Deep Learning—Quantified Calcium Scores for Automatic Cardiovascular Mortality Prediction at Lung Screening Low-Dose CT

Bob D. de Vries, PhD • Nikolas Lessmann, PhD • Pim A. de Jong, MD, PhD • Ivana Išgum, PhD

From the Department of Biomedical Engineering and Physics (B.D.d.V., I.I.), Cardiovascular Institute (B.D.d.V., I.I.), and Department of Radiology and Nuclear Medicine (I.I.), Amsterdam University Medical Center, Meibergdreef 9, Amsterdam, the Netherlands; and Image Sciences Institute (B.D.d.V., N.L., I.I.) and Department of Radiology (P.A.d.J., I.I.), University Medical Center Utrecht, Utrecht, the Netherlands. Received November 26, 2019; revision requested February 4, 2020; revision received January 5, 2021; accepted January 20. Address correspondence to B.D.d.V. (e-mail: b.d.devos@amsterdamumc.nl).

Conflicts of interest are listed at the end of this article.

Radiology: Cardiothoracic Imaging 2021; 3(2):e190219 • <https://doi.org/10.1148/ryct.2021190219> • Content codes: **AI** **CA** **CT**

Purpose: To examine the prognostic value of location-specific arterial calcification quantities at lung screening low-dose CT for the prediction of cardiovascular disease (CVD) mortality.

Materials and Methods: This retrospective study included 5564 participants who underwent low-dose CT from the National Lung Screening Trial between August 2002 and April 2004, who were followed until December 2009. A deep learning network was trained to quantify six types of vascular calcification: thoracic aorta calcification (TAC); aortic and mitral valve calcification; and coronary artery calcification (CAC) of the left main, the left anterior descending, and the right coronary artery. TAC and CAC were determined in six evenly distributed slabs spatially aligned among chest CT images. CVD mortality prediction was performed with multivariable logistic regression using least absolute shrinkage and selection operator. The methods were compared with semiautomatic baseline prediction using self-reported participant characteristics, such as age, history of smoking, and history of illness. Statistical significance between the prediction models was tested using the nonparametric DeLong test.

Results: The prediction model was trained with data from 4451 participants (median age, 61 years; 37.9% women) and then tested on data from 1113 participants (median age, 61 years; 37.9% women). The prediction model using calcium scores achieved a C statistic of 0.74 (95% CI: 0.69, 0.79), and it outperformed the baseline model using only participant characteristics (C statistic, 0.69; $P = .049$). Best results were obtained when combining all variables (C statistic, 0.76; $P < .001$).

Conclusion: Five-year CVD mortality prediction using automatically extracted image-based features is feasible at lung screening low-dose CT.

© RSNA, 2021

Cardiovascular disease (CVD) is the leading cause of mortality worldwide (1) and is also the leading cause of mortality in heavy smokers (> 30 pack-years) instead of lung cancer (2). The strongest predictor for CVD mortality and other adverse events is the presence of arterial calcium (3–8), which is highly prevalent in heavy smokers (3,9). Given that the low-dose CT performed in lung cancer screening allows extracting information about calcification in the heart and aorta, this information might be exploited to supplement lung cancer screening with CVD screening. It has been shown that lung screening with low-dose CT of heavy smokers reduces lung cancer mortality (2). Correspondingly, screening for CVD in asymptomatic lung screening participants might lead to a further mortality reduction.

Numerous studies have shown the prognostic value of arterial calcium quantified at lung screening low-dose CT. Specifically, coronary artery calcium (CAC), thoracic aorta calcium (TAC), aortic valve calcification, and mitral valve calcification (3,10–15) have been quantified in previous studies. To the best of our knowledge, the previous studies that investigated arterial calcium in relation to CVD mortality used information extracted from CT images as well as other risk factors, such as cholesterol levels and blood

pressure, and self-reported clinical data, such as history of illness. However, to complement lung screening with CVD screening, its implementation should preferably be automatic, fast, and only cause minimal extra workload.

