

Predicting Malignancy Risk of Screen-Detected Lung Nodules—Mean Diameter or Volume



Martin Tammemagi, PhD,^a Alex J. Ritchie, MD,^b Sukhinder Atkar-Khattra, BSc,^c Brendan Dougherty, MD,^d Calvin Sanghera, BSc,^c John R. Mayo, MD,^e Ren Yuan, MD,^c Daria Manos, MD,^f Annette M. McWilliams, MD,^g Heidi Schmidt, MD,^h Michel Gingras, MD,ⁱ Sergio Pasian, MD,ⁱ Lori Stewart, MD,^j Scott Tsai, MD,^j Jean M. Seely, MD,^k Paul Burrowes, MD,^l Rick Bhatia, MD,^m Ehsan A. Haider, MD,^j Colm Boylan, MD,^j Colin Jacobs, PhD,ⁿ Bram van Ginneken, PhD,ⁿ Ming-Sound Tsao, MD,^h Stephen Lam, MD,^{c,*} and the Pan-Canadian Early Detection of Lung Cancer Study Group

^aDepartment of Health Sciences, Brock University, St. Catharines, Ontario, Canada

^bRoyal Brisbane and Women's Hospital, Brisbane, Australia

^cBritish Columbia Cancer, Vancouver, British Columbia, Canada

^dFlinders Medical Centre, Adelaide, Australia

^eDepartment of Radiology, Vancouver Coastal Health, Vancouver, British Columbia, Canada

^fDalhousie University, Halifax, Nova Scotia, Canada

^gFiona Stanley Hospital and Sir Charles Gairdner Hospital, Perth, Western Australia, Australia

^hUniversity Health Network and Princess Margaret Cancer Centre, Toronto, Ontario, Canada

ⁱUniversity Institute of Cardiology and Pneumology of Quebec, Quebec, Canada

^jDepartment of Diagnostic Imaging, Henderson Hospital, Hamilton, Ontario, Canada

^kOttawa Hospital Research Institute and the University of Ottawa, Ottawa, Ontario, Canada

^lUniversity of Calgary, Foothills Medical Centre, Calgary, Alberta, Canada

^mMemorial University, St. John's, Newfoundland, Canada

ⁿRadboud University Medical Center, Nijmegen, The Netherlands

Received 30 July 2018; revised 7 October 2018; accepted 9 October 2018

Available online - 24 October 2018

ABSTRACT

Objective: In lung cancer screening practice low-dose computed tomography, diameter, and volumetric measurement have been used in the management of screen-detected lung nodules. The aim of this study was to compare the performance of nodule malignancy risk prediction tools using diameter or volume and between computer-aided detection (CAD) and radiologist measurements.

Methods: Multivariable logistic regression models were prepared by using data from two multicenter lung cancer screening trials. For model development and validation, baseline low-dose computed tomography scans from the Pan-Canadian Early Detection of Lung Cancer Study and a subset of National Lung Screening Trial (NLST) scans with lung nodules 3 mm or more in mean diameter were analyzed by using the CIRRUS Lung Screening Workstation (Radboud University Medical Center, Nijmegen, the Netherlands). In the NLST sample, nodules with cancer had been matched on the basis of size to nodules without cancer.

Results: Both CAD-based mean diameter and volume models showed excellent discrimination and calibration, with similar areas under the receiver operating characteristic curves of 0.947. The two CAD models had predictive performance similar to that of the radiologist-based model.

*Corresponding author.

Disclosure: Dr. Jacobs reports grants and a research grant and royalties for development of the lung screening workstation Veolity from MeVis Medical Solutions AG, Bremen, Germany, outside the submitted work. Dr. Tsao reports personal fees from BMS, AstraZeneca, Pfizer, Merck, and Abbvie outside the submitted work. The remaining authors declare no conflict of interest.

Trial Registration: Pan-Canadian Early Detection of Lung Cancer Study (NCT00751660) and National Lung Screening Trial (NCT00047385).

Address for correspondence: Stephen Lam, MD, British Columbia Cancer Agency, 675 West 10th Ave., Vancouver, BC, Canada V5Z 1L1. E-mail: slam@bccancer.bc.ca

© 2018 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.

ISSN: 1556-0864

<https://doi.org/10.1016/j.jtho.2018.10.006>

In the NLST validation data, the CAD mean diameter and volume models also demonstrated excellent discrimination: areas under the curve of 0.810 and 0.821, respectively. These performance statistics are similar to those of the Pan-Canadian Early Detection of Lung Cancer Study malignancy probability model with use of these data and radiologist-measured maximum diameter.

Conclusion: Either CAD-based nodule diameter or volume can be used to assist in predicting a nodule's malignancy risk.

© 2018 International Association for the Study of Lung Cancer. Published by Elsevier Inc. All rights reserved.

