	p = 0.60
Yes	1818 (97.0%)	56 (3.0%)		1797 (97.6%)	45 (2.4%)		2181 (94.5%)	126 (5.5%)	p < 0.001	678 (96.2%)	27 (3.8%)	
Family history of LCA				33569 (98.3%)	564 (1.7%)		39494 (96.6%)	1386 (3.4%)		12127 (96.9%)	393 (3.1%)	
No	33365 (98.4%)	541 (1.6%)	p < 0.001	4505 (97.1%)	139 (3.0%)	p < 0.001	11111 (95.7%)	497 (4.3%)	p < 0.001	1763 (94.7%)	99 (5.3%)	p < 0.001
Yes	4393 (97.1%)	130 (2.9%)										
COPD												
No	35388 (98.5%)	556 (1.5%)		36035 (98.4%)	596 (1.6%)		48770 (96.6%)	1724 (3.4%)		12452 (96.9%)	402 (3.1%)	
Yes	3451 (96.1%)	142 (3.9%)	p < 0.001	3442 (95.8%)	152 (4.2%)	p < 0.001	2497 (92.9%)	192 (7.1%)	p < 0.001	2051 (94.3%)	124 (5.7%)	p < 0.001
Smoking status												
Former	31600 (98.8%)	385 (1.2%)		31971 (98.7%)	422 (1.3%)		26826 (97.2%)	764 (2.8%)		8740 (97.2%)	250 (2.8%)	
Current	7600 (95.9%)	324 (4.1%)	p < 0.001	7695 (95.9%)	331 (4.1%)	p < 0.001	24460 (95.5%)	1152 (4.5%)	p < 0.001	5830 (95.4%)	279 (4.6%)	p < 0.001
Smoking intensity (cig/day)	24.8 {14.7}	30.2 {15.2}	24.9 {14.7} p < 0.0001	24.5 {14.4}	29.6 {15.4}	24.6 {14.5} p < 0.0001	28.4 {11.5}	29.6 {11.7}	28.4 {11.5} p < 0.0001	30.7 {13.4}	31.5 {14.0}	30.7 {13.6} p = 0.15
Smoking duration (years)	27.5 {13.8}	39.9 {10.5}	27.7 {13.8} p < 0.0001	27.3 {13.8}	40.1 {10.8}	27.6 {13.9} p < 0.0001	39.7 {7.3}	44.2 {7.0}	39.8 {7.3} p < 0.0001	39.5 {7.2}	44.1 {6.7}	39.7 {7.2} p < 0.0001
Smoking quit time in former smokers (years)	20.3 {12.1}	12.4 {10.5}	20.2 {12.1} p < 0.0001	20.4 {12.0}	12.7 {10.7}	20.3 {12.0} p < 0.0001	7.3 {4.8}	6.6 {4.8}	7.3 {4.8} p < 0.0001	7.6 {4.5}	6.2 {4.6}	7.5 {4.5} p < 0.0001
Predictive Performance												
Lung cancer incidence per 10,000 per year in former, current smokers			20.7, 72.4			22.5, 73.3			48.0, 79.9			48.8, 81.5
Mean lung cancer probability	0.0169	0.0453	0.0174 p < 0.0001	0.0167	0.0447	0.0172 p < 0.0001	0.0316	0.0519	0.0323 p < 0.0001	0.0333	0.0553	0.0340 p < 0.0001
AUC (95% CI)§	0.803 (0.788 to 0.817)			0.797 (0.782 to 0.813)			0.701 (0.689 to 0.712)			0.710 (0.689 to 0.732)		

Table S1. Continued.

Abbreviations: AUC, receiver operator characteristic area under the curve, COPD, chronic obstructive pulmonary disease; CXR, chest x-ray; LDCT, low-dose computed tomography; NLST, National Lung Screening Trial; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial, +ve, positive.

* NLST smoking criteria for study entry included 30 or more pack-years smoked and less than 15 years quit-time in former smokers.

† Lung cancer detected in 6 years of follow-up from randomization.

‡ For categorical variables, row percentages are presented in parentheses. For continuous variables, standard deviations are presented in pointed brackets. Reported p-values are for Fisher's exact test for dichotomous or unordered categorical variables, t-test for continuous variables, and nonparametric test of trend for ordinal data.

