

Supplementary material: External validation and Re-Calibration of the Brock Model to Predict Probability of Cancer in Pulmonary Nodules using NLST Data

Supplementary Figures legends

- **Figure 1S** – Description of the Brock model version 2b, “full model, with spiculation” [1]. We can interpret, for example, that the OR for a change of 1 unit (i.e. 1 year) in *age* is $e^{0.0287} = 1.03$, representing about a 3% increase in the odds of being diagnosed with lung cancer. Similarly, for participants with a spiculated nodule, the odds of being diagnosed was 2.17 times greater than the odds for those without spiculation (with all other covariates being constant).

* A back-transformation has to be applied to the odds ratio for nodule size. According to Elswick et al [2] the odds ratio for a change of 1 unit in nodule size is $\exp \left[\hat{\beta} \left\{ \left(\frac{\text{nodule size} + 1}{10} \right)^{-0.5} - \left(\frac{\text{nodule size}}{10} \right)^{-0.5} \right\} \right]$ with a confidential interval $\exp \left[\hat{\beta} \left\{ \left(\frac{\text{nodule size} + 1}{10} \right)^{-0.5} - \left(\frac{\text{nodule size}}{10} \right)^{-0.5} \right\} \pm z_{1-\alpha/2} \left\{ \left(\frac{\text{nodule size} + 1}{10} \right)^{-0.5} - \left(\frac{\text{nodule size}}{10} \right)^{-0.5} \right\} \widehat{SE}(\hat{\beta}) \right]$ with $\hat{\beta} = -5.385$ and $\widehat{SE}(\hat{\beta})$ the estimated standard error of $\hat{\beta}$.

- **Figure S2** - Sensitivity and specificity of the original model in the NLST dataset when the decision threshold is set to 0.06 [3].

Table S1 - Relation between covariates from Brock model [1] and those from the NLST database [4], using "Participants" and "SCT abnormalities" datasets.

	Covariates	Covariates from NLST dataset
Covariates from Brock model	Age	Participant.age
	Sex	Participant.gender
	Family history of lung cancer	Participant.famfather, Participant.fammother, Participant.fambrother, Participant.famsister, Participant.famchild
	Emphysema	Sct_abnormalities.sct_ab_desc = 59 ("Emphysema")
	Nodule size (mm)	Sct_abnormalities.SCT_LONG_DIA
	Nodule type	Sct_abnormalities.SCT_PRE_ATT
	Upper lobe	Sct_abnormalities.SCT_EPI_LOC
	Spiculation	Sct_abnormalities.SCT_MARGINS
	Nodule count	Count of Sct_abnormalities.sc_ab_desc = 51 by patient
Covariates used to create the event covariate	Location of the nodule	Sct_abnormalities.sct_epi_loc
	Location of the cancer	Participant.loclin, Participant.locflow, Participant.locrup, Participant.locrflow, Participant.locrmid, Participant.locrup
	Results of the 2 nd and 3 rd screening	Participant.scr_iso, Participant.scr_iso2
		- 2: "Negative screen, minor abnormalities not suspicious for lung cancer"
		- 3: "Negative screen, significant abnormalities not suspicious for lung cancer"
		- 4: "Positive, Change Unspecified, nodule(s) \geq 4 mm or enlarging nodule(s), mass(es), other non-specific abnormalities suspicious for lung cancer"
		- 13: "Not Expected - Cancer before screening window"
		- 14: "Not Expected - Death before screening window"
		- 15: "Not Compliant - Refused a screen"
		- 23: "Not Expected - Cancer in screening window"
	- 24: "Not Expected - Death in screening window"	
	Days from randomization to first diagnosis of lung cancer	Participant.candx_days
	Days from randomization to date last known alive	Participant.fup_days
	Days since randomization at time of screening for the first exam	Participant.scr_days0
	Study year of screen	Sct_abnormalities.study_yr
	Abnormality description	Sct_abnormalities.sc_ab_desc (if = 51: Non-calcified nodule or mass (opacity \geq 4mm diameter))

Smoking status at randomization	Participant. <i>cigsmok</i>
Pack-year calculated as: (Total Years Smoked x Cigarettes Per Day / 20)	Participant. <i>pkyr</i>
Asbestos: Ever worked with for 1 or more years?	Participant. <i>wrkasbe</i>

Table S2 - Distribution of nodules covariates (both existing covariates from the Brock model and newly identified covariates), according to cancer status, and comparison in the NLST dataset. Mean and standard deviation are reported for quantitative covariates; number and percentage are reported for qualitative covariates. Student test or Wilcoxon-Mann-Whitney test or χ^2 test used when appropriate.