The purpose of this study was to examine the prognostic value of automatically derived arterial calcification quantities from lung screening low-dose CT for the prediction of 5-year CVD mortality. We trained and tested a method for 5-year CVD mortality prediction using automatically determined calcium quantities originated from six vascular beds. Additionally, we determined the locations of these calcifications in six evenly distributed slabs that were spatially corresponding among low-dose CT scans.

Materials and Methods

Population and Sample Selection

A retrospective multicenter nested case-control study was designed using data obtained from the National Lung Screening Trial (NLST). The design of the NLST has been described elsewhere, and each of the 33 sites had institutional review board approval at the time of the study (2). Briefly, the NLST was designed to compare lung cancer screening with chest radiography to screening

Abbreviations

CAC = coronary artery calcium, CalcScore model = model using calcium scoring, CVD = cardiovascular disease, LASSO = least absolute shrinkage and selection operator, MESA = Multi-Ethnic Study of Atherosclerosis, NLST = National Lung Screening Trial, ParChars = model using participant characteristics, ParChars+CAC = model using participant characteristics and total coronary artery calcium, ParChars+CalcScore = model using participant characteristics and calcium scoring, ROC = receiver operating characteristic, TAC = thoracic aorta calcium

Summary

Real-time mortality prediction at lung screening chest CT is feasible using only information from automatic deep learning-quantified arterial calcium.

Key Points

- Cardiovascular mortality can be predicted using deep learning-quantified location-specific arterial calcium (C statistic, 0.74).
- Addition of self-reported risk factors inhibits the fully automatic approach but improves mortality prediction (C statistic, 0.76; $P < .001$).

with low-dose chest CT. The NLST included a cohort of participants between 55 and 74 years of age with a smoking history of at least 30 pack-years. Exclusion criteria were: previous diagnosis of lung cancer, chest CT within 18 months before enrollment, hemoptysis, or an unexplained weight loss of more than 6.8 kg (15 lb) in the preceding year. Participants were enrolled in the study between August 2002 and April 2004 and followed until December 2009. Of 53 454 participants, 26 722 underwent low-dose chest CT screening with a median follow-up time of 6.5 years in one of 33 participating medical centers in the United States.

Participants who were lost to follow-up were excluded. Cause of death was determined from death certificates using codes of the International Classification of Disease, 10th edition. End-points for CVD mortality were defined using codes 100–199; those with other mortality causes were excluded from the analysis given the likelihood of competing risks (16).

All CVD deaths within 5 years after undergoing the baseline CT scan were included as cases. This time interval showed an adequate balance of sufficient number of participants and acceptable time to event. We supplemented the deceased cases with twice as many randomly selected controls, which were participants who did not meet the primary endpoint. The controls included participants who died of CVD after the 5-year time point. The cohort was randomly divided into a training cohort, with 80% of the participants used to train the models, and a test cohort, with 20% of the participants for final evaluation of the models. Details of the selection are shown in Figure 1.

Available self-reported clinical data known to be predictive of CVD risk (17) were extracted from the data set. These data were age, sex, pack-years of smoking, current smoking status, and history of heart disease, stroke, hypertension, and/or diabetes.

CT Scan Acquisition

Baseline images were selected from the available set of low-dose chest CT images. Low-dose CT was performed without

electrocardiography synchronization and contrast agents using a wide variety of CT scanner models from multiple vendors (2). Low-dose CT was performed with 120–140 kVp tube voltage and with tube currents varying between 40 and 80 mAs. Because the section thickness and increment varied between scans, all images were resampled to 3-mm section thickness with 1.5-mm increments, as is common for calcium scoring (18).

Calcium Quantification

Arterial calcium was quantified using a previously described fully automatic deep learning method that directly quantified calcium from input image sections (19). Briefly, the method employs two convolutional neural networks. The first network aligns and crops a low-dose chest CT image at hand to a previously created atlas image to mimic the field of view of cardiac CT. The second network directly quantifies calcium (ie, determines Agatston scores) in axial image sections extracted from the aligned low-dose CT image.