Keywords: Screening; Lung nodule malignancy risk; Lung nodule malignancy probability; Diameter or volumetric measurement

Introduction

Following recommendations by the U.S. Preventive Services Task and Centers for Medicare and Medicaid Services and the Canadian Task Force on Preventive Health Care that support use of low-dose computed tomography (LDCT) of the chest to decrease lung cancer mortality,¹⁻³ LDCT screening is being implemented in the United States and Canada. One concern with population screening is the large number of lung nodules that are identified by LDCT, most of which are benign. A number of lung nodule management guidelines and nodule malignancy risk prediction tools have been developed to guide management. When two or more scans are available, the management of lung nodules is straightforward because nodule growth is the most important indicator of malignancy. The first screening LDCT is more challenging because it has important implication in the rate of early-recall imaging studies for intermediate or suspicious nodules and the frequency of the next screening LDCT, such as an annual or biennial repeat. The PanCan nodule malignancy probability model (PanCan model [synonyms are the McWilliams model, Brock model, and Tammemagi model]) was developed on the basis of nodules detected and followed from the first screen in the prospective Pan-Canadian Early Detection of Lung Cancer Study (PanCan study).⁴ This model was initially validated externally with use of British Columbia Cancer screening data, and it has subsequently been validated in multiple studies in different settings by different research groups.⁵⁻⁹ The PanCan risk calculator has been recommended for use in some settings by the British Thoracic Society Guidelines for the Investigation and Management of Nodules and the American College of Radiology's Lung Imaging Reporting and Data System.¹⁰⁻¹²

In the PanCan model, the nodule size used for risk prediction was the maximum length as measured manually by radiologists. Published guidelines use two-dimensional perpendicular mean diameter¹² or volumetric measurement¹³ or a combination of both.^{10,11} Volumetric measurement uses computer-aided detection (CAD)¹³ tools whereas measurement of mean diameter can also be performed manually with electronic calipers. CAD-based measurements reduce inter-observer variation but require dedicated software tools. The aim of this study was to compare the performance of nodule malignancy risk prediction tools using diameter or volume and between and radiologist measurements.

Methods

The current study uses data from the PanCan study, which has been described elsewhere.¹⁴ In brief, the PanCan study was a single-arm prospective lung cancer screening study in which 2537 individuals at high risk for lung cancer were screened for lung cancer using LDCT in eight centers across Canada. Entry into the study was based on having a 2% or higher 6-year risk according to a prototype of the Prostate, Lung and Ovarian Cancer Screening Trial model 2012 risk prediction model.¹⁵ In the current study, baseline PanCan study scans that had radiologist-detected nodules were reread using the CIRRUS CAD system (CIRRUS Lung Screening, Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands). The list of malignant lung nodules reported previously⁴ was updated.

To validate the models, we used CAD read data from a subset of NLST scans. The data set was obtained from the NLST repository. The sampling of the NLST data was driven by alternative study questions to measure airways and lung parenchyma features associated with lung cancer or no lung cancer. The samples included three categories: (1) LDCT scans from all participants in whom lung cancer had been diagnosed within 5 years of the baseline scan on the basis of size of the largest recorded lung nodule (if present); (2) baseline LDCT scans from participants with no lung cancer who were matched by year of entry, study site, smoking status (current versus former smokers), and nodule closest in size; and (c) prior LDCT scans from individuals with interval lung cancer.

Semiautomated Reading of Screening LDCT Scans

A research workstation specifically designed to read screening LDCT scans (CIRRUS Lung Screening, Diagnostic Image Analysis Group) was used. CIRRUS Lung

Screening integrates three CAD software packages that identify solid and subsolid lung nodules. The performance of the CIRRUS Lung Screening software has been previously reported.^{16–18} The lung nodules were automatically segmented. The radiologist checked the results and edited the segmentation if needed. False-positive nodules were removed. False-negative nodules 4 mm or larger were added by the human observer for segmentation by CIRRUS. Calcified or noncalcified lung nodules 3 mm or smaller were excluded from analysis. Perifissural nodules defined as small solid nodules adjacent to pleural fissures that are thought to represent intrapulmonary lymph nodes were also excluded from analysis in keeping with the Fleischner Society guideline and other published studies.^{4,19–21} The radiologists who read the scans made the decision.

Statistical Methods

Descriptive statistics were prepared by using contingency table analysis for categorical data and Fisher's exact test. The 95% confidence intervals (CIs) for proportions were estimated by using the binomial exact method. Comparisons of nonnormal and normal continuous data were made by using the nonparametric test of trend²² and Student's *t* test, respectively.