Nodules Covariates	Malignant nodules (N=222)	Benign nodules (N=7629)	Total (N=7851)	P-values
Nodule size (mm)	18.45 (13.89)	6.56 (4.47)	6.89 (5.36)	< 0.01
Nodule type				< 0.01
solid	195 (87.84 %)	6993 (91.66 %)	7188 (91.56 %)	
non-solid	10 (4.5 %)	475 (6.23 %)	485 (6.18 %)	
part-solid	17 (7.66 %)	161 (2.11 %)	178 (2.27 %)	
Nodule location				< 0.01
Right upper lobe	90 (40.54 %)	1606 (21.05 %)	1696 (21.60 %)	
Right middle lobe	11 (4.95 %)	1174 (15.39 %)	1185 (15.09 %)	
Right lower lobe	39 (17.57 %)	1920 (25.17 %)	1959 (24.95 %)	
Left upper lobe	56 (25.23 %)	949 (12.44 %)	1005 (12.80 %)	
Lingula	0 (0 %)	339 (4.43 %)	339 (4.32 %)	
Left lower lobe	26 (11.71 %)	1595 (20.91 %)	1621 (20.65 %)	
Other	0 (0 %)	46 (0.60 %)	46 (0.59 %)	
Nodule count at baseline	1.87 (1.1)	2.46 (1.65)	2.44 (1.64)	< 0.01
Spiculation	159 (71.62 %)	871 (11.42 %)	1030 (13.12 %)	< 0.01
Brock score	0.32 (0.25)	0.03 (0.08)	0.04 (0.10)	< 0.01
Participants Covariates	Cancer (N= 194)	Benign (N= 4804)	Total (N= 4998)	P-values
Age (years)	63.74 (5.35)	61.95 (5.10)	62.02 (5.12)	< 0.01
Gender				1
Female	75 (38.66 %)	1862(38.76 %)	1937 (38.76 %)	
Male	119 (61.34 %)	2942 (61.24 %)	3061 (61.24 %)	
Family history of lung cancer	58 (29.90 %)	1090 (22.69 %)	1148 (22.97 %)	0.02
Emphysema	83 (42.78 %)	1569 (32.66 %)	1652 (33.05 %)	< 0.01
BMI	26.68 (5.09)	27.68 (4.78)	27.64 (4.8)	< 0.01
Smoking status at randomization				0.19
Former	90 (46.39 %)	2472 (51.46 %)	2562 (51.26 %)	
Current	104 (53.61 %)	2332 (48.54 %)	2436 (48.74 %)	
Pack years	69.81 (33.68)	56.82 (24.47)	57.32 (25.01)	< 0.01
Asbestos	7 (3.61 %)	215 (4.48 %)	222 (4.44 %)	0.69

Table S3 – Examples of threshold for the original and the re-calibrated models in the NLST dataset and their impact on sensitivity, specificity, negative predictive value and positive predictive value. Confidence intervals have been calculated according to [5].

Models	Measures	Thresholds				
		3.4%	6%	6.5%	10%	30%
Original	Sensitivity (%)	85.59 [80.97-90.21]	80.18 [74.94-85.42]	80.18 [74.94-85.42]	74.32 [68.58-80.07]	48.65 [42.07-55.22]
	Specificity (%)	82.49 [81.64-83.34]	88.74 [88.03-89.45]	89.63 [88.95-90.31]	92.52 [91.93-93.11]	97.57 [97.23-97.92]
	Negative predictive value (%)	99.50 [99.34-99.66]	99.36 [99.19-99.53]	99.36 [99.2-99.53]	99.20 [99.02-99.38]	98.50 [98.31-98.69]
	Positive predictive value (%)	12.41 [11.62-13.2]	17.12 [15.83-18.4]	18.31 [16.92-19.7]	22.36 [20.44-24.27]	36.73 [32.18-41.29]
Re-calibrated	Sensitivity (%)	82.43 [77.43-87.44]	75.68 [70.03-81.32]	74.32 [68.58-80.07]	69.37 [63.31-75.43]	36.94 [30.59-43.29]
	Specificity (%)	88.49 [87.78-89.21]	92.55 [91.97-93.14]	92.90 [92.32-93.47]	94.98 [94.49-95.47]	98.68 [98.42-98.93]
	Negative predictive value (%)	99.43 [99.26-99.59]	99.24 [99.07-99.42]	99.20 [99.02-99.38]	99.07 [98.89-99.25]	98.17 [97.99-98.35]
	Positive predictive value (%)	17.25 [16.01-18.49]	22.83 [20.91-24.74]	23.34 [21.33-25.34]	28.68 [26-31.36]	44.81 [38.4-51.21]

Covariates	β coefficients	Odds ratio = exp(β) (95% Confidence Interval)	P-values
Intercept	-6.7892		
Age (year)	0.0287	1.03 [0.99-1.07]	0.16
Sex, female vs male	0.6011	1.82 [1.12-2.97]	0.02
Family history of lung cancer, yes vs no	0.2961	1.34 [0.83-2.17]	0.23
Emphysema, yes vs no	0.2953	1.34 [0.78-2.33]	0.29
Nodule size (mm)	-5.3854	*	< 0.001
Nodule type:			
Solid	Reference		-
Nonsolid or with ground-glass opacity	-0.1276	0.88 [0.48-1.62]	0.68
Part-solid	0.377	1.46 [0.74-2.88]	0.28
Nodule location, upper vs middle or lower lobe	0.6581	1.93 [1.14-3.27]	0.02
Nodule count	-0.0824	0.92 [0.85-1]	0.049
Spiculation, yes vs no	0.7729	2.17 [1.16-4.05]	0.02

Figure 1S - Description of the Brock model version 2b, “full model, with spiculation” [1]. We can interpret, for example, that the OR for a change of 1 unit (i.e. 1 year) in *age* is $e^{0.0287} = 1.03$, representing about a 3% increase in the odds of being diagnosed with lung cancer. Similarly, for participants with a spiculated nodule, the odds of being diagnosed was 2.17 times greater than the odds for those without spiculation (with all other covariates being constant).