We trained the second convolutional neural network for calcium scoring at six different sites. In each axial section, the method quantifies TAC, aortic valve calcification, mitral valve calcification, and CAC in three different arteries: the right coronary artery, the left circumflex artery, and left anterior descending artery. Differentiation of CAC in the left main from CAC in the left anterior descending artery is infeasible due to lack of intravenous contrast enhancement, absence of electrocardiography synchronization, and relatively low image resolution. Thus, CAC in the left main and left anterior descending artery were combined. In addition, differentiation between aortic and mitral valve calcification and their annuli was not possible, hence these were included in aortic valve calcification and mitral valve calcification.

To establish the prognostic value of spatial distributions of CAC and TAC, predicted CAC and TAC scores were binned into evenly spaced slabs. Binning multiple sections into a fixed number of slabs ensures a fixed number of variables that can be used for subsequent prediction. Because the images were aligned before calcium scoring, the per-slab CAC and TAC scores were spatially corresponding among participants. As illustrated in Figure 2, slabs 1 through 6 were positioned from inferior to superior positions along the axial direction. Any number of slabs might be chosen, but we chose six, providing an optimal trade-off between slab resolution and prevalence of the calcium scores. Because aortic valve calcification and mitral valve calcification are very local, total scores were used.

Prediction Model

We derived a multivariable logistic regression model for 5-year CVD mortality prediction using the automatically obtained per-slab CAC and TAC scores and total mitral valve calcification and aortic valve calcification scores. The full pipeline is shown in Figure 3. These scores were log transformed to correct for skewness of the distributions. Additionally, continuous variables were normalized to zero mean and unit variance using the mean and variance computed on the training cohort. The least absolute shrinkage and selection operator (LASSO) was employed to create a sparse

Figure 1: Flowchart illustrates the selection of the studied cohort of National Lung Screening Trial (NLST) participants. Table 1 lists additional information of the data. DICOM = Digital Imaging and Communications in Medicine.

model (20) (ie, LASSO uses L1 regularization to force coefficients toward zero). As a consequence, overfitting is averted, and only a few coefficients related to the problem at hand persist. The training cohort was used in a threefold cross-validation setting to select the optimal L1 regularization weight (ie, λ). To account for overfitting, the one-standard-error rule was used to select the best model (21). The test cohort was only used for final evaluation of the model and not during model development.

We compared our prediction model (called CalcScore) with three other models that use the aforementioned approach but with different variables: (a) a baseline model using self-reported participant characteristics (ie, CVD risk factors only; called ParChars); (b) a model using self-reported participant characteristics and automatically determined total CAC (called ParChars+CAC); and (c) a model using self-reported participant characteristics and the location-based calcium scores determined by our method (called ParChars+CalcScore).

Statistical Analysis

The models were evaluated using receiver operating character-

istic (ROC) curves. Although, ROC curve analysis requires balanced cases and controls, imbalance in this case-control study was not an issue because the same test cohort was used for all models. As a consequence, ROC curve analysis could reliably be used to compare prediction performance of the models. The concordance statistic (C statistic) was used to evaluate the discriminative performance of the models. Statistical significance between prediction models was evaluated using the nonparametric DeLong test (22).

All statistical analyses were performed with R (version 3.4.1, R Foundation for Statistical Computing). The package glmnet (21) was employed using threefold cross-validation to determine an optimal model used for final evaluation on the test cohort. The package pROC (23) was used for ROC curve analysis.

Processing Time

Predictions were obtained in 0.5 second on a state-of-the-art graphics processing unit (NVIDIA GTX 1080 Ti) or in 15 seconds on a single core of

a central processing unit (Intel Xeon 6128 3.40 GHz) for all stages of analysis, including image alignment and automatic calcium quantification.