Multivariable logistic regression models evaluated the same set of predictors as were included in the original PanCan study model.⁴ The outcome was lung cancer, and the radiological predictor characteristics were CAD-derived. Variables with effect estimates that approached the null and did not improve prediction were excluded from the final models. Predictor variables with effect estimates that were in the same direction and of a magnitude similar to those observed in the original model were retained in the CAD models even when they were not statistically significant.²³ The nodule two-dimensional perpendicular mean diameter and volume were right-skewed and were natural log-transformed. Nonlinear effects of continuous variables were evaluated by using locally weighted scatterplot smoothing plots and multivariable fractional polynomials.²⁴ Because some individuals had multiple nodules, the variances of effect estimates were adjusted for clustering of data within individuals by using the Huber and White robust "sandwich" variance estimator.²⁵

A model's overall predictive performance was assessed by using the Brier score.²⁶ Discrimination (ability to classify correctly) was measured by using the receiver operator characteristic area under the curve (AUC). Calibration (whether the model-predicted probabilities matched the observed probabilities) was assessed by using the Spiegelhalter *z* statistic *p* value (with significance indicating poor calibration) and by comparing

the mean and observed overall probabilities of outcomes. Brier scores and AUCs have been presented with bootstrap bias-corrected 95% CIs with bootstrapping using 1000 resamplings.²⁷ The predictive performances of the CAD mean diameter and volume models were compared with that of the original PanCan model in the current development and validation samples.

All reported *p* values are two sided, unless otherwise indicated. Statistics and figures were prepared by using Stata 15.1 MP software (StataCorp, College Station, TX).

Results

The characteristics of the study participants and nodules in the development and validation samples are described in [Table 1](#).

In the PanCan data included in the analysis, there were 1711 individuals who had 5779 nodules and 117 lung cancers (2.0% of nodules). The PanCan CAD mean size diameter and volume models are described in [Table 2](#). Both models include sex, family history of lung cancer, emphysema, nodule type, spiculation, nodule location, nodule count, and nodule size. Sex and family history of lung cancer and emphysema were not statistically significant, but they had effect estimates in a direction and magnitude consistent with those of the original PanCan model. Age, which is a predictor present in the original PanCan model, was excluded from the current models because its effect estimate approached the null value and it did not contribute to prediction. In the original PanCan model, pure nonsolid (ground glass) nodules carried a reduced risk compared with solid nodules. In the current data, the risk of ground glass nodules was found to not differ substantially from that of solid nodules and they were pooled together. As was observed in our original PanCan model, the relationship between nodule size and lung cancer probability is nonlinear. [Figure 1](#) describes the nonlinear relationships between mean diameter and volume and lung cancer risk according to the two PanCan CAD models.

In the PanCan study data, both PanCan CAD models demonstrated excellent overall prediction (Brier scores of 0.015 for both), discrimination (AUC = 0.947 for both mean diameter and volume models [see [Table 2](#) and [Supplementary Fig. 1](#)]), and calibration (Spiegelhalter *p* > 0.35 for both).

In the NLST data set there were 6009 nodules in 3680 individuals; 460 nodules (7.7%) were lung cancer. In the external validation in NLST data, the PanCan CAD mean diameter and volume models demonstrated excellent discrimination (AUC = 0.810 and 0.821, respectively).

When the models were restricted to 8- to 15-mm nodules, the overall predictive performance, including discrimination and calibration of the PanCan CAD mean

Table 1. Characteristics of Individuals and Nodules in the Analytic Development PanCan Study Sample and in the Validation NLST Subsample, Stratified by Lung Cancer Status