§ The 95% confidence intervals are bootstrap bias corrected with bootstrapping using 1000 re-samplings.

Table S2. Net Reclassification Improvement analysis of alternative models in the PLCO control ever-smokers

eTable 1a. Classification of PLCO ever-smoker controls by alternative models - Full model with smoking intensity as non-linear versus linear

PLCO control cases

	Full model with linear smoking intensity			
Full model with nonlinear smoking intensity	Low <1%	Intermediate ≥1 to <2%	High ≥2%	Totals
Low risk* <1%	54	17	0	71
Intermediate risk ≥1 to <2%	15	72	12	99
High risk ≥2%	0	23	437	460
Totals	69	112	449	630

Net Reclassification Improvement in cases = 1.4%

PLCO non-cases

	Full model with linear smoking intensity			
Full model with nonlinear smoking intensity	Low <1%	Intermediate ≥1 to <2%	High ≥2%	Totals
Low risk <1%	17753	1344	127	19224
Intermediate risk ≥1 to <2%	1020	5118	966	7104
High risk ≥2%	0	919	8639	9558
Totals	18773	7381	9732	35886

Net Reclassification Improvement in non-cases = 1.4%
Net Reclassification Improvement overall = 2.1% (p = 0.016)

eTable 1b. Classification of PLCO ever-smoker controls by alternative models – Full model versus model without personal history of cancer and race/ethnicity

PLCO control cases

	Full model without personal history of cancer and race/ethnicity			
Full model with personal history of cancer and race/ethnicity	Low <1%	Intermediate ≥1 to <2%	High ≥2%	Totals
Low risk* <1%	64	7	0	71
Intermediate risk ≥1 to <2%	4	85	10	99
High risk ≥2%	0	13	447	460
Totals	68	105	457	630

Net Reclassification Improvement in cases = 0%

PLCO non-cases

	Full model without personal history of cancer and race/ethnicity			
Full model with personal history of cancer and race/ethnicity	Low <1%	Intermediate ≥1 to <2%	High ≥2%	Totals
Low risk <1%	18622	591	11	19224
Intermediate risk ≥1 to <2%	375	6213	516	7104
High risk ≥2%	11	399	9148	9558
Totals	19008	7203	9675	35886

Net Reclassification Improvement in non-cases = 0.9%
Net Reclassification Improvement overall = 0.9% (p = 0.16)

Abbreviation: PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.
* Risk is the probability of developing lung cancer from study entry to end of six years of follow-up.

Table S3. Beta coefficients for comparable Cox and logistic regression model parameters developed in PLCO control ever-smokers (N = 36,286)

Variables	Cox model beta coefficients	Logistic model beta coefficient (Table 3 in text)
Age* (per one year)	.0792005	.0778868
Race/ethnicity		
White	Referent group	Referent group
Black	.4224289	.3944778
Hispanic	-.7186485	-.7434744
Asian	-.4673233	-.466585
American Indian or Alaskan Native	-33.97007	0
Native Hawaiian or Pacific Islander	.9443303	1.027152
Education* (per 7 levels)	-.0797203	-.0812744
Body mass index* (weight in kg/height in m^2)	-.0277547	-.0274194
Chronic obstructive pulmonary disease (yes vs. no)	.3849946	.3553063
Personal history of cancer (yes vs. no)	.4408969	.4589971
Family history of lung cancer (yes vs. no)	.5669498	.587185
Smoking status (current vs. former)	.2499022	.2597431
Smoking intensity† (average cigarettes/day)	-1.808796†	-1.822606†
Smoking duration* (per year)	.0317864	.0317321
Smoking quit time* (per year)	-.0306086	-.0308572
Model constant	NA	-4.532506

Abbreviations: CI, confidence interval; kg, kilograms; m, meter; NLST, National Lung Screening Trial, PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

* Age is centered on 62 years, education is centered on level 4, body mass index is centered on 27, smoking duration was centered on 27 years, and smoking quit time was centered on 10 years.

† Smoking intensity (average number of cigarettes smoked per day) had a nonlinear association with lung cancer and this variable was transformed. For this reason, the odds ratio is not directly interpretable in a meaningful fashion. The variable *smoking intensity* was first scaled by dividing by 10, this term was then exponentiated by the power -1, and this term was then centered by subtracting 0.4021657707 in the Cox model and by subtracting 0.4021541613 in the logistic regression model.