Results

Population and Sample Selection

In total, 5861 lung screening participants were selected from the NLST. Participants with missing information were removed from the data sets ($n = 197$). Using uniform random selection, the cohort was split into a cohort for training ($n = 4451$) and a cohort for testing ($n = 1113$) of the derived models. Demographic data of the training and test cohorts are shown in Table 1.

Prevalence of Arterial Calcification

Calcium distributions as predicted by automatic calcium scoring are listed in Table 2. The most prevalent were left anterior descending artery calcifications and TAC in slab 2 in 66.0% (3667 of 5564) and 57.4% (3191 of 5564) of the

Figure 2: Projections of all aligned chest CT scans show feasibility of slab-based quantification of calcium, resulting in an average image. For alignment, only translation, rotation, and scaling were allowed, resulting in a blurry image, because not all anatomy is exactly the same across participants. From left to right, the center axial, sagittal, and coronal sections are shown. Note that field of view is similar to cardiac CT, which is a consequence of image alignment by the used automatic calcium scoring method. Image alignment allowed the determination of calcification distributions into slabs as a proxy measure for proximal and distal calcifications, of which the borders are indicated by the horizontal lines.

Figure 3: Full pipeline from input to output. x_n = resulting calcium scores. LASSO = least absolute shrinkage and selection operator.

participants, respectively. The distributions of calcification were similar between the training cohort and the test cohort. Figure 4 shows two randomly selected cases (participants who died within 5 years of low-dose CT) and two randomly selected controls (participants who did not meet primary endpoint) from the test set, each with corresponding automatically determined calcium scores.

CVD Mortality Prediction

Figure 5 presents the coefficients of the model using here-proposed per-slab calcium scores. It shows that nine of 28 location-specific calcium scores had nonzero coefficients. The enforced sparsity by using LASSO in the prediction model selected

Table 1: Demographics of Training Cohort for Model Development and Holdout Test Cohort

Participant Characteristic	Training Cohort ($n = 4451$)	Test Cohort ($n = 1113$)
Age (y)*	61 (57–66)	61 (58–65)
No. of women	1687 (37.9)	422 (37.9)
Current smoker	2209 (49.6)	558 (50.1)
Pack-years*	50 (40–70)	51 (40–70)
History of heart disease	668 (15.0)	152 (13.7)
History of hypertension	1592 (35.8)	412 (37.0)
History of stroke	159 (3.6)	34 (3.1)
History of diabetes	494 (11.1)	143 (12.8)
Cause of mortality		
All cause	890 (20.0)	243 (21.8)
Cardiovascular disease	262 (5.9)	87 (7.8)

Note.—Unless otherwise noted, values are numbers with percentages in parentheses.

*Values are given as median with interquartile range in parentheses.

Table 2: Calcium Distribution Predicted by Automatic Calcium Scoring of Training Cohort for Model Development and Holdout Test Cohort

Parameter	Training Cohort (<i>n</i> = 4451)		Test Cohort (<i>n</i> = 1113)	
	Agatston Score	Participants (%)	Agatston Score	Participants (%)
Slab 1				
LM + LAD	189 (29–855)	21.2	170 (31–798)	22.4
LCX	23 (3–113)	4.0	29 (5–120)	4.9
RCA	20 (3–165)	2.1	25 (2–43)	2.8
Aorta	585 (86–2601)	45.5	599 (59–2175)	45.7
Slab 2				
LM + LAD	672 (119–2409)	66.3	596 (153–2172)	64.3
LCX	194 (32–738)	31.9	187 (24–693)	30.9
RCA	156 (24–624)	23.2	109 (18–507)	22.5
Aorta	612 (76–2459)	57.4	538 (83–2157)	57.1
Slab 3				
LM + LAD	150 (19–609)	22.8	92 (8–478)	23.2
LCX	204 (43–694)	27.8	162 (33–654)	27.4
RCA	277 (47–1260)	39.6	304 (43–955)	37.0
Aorta	547 (90–2138)	49.3	487 (69–1880)	49.6
Slab 4				
LM + LAD	22 (3–144)	4.1	48 (5–110)	3.4
LCX	58 (7–245)	9.1	30 (3–249)	9.6
RCA	221 (39–1015)	29.2	150 (37–822)	28.2
Aorta	557 (91–2088)	38.5	577 (98–2014)	37.5
Slab 5				
LM + LAD	20 (4–97)	2.0	15 (2–97)	1.9
LCX	8 (2–39)	2.7	8 (3–39)	3.1
RCA	179 (25–762)	21.5	150 (26–684)	21.1
Aorta	556 (89–2059)	37.0	506 (89–1990)	36.0
Slab 6				
LM + LAD	5 (1–18)	1.2	4 (1–11)	0.5
LCX	1 (1–9)	0.3	1 (1–1)	0.1
RCA	86 (8–359)	5.2	61 (8–179)	5.7
Aorta	574 (99–1865)	37.5	552 (89–1846)	36.2
Valves				
Aortic	28 (4–325)	24.6	27 (6–289)	24.3
Mitral	39 (6–286)	11.9	24 (4–376)	9.9