Data	No Lung Cancer	Lung Cancer	p Value	Total
PanCan data				
Person level	n = 1600	n = 111		N = 1711
Age, y	62.5 (SD 5.8)	63.4 (SD 5.7)	0.098	62.5 (SD 5.8)
Sex				
Female	741 (46.3%)	61 (54.9%)		802 (46.9%)
Male	859 (53.7%)	50 (45.1%)	0.094	909 (53.1%)
Family history of lung cancer				
No	1069 (66.8%)	69 (62.2%)		1138 (66.5%)
Yes	531 (33.2%)	42 (37.8%)	0.349	573 (33.5%)
Emphysema				
No	653 (40.8%)	29 (26.1%)		682 (39.9%)
Yes	947 (59.2%)	82 (73.9%)	0.002	1029 (60.1%)
Pack-years smoked	54.9 (SD 24.0)	54.8 (SD 23.4)	0.967	54.9 (SD 24.0)
Nodule level	n = 5662	n = 117		N = 5779
Spiculation				
No	5517 (97.4%)	79 (67.5%)		5596 (96.8%)
Yes	145 (2.6%)	38 (32.5%)	0.001	183 (3.2%)
Upper lobe location				
No	3021 (52.4%)	44 (37.6%)		3065 (53.0%)
Yes	2641 (46.6%)	73 (62.4%)	0.001	2714 (47.0%)
Nodule type				
Nonsolid	983 (17.4%)	26 (22.2%)		1009 (17.5%)
Semisolid	278 (4.9%)	33 (28.2%)		311 (5.4%)
Solid	4401 (77.7%)	58 (49.6%)	0.001	4459 (77.2%)
Nodule count, median	5 (IQR 3-8)	4 (IQR 3-8)	0.283	5 (IQR 3-8)
Nodule diameter, mm, mean	5.4 (SD 3.2)	16.8 (SD 10.2)	<0.001	5.6 (SD 3.8)
Nodule volume, mm, mean	169.0 (SD 1261.3)	4669.7 (SD 13304.6)	<0.001	260.2 (SD 2348.2)
NLST data				
Person level	n = 3239	n = 441		N = 3680
Age	62.1 (SD 5.2)	63.7 (SD 5.3)	0.001	62.3 (SD 5.3)
Sex				
Female	1426 (44.0.7%)	180 (40.8%)		1606 (43.6%)
Male	1813 (56.0%)	261 (59.2%)	0.219	2074 (56.4%)
Family history of lung cancer				
No	2514 (77.6%)	324 (73.5%)		2838 (77.1%)
Yes	725 (22.4%)	117 (26.5%)	0.053	842 (22.9%)
Emphysema				
No	2131 (65.8%)	253 (57.4%)		2384 (64.8%)
Yes	1108 (34.2%)	188 (42.6%)	0.001	1296 (35.2%)
Nodule level	n = 5549	n = 460		N = 6009
Spiculation				
No	4973 (89.6%)	286 (62.2%)		5259 (87.5%)
Yes	576 (10.4%)	174 (37.8%)	0.001	750 (12.5%)
Upper lobe location				
No	3205 (57.8%)	183 (39.8%)		3388 (56.4%)
Yes	2344 (42.2%)	277 (60.2%)	0.001	2621 (43.6%)
Nodule type				
Nonsolid	1451 (26.2%)	24 (5.2%)		1475 (24.5%)
Semisolid	524 (9.4%)	109 (23.7%)		633 (10.5%)
Solid	3574 (64.4%)	327 (71.1%)	0.001	3901 (64.9%)
Nodule count (median)	2 (IQR 1-3)	1 (IQR 1-1)	0.001	2 (IQR 1-3)
Nodule diameter, mm (mean)	7.9 (SD 5.5)	16.8 (SD 12.1)	0.001	8.6 (SD 6.6)
Nodule volume (mean)	605.6 (SD 3031.0)	6289.4.7 (SD 18214.3)	0.001	1040.7 (SD 6009.4)

IQR, interquartile range; NLST, National Lung Screening Trial; PanCan, Pan-Canadian Early Detection of Lung Cancer Study.

Table 2. Parameter Estimates for the PanCan Model Based on CAD Reading of Scans and Nodule Size Estimated by Mean Diameter or Volume

Predictor	Mean Diameter Model ^a		Volume Model ^b	
	OR (95% CI), p Value	Beta Coefficient	OR (95% CI), p Value	Beta Coefficient
Sex (F vs. M)	1.45 (0.95-2.3), p = 0.09	.3749324	1.44 (0.93-2.22), p = 0.10	.3641748
Family history lung cancer	1.38 (0.89-2.13), p = 0.15	.3203023	1.35 (0.87-2.10), p = 0.19	.2980618
Emphysema (yes vs. no)	1.33 (0.83-2.13), p = 0.23	.2878758	1.29 (0.80-2.10), p = 0.30	.2535959
Nodule type (semisolid vs. other)	2.01 (1.19-3.40), p = 0.009	.7004866	2.10 (1.23-3.59), p = 0.006	.7439433
Nodule spiculation (yes vs. no)	2.64 (1.53-4.54), p < 0.001	.9698875	2.59 (1.48-4.52), p = 0.001	.9501913
Nodule location (upper vs. other)	1.65 (1.06-2.58), p = 0.03	.5029189	1.65 (1.05-2.60), p = 0.03	.5012571
Nodule count ^T	0.92 ^a (0.88-0.96), p < 0.001	-.0852886 ^a	0.92 [⊥] (0.88-0.96), p < 0.001	-.0865209 [⊥]
Nodule size ^T	NA, ^a p < 0.001	-16.12322 ^a	NA [⊥] , p < 0.001	-9.228516 [⊥]
constant		-6.535526		-6.443251
Predictive performance statistics for the PanCan development data (N = 5779)				
Brier score	0.015 (0.013-0.018) ^c		0.015 (0.013-0.018) ^c	
AUC	0.947 (0.921-0.964) ^c		0.947 (0.919-0.964) ^c	
Observed/expected means	0.0202/0.0202 = 1.000		0.0202/0.0202 = 1.000	
Spiegelhalter z statistic	p = 0.425		p = 0.364	
Predictive performance statistics for the NLST validation data (N=6009)				
Brier score	0.064 (0.059-0.068) ^c		0.0594 (0.055-0.064) ^c	
AUC	0.810 (0.786-0.833) ^c		0.821 (0.797-0.843) ^c	
Observed/expected means	0.0766/0.0825 = 0.928		0.0766/0.0826 = 0.927	
Spiegelhalter z statistic	p < 0.001		p < 0.001	

^aIn the mean nodule model, the following transformations apply: Nodule count^T = Nodule count - 4 and Mean nodule diameter^T = ([LN Mean nodule diameter⁻¹]) - .6222482239.