* A back-transformation has to be applied to the odds ratio for nodule size. According to Elswick et al [2] the odds ratio for a change of 1 unit in nodule size is $\exp \left[\hat{\beta} \left\{ \left(\frac{\text{nodule size} + 1}{10} \right)^{-0.5} - \left(\frac{\text{nodule size}}{10} \right)^{-0.5} \right\} \right]$ with a confidential interval $\exp \left[\hat{\beta} \left\{ \left(\frac{\text{nodule size} + 1}{10} \right)^{-0.5} - \left(\frac{\text{nodule size}}{10} \right)^{-0.5} \right\} \pm z_{1-\alpha/2} \left\{ \left(\frac{\text{nodule size} + 1}{10} \right)^{-0.5} - \left(\frac{\text{nodule size}}{10} \right)^{-0.5} \right\} \widehat{SE}(\hat{\beta}) \right]$ with $\hat{\beta} = -5.385$ and $\widehat{SE}(\hat{\beta})$ the estimated standard error of $\hat{\beta}$.

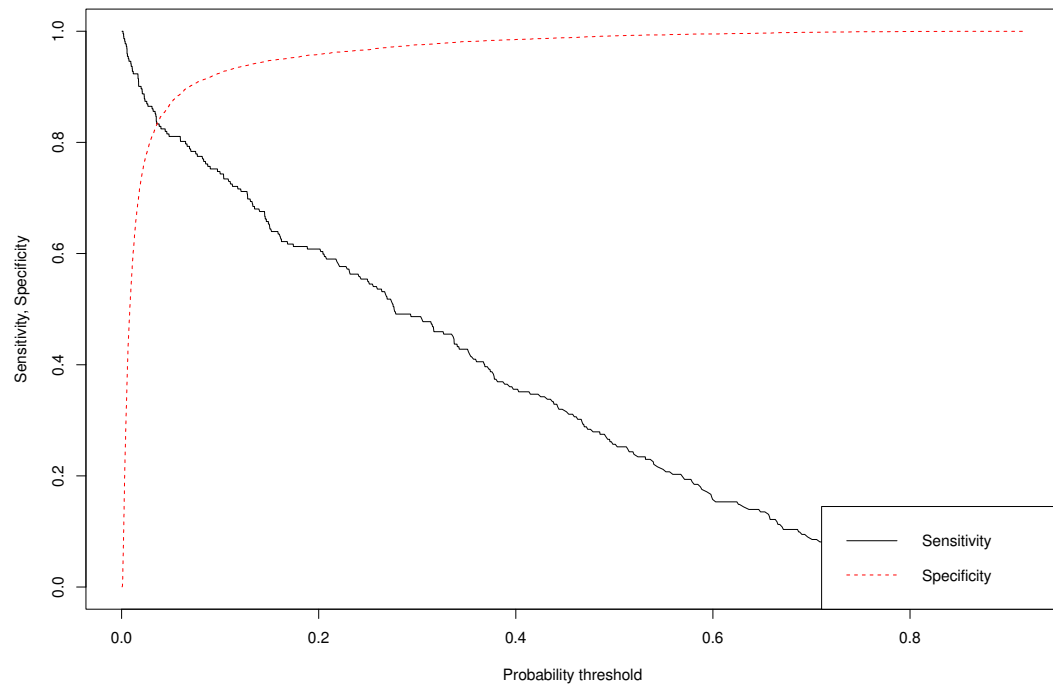


Figure S2 - Sensitivity and specificity of the original model in the NLST dataset when the decision threshold is set to 0.06 [3].

References:

- 1 McWilliams A, Tammemagi MC, Mayo JR, *et al.* Probability of cancer in pulmonary nodules detected on first screening CT. *New England Journal of Medicine* 2013;**369**:910–919.
- 2 Elswick RK, Schwartz PF, Welsh JA. Interpretation of the odds ratio from logistic regression after a transformation of the covariate vector. *Statistics in Medicine* 1997;**16**:1695–703. doi:10.1002/(SICI)1097-0258(19970815)16:15<1695::AID-SIM601>3.0.CO;2-V
- 3 Youden WJ. Index for rating diagnostic tests. *Cancer*;3:32–5. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
- 4 The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine* 2011;**365**:395–409. doi:10.1056/NEJMoa1102873
- 5 Mercaldo ND, Zhou X-H. Confidence Intervals for Predictive Values Using Data from a Case Control Study. ;:15.