Note.—Given that the majority of participants had no calcium in the specified regions, Agatston scores, given as medians with interquartile ranges in parentheses, were only calculated among participants with calcium of a specific category. The adjacent percentages indicate the percentage of participants having calcium of the specific categories. Calcification quantities are provided for the aortic and mitral valves and for six slabs each with four categories: left main and left anterior descending artery (LM+LAD), left circumflex artery (LCX), right coronary artery (RCA), and the thoracic aorta. There were no statistically significant differences between the cohorts.

the variables (ie, features) most predictive for 5-year mortality prediction. The analysis shows that importance of calcifications in the final prediction model varies depending on the origin of the calcium. The dependence of origin is particularly apparent for TAC in slab 1, CAC in the left circumflex artery and left anterior descending artery in slab 2, and CAC in the right coronary artery in slab 5, as these

factors contribute the most to the final prediction model.

Evaluation was performed using the holdout test set. Our proposed model (CalcScore) achieved a C statistic of 0.74 (95% CI: 0.69, 0.79). The ParChars model achieved a C statistic of 0.69 (95% CI: 0.64, 0.74). The ParChars+CAC model obtained a C statistic of 0.72 (95% CI: 0.67, 0.77). Finally, the ParChars+CalcScore model had a C statistic of 0.76 (95% CI:

Figure 4: From top to bottom, maximum intensity projections of four randomly selected images from the evaluation set. The top two rows show cases that contain extensive arterial calcification (mainly the aortic arch), and the bottom two rows show controls containing none or limited arterial calcification. The left and right columns show maximum intensity projections of respective sagittal views and coronal views. The middle column lists Agatston scores if present in the slab. Note that none of the examples contained valve calcifications. Images have been automatically aligned such that they mimic the field of view of cardiac CT. However, for visualization purposes, large parts of the ribs and the vertebrae have been manually cropped from these examples. LAD = left anterior descending artery, LCX = left circumflex artery, RCA = right coronary artery.

0.71, 0.80). Figure 6 shows the obtained ROC curves.

Results of the DeLong tests for comparison of the prediction models are listed in Table 3. The ParChars+CAC model did not significantly improve when compared with the ParChars model ($P = .11$). However, a significant difference was shown between the ParChars model and the CalcScore model ($P = .049$). The best performing model was the ParChars+CalcScore model ($P < .001$). There was no difference between the CalcScore model and ParChars+CAC ($P = .27$), meaning that the prediction model using solely imaging-based features (ie, slab-based

per-artery calcium scores) is equally good as a prediction model that relies on standardly used self-reported participant characteristics and total CAC.

Discussion

We developed and validated a method for real-time 5-year CVD mortality prediction using only information automatically extracted from lung screening chest CT images. The method used an automatic deep learning approach for direct calcium scoring to determine site-specific calcium scores. Cal-

Figure 5: Feature importance of the derived model using only location-specific calcium scores as image features. Feature importance is determined by scaling the coefficients obtained by least absolute shrinkage and selection operator regression. For reference, the average coronal center section of all aligned chest CT scans is shown. LAD = left anterior descending artery, LCX = left circumflex artery, LM = left main, RCA = right coronary artery.