^bIn the volume model, the following transformations apply: Nodule count^T = Nodule count - 4 and Volume^T = ([LN volume/10]⁻⁵) - 1.619158938.

^cBootstrap bias - corrected 95% confidence intervals.

PanCan, Pan-Canadian Early Detection of Lung Cancer Study; CAD, computer-aided detection; F, female; M, male; CI, confidence interval; AUC, area under the curve; LN, natural log.

diameter and volume models, remained good (see [Supplementary Table 1](#)).

That both PanCan CAD model risk estimates separate lung cancer from noncancer nodules is graphically

demonstrated in the PanCan study data ([Fig. 2](#)) and NLST data ([Supplementary Fig. 2](#)). [Table 3](#) and [Supplementary Figure 3](#) demonstrate that the sensitivity, specificity, and positive predictive value (PPV) for the PanCan CAD mean

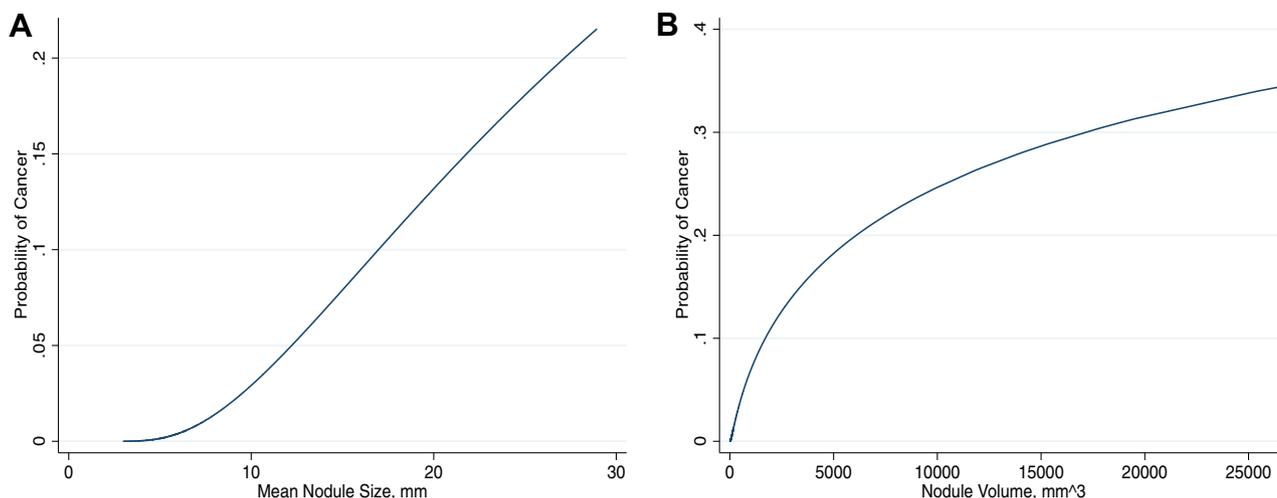


Figure 1. The relationship between mean nodule diameter and probability of lung cancer as determined by the Pan-Canadian Early Detection of Lung Cancer Study computer-aided detection (CAD) mean diameter model (A) and between nodule volume and probability of lung cancer as determined by the Pan-Canadian Early Detection of Lung Cancer Study CAD volume model (B.)