Figure 6: Receiver operating characteristic (ROC) curves for derived models predicting 5-year cardiovascular disease mortality. ROC curves were obtained from the test cohort. Models were derived using participant characteristics (ParChars), a combination of participant characteristics and total coronary calcium scores (ParChars+CAC), our proposed location-specific calcium scores (CalcScore), and finally by a combination of participant characteristics data and location-specific calcium scores (ParChars+CalcScore). Please note that 95% confidence regions of all ROC curves overlap heavily and are thus not shown.

cification was quantified in six vascular beds: in the thoracic aorta, the mitral valve, the aortic valve, in the right coronary artery, the left circumflex artery, and in the left anterior descending artery and left main. Per-section calcium scores were binned into six evenly and spatially distributed slabs that were matched among participant CT scans. Atlas-based registration resulted in aligned anatomy, and as cardiovascular anatomy is

quite uniform across patients, we postulated that proximal arteries are predominantly superior to the distal arteries. This led us to the presumption that slabs can be considered as a proxy measure for distal and proximal CAC and TAC distributions. The results show that these slab-based Agatston scores provide a better prediction model than self-reported clinical data.

There was no significant difference in the performance between our prediction model (CalcScore) and the model using total CAC scores and traditional risk factors (age, sex, pack-years of smoking, current smoking status, history of heart disease, stroke, hypertension, and diabetes; the ParChars+CAC model).

These findings are in line with a previous study predicting CVD events in men who underwent lung cancer screening (24), where the model using image-derived information had similar performance compared with the model with image and self-reported clinical data. In contrast, the model used here was derived using a larger and more heterogeneous population.

The prediction model that used a combination of our proposed features and self-reported clinical data (ParChars+CalcScore model) achieved the best performance. However, inclusion of self-reported data hinders simplicity of implementation because of increased administrative overhead, and relying on self-reported features may decrease reproducibility. In contrast, a model that solely uses image information (CalcScore model) is fully automatic and therefore simpler to implement in a screening setting. Our CalcScore model, using only image-derived calcium scores, was a significant predictor of CVD mortality.

Use of LASSO allowed us to inspect the sites of calcification most important for mortality prediction. This resulted in three findings. First, this analysis revealed that proximal TAC was most important, which concurs with the Multi-Ethnic Study of Atherosclerosis (MESA), where TAC in the ascending aorta was shown to be predictive for CVD risk (25). However, our method cannot discern TAC within a slab. Second, previous studies have shown that aortic valve calcification did not improve CVD event prediction when combined with traditional risk factors (13,14,26). However, in our study, aortic valve calcification was an important variable and favored by the model over the majority of CAC sites. Third, mitral valve calcification did not contribute to prediction, likely because of the low occurrence of mitral valve calcification or that it is difficult to separate from CAC in the distal left circumflex artery. Previously, it has been shown that manual expert detection of mitral valve calcification at low-dose CT is subject to large interobserver variability (27). Nevertheless, further studies with electrocardiography-synchronized images with higher resolution are required to confirm

this finding.

There was limited additional prognostic power when total CAC scores were added to self-reported participant characteristics. This might be due to the cohort used in this study in which the majority of the participants had CAC. As a consequence, CAC presence might provide insufficient prognostic power for mortality prediction, in contrast to a more homogeneous cohort, like that of MESA (28).

Predictive variables were CAC in the proximal left main, left anterior descending artery, left circumflex artery, and right coronary artery. While it is known that CAC in the right coronary artery has prognostic value for cardiovascular events in participants of MESA (28), the results here indicate that especially distal right coronary artery had prognostic value, which has been related to advanced CVD (29). It is unclear why in our model CAC in the proximal left circumflex artery was selected in favor of CAC in left main and left anterior descending