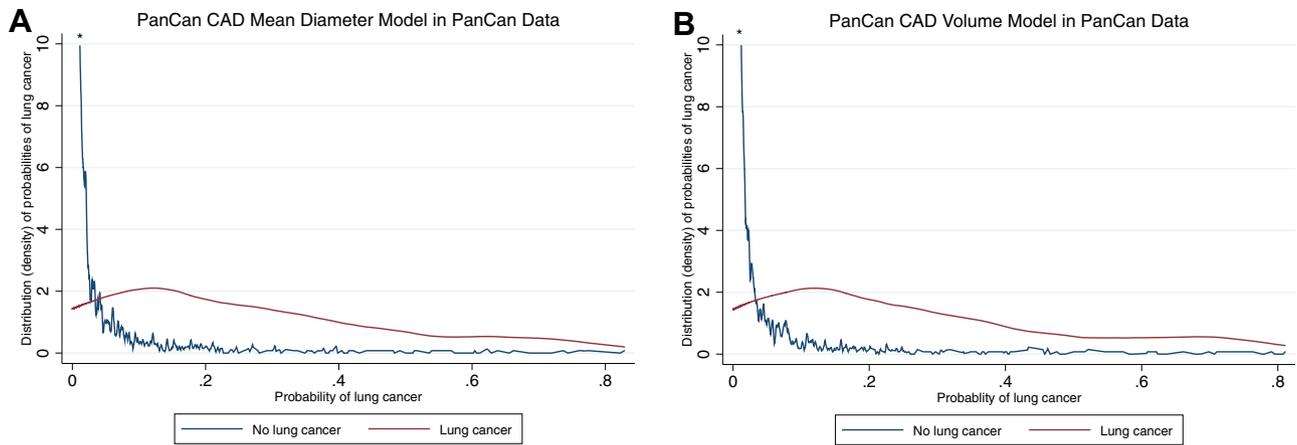


Figure 2. Distributions of estimated lung cancer probabilities in the Pan-Canadian Early Detection of Lung Cancer Study participants' nodules stratified by lung cancer status as determined by the Pan-Canadian Early Detection of Lung Cancer Study computer-aided detection (CAD) mean diameter and volume models. The densities for no lung cancer are truncated to improve visualization of the distributions.

diameter and volume models did not differ substantially over the range of risk estimates in the PanCan data. A probability threshold of 0.05 has sensitivities, specificities, and PPVs of 83.8%, 92.8%, and 19.4% and 84.6%, 93.1%, and 20.1% for the mean diameter and volume models, respectively. A probability threshold of 0.06 has sensitivities, specificities, and PPVs of 78.6%, 93.8%, and 20.6% and 80.3%, 94.0%, and 21.7%, respectively.

When predictive performance was evaluated in nodules that were 20 mm or smaller, discrimination was excellent in the PanCan data and moderate in the NLST data and overall calibration when assessed by observed and/or predicted mean probabilities was very good in both data sets, although the significant Spiegelhalter *p* value indicates that some estimates are different from the observed values (Supplementary Table 2).

When the original PanCan model parameters were applied to this study's CAD-derived data, the AUC was

0.940 (95% CI: 0.912–0.959) versus 0.947 for the Pan-Can CAD models. Although the differences between the original PanCan maximum diameter model⁴ and the PanCan CAD mean diameter and volume models were statistically significantly different (*p* = 0.01 and *p* = 0.002, respectively), their difference in magnitude was not clinically meaningful. When the original PanCan model parameters were applied to this study's data, the Brier score was 0.016, the respective observed and predicted mean probabilities were 0.0202 and 0.0328, and the Spiegelhalter *z* statistic *p* value was less than 0.001. These prediction statistics indicate that the Pan-Can CAD models provide prediction equal to or better than the original PanCan model in these training and testing data.

Spreadsheet calculators for the PanCan CAD mean diameter and volume models are available free of charge to noncommercial users at <https://brocku.ca/lung-cancer-screening-and-risk-prediction/risk-calculators/>.

Table 3. PanCan CAD Mean Diameter and Volume Model Sensitivities, Specificities, and Positive Predictive Values by Different Model Probability Thresholds, in PanCan Data

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.20	0.40	0.60
Mean diameter model													
Sensitivity	0.949	0.897	0.880	0.855	0.838	0.786	0.778	0.744	0.718	0.701	0.521	0.222	0.120
Specificity	0.800	0.869	0.897	0.915	0.928	0.938	0.947	0.953	0.959	0.962	0.985	0.996	0.998
PPV	0.088	0.124	0.151	0.172	0.194	0.206	0.233	0.248	0.263	0.277	0.409	0.520	0.583
NPV	0.999	0.998	0.997	0.997	0.996	0.995	0.995	0.995	0.994	0.994	0.990	0.984	0.982
Volume model													
Sensitivity	0.957	0.906	0.863	0.855	0.846	0.803	0.744	0.735	0.718	0.692	0.504	0.231	0.137
Specificity	0.794	0.870	0.902	0.919	0.931	0.940	0.948	0.956	0.961	0.963	0.985	0.995	0.998
PPV	0.088	0.126	0.154	0.178	0.201	0.217	0.228	0.257	0.274	0.279	0.404	0.474	0.640
NPV	0.999	0.998	0.997	0.997	0.997	0.996	0.994	0.994	0.994	0.993	0.990	0.984	0.982

CAD, computer-aided detection; PanCan, Pan-Canadian Early Detection of Lung Cancer Study; PPV, positive predictive value; NPV, negative predictive value.

Discussion

In this study, the CAD mean diameter and volume models demonstrated excellent overall prediction in the development and validation data, predictive abilities similar to those of the previously validated PanCan model that was based on radiologist-read LDCT scans and maximum nodule size, and similar predictive abilities between the mean diameter and volume models. Both CAD mean diameter and volume models had AUCs of 0.947 and Brier scores of 0.015 in the PanCan data and AUCs of 0.810 and 0.821 and Brier scores of 0.064 and 0.0594 in the NLST data, respectively. In addition, the CAD models demonstrated reasonably high predictive performance when limited to nodules 8 mm to 15 mm in diameter (see [Supplementary Table 1](#)). In external validation, the previously published PanCan model demonstrated higher AUCs in its validation data than the CAD-based models demonstrated in the current validation data. However, this is in large part explained by limitations in the NLST sample used for validation.

The NLST validation data set was developed for purposes other than the current study and was not ideal for validation of prediction models. Sampling NLST scans matched nodules without cancer on size to nodules that were cancer. This forced cancer nodules to be similar in size to noncancer nodules. Nodule size is the single most important predictor in models, and matching on nodule size reduces ability to predict cancer. In the PanCan data, each nodule was followed prospectively over time to identify which specific nodule was cancer; this was not the case in the NLST. Clinical judgement based on available data was used to decide whether a nodule was cancer. But in many cases, diagnosis of lung cancer came from a participant or proxy survey report or from a death registry. For the purposes of this study, it was assumed that the largest nodule was the one with development of cancer. In the PanCan study, 20% of cancer did not occur in the largest nodule. These issues add to measurement error in the NLST sample and biases prediction validation downward. NLST subset sampling included all interval cancers, which led to relative oversampling for prediction purposes. For interval cancers, it would not be possible to predict lung cancer on the basis of nodule characteristics because radiologically detected causal nodules were not detected on preceding scans. The NLST sample prevalence of lung cancer was higher than that observed in screening trials or programs. This is expected to lead to the appearance of poorer calibration when screening population-developed models are tested in lung cancer-enriched subset samples of the NLST. For the current study, the NLST subset sample was selected because of the limited

availability of similar high-quality data sets. They are time-consuming and expensive to produce. Regardless of the numerous limitations of the NLST sample for prediction validation, the two CAD-based models demonstrated moderate to excellent prediction in NLST subset validation.

The current study has potential limitations. Modeling was based on the CIRRUS CAD system. Whether model performance will differ with use of other CAD systems is unclear and needs to be evaluated. It has been shown that different software packages can result in significant differences in volumetric measurements.²⁸ Until scanners and acquisition protocols can be standardized internationally, it is recommended that volumetric measurement be done by using the same scanner and acquisition protocol for comparison of serial imaging studies.²⁹ We did not explore other methods of image analysis such as the relative roundness of a lesion or the relationship between automated mean diameter and volume because the purpose of our study was to compare mean diameter with volumetric measurement to predict malignancy potential. Mean diameter measurements can be done manually when CAD software is not available.

The models in the current study are not applicable to new nodules in annual repeat screening in which the malignancy risk of nodules smaller than 8 mm is higher than that of baseline nodules.³⁰ Currently, there are no validated prediction models for new nodules. In the American College of Radiology's Lung Imaging Reporting and Data System, the nodule size threshold for early-recall LDCT for new nodules is 4 mm or larger instead of 6 mm or larger for baseline scans.¹²

Semisolid nodules were measured and included in the models by their overall size, not by the size of the solid component. In our original PanCan model, nonsolid nodules were found to be at reduced risk compared with solid nodules, and this observation was consistent with those found by others.¹² In contrast, in the current study, nonsolid nodules were found to be at similar risk as solid nodules. Han et al. found that for detecting malignancy in pure nonsolid nodules, maximum cross-sectional area was superior to volume and other quantitative CT measures.³¹ The extent to which the findings for nonsolid nodules in our current study reflect measurement errors resulting from inaccurate measurement of nonsolid nodules by a CAD system remains to be determined.

Although a risk prediction model does not by itself constitute a nodule management protocol, an accurate risk prediction model can and possibly should be an important component of a nodule management protocol. This study provides support for using CAD-based nodule

measurements in risk assessment and also provides additional support for using radiologist-based measurements because they are not inferior to CAD-based measurements. Preference can depend on local practice and availability of CAD software. The differential distribution of characteristics of malignant and nonmalignant lung nodules allows rational selection of malignancy risk threshold to guide management.³² However, evaluating the clinical utility of lung nodule management tools or systems requires prospective studies using quality indicators by frequency and result of early-recall computed tomography, positron emission tomography-computed tomography studies, lung cancer detection rate, stage shift, interval cancers, and harm (for example, from unnecessary biopsies and surgical procedures) in different settings with different CAD systems.

Acknowledgments

This study was funded by the Terry Fox Research Institute and British Columbia Cancer Foundation. The funding sources had no involvement in the study design, data collection, or analysis and interpretation of data.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at <https://doi.org/10.1016/j.jtho.2018.10.006>.

References

- Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the U.S. Preventive Services Task Force recommendation. *Ann Intern Med.* 2013;159:411-420.
- U.S. Centers for Medicare and Medicaid Services. Decision memo for screening for lung cancer with low dose computed tomography (LDCT) (CAG-00439N). [https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274&NcaName=Screening+for+Lung+Cancer+with+Low+Dose+Computed+Tomography+\(LDCT\)&MEDCACId=68&lsPopup=y&bc=AAAAA AAAgAAAA%3d%3d&](https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274&NcaName=Screening+for+Lung+Cancer+with+Low+Dose+Computed+Tomography+(LDCT)&MEDCACId=68&lsPopup=y&bc=AAAAA AAAgAAAA%3d%3d&). Accessed August 31, 2015.
- Lewin G, Morissette K, Dickinson J, et al. Recommendations on screening for lung cancer. *CMAJ.* 2016;188:425-432.
- McWilliams A, Tammemagi M, Mayo J, et al. Probability of cancer in pulmonary nodules detected on first screening computed tomography. *N Engl J Med.* 2013;369(10):910-919.
- Al-Ameri A, Malhotra P, Thygesen H, et al. Risk of malignancy in pulmonary nodules: a validation study of four prediction models. *Lung Cancer.* 2015;89:27-30.
- Winkler Wille MM, van Riel SJ, Saghir Z, et al. Predictive accuracy of the PanCan lung cancer risk prediction model -external validation based on CT from the Danish Lung Cancer Screening Trial. *Eur Radiol.* 2015;25:3093-3099.
- Zhao H, Marshall HM, Yang IA, et al. Screen-detected subsolid pulmonary nodules: long-term follow-up and application of the PanCan lung cancer risk prediction model. *Br J Radiol.* 2016;89:20160016.
- van Riel SJ, Ciompi F, Jacobs C, et al. Malignancy risk estimation of screen-detected nodules at baseline CT: comparison of the PanCan model, Lung-RADS and NCCN guidelines. *Eur Radiol.* 2017;27:4019-4029.
- Nair VS, Sundaram V, Desai M, Gould MK. Accuracy of models to identify lung nodule cancer risk in the National Lung Screening Trial. *Am J Respir Crit Care Med.* 2018;197:120-123.
- Callister ME, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax.* 2015;70(suppl 2):ii1-ii54.
- Baldwin DR, Callister MEGuideline Development Group. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax.* 2015;70:794-798.
- ACR American College of Radiology. Lung CT screening reporting and data system (Lung-RADS). <http://www.acr.org/Quality-Safety/Resources/LungRADS>. Accessed April 27, 2017.
- Horeweg N, van der Aalst CM, Vliegenthart R, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J.* 2013;42:1659-1667.
- 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;18:1523-1531.
- Tammemagi MC, Katki HA, Hocking WG, et al. Selection criteria for lung-cancer screening. *N Engl J Med.* 2013;368:728-736.
- Murphy K, van Ginneken B, Schilham AM, de Hoop BJ, Gietema HA, Prokop M. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Med Image Anal.* 2009;13:757-770.
- Jacobs C, Sanchez CI, Saur SC, Twellmann T, de Jong PA, van Ginneken B. Computer-aided detection of ground glass nodules in thoracic CT images using shape, intensity and context features. *Medical Image Comput Assist Interv.* 2011;14(pt 3):207-214.
- Jacobs C, van Rikxoort EM, Twellmann T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Med Image Anal.* 2014;18:374-384.
- Ahn MI, Gleeson TG, Chan IH, et al. Perifissural nodules seen at CT screening for lung cancer. *Radiology.* 2010;254:949-956.
- de Hoop B, van Ginneken B, Gietema H, Prokop M. Pulmonary perifissural nodules on CT scans: rapid growth is not a predictor of malignancy. *Radiology.* 2012;265:61-616.
- MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology.* 2017;284:228-243.

22. Cuzick J. A Wilcoxon-type test for trend. *Stat Med*. 1985;4:87-90.
23. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. 2015. New York: Springer-Verlag.
24. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Hoboken, NJ: John Wiley & Sons; 2008.
25. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*. 1980;48:817-838.
26. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol*. 2010;63(8):938-939 [author reply: 939].
27. Pepe MS, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. *Stata J*. 2009;9:1-16.
28. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner Society. *Radiology*. 2017;285:584-600.
29. Rydzak CE, Armato SG, Avila RS, Mulshine JL, Yankelevitz DF, Gierada DS. Quality assurance and quantitative imaging biomarkers in low-dose CT lung cancer screening. *Br J Radiol*. 2017:20170401.
30. Pinsky PF, Gierada DS, Nath PH, Munden R. Lung cancer risk associated with new solid nodules in the National Lung Screening Trial. *AJR Am J Roentgenol*. 2017;209:1009-1014.
31. Han L, Zhang P, Wang Y, et al. CT quantitative parameters to predict the invasiveness of lung pure ground-glass nodules (pGGNs). *Clin Radiol*. 2018;73:504.e501-504.e507.
32. Tammemagi MC, Lam S. Screening for lung cancer using low dose computed tomography. *BMJ*. 2014;348:g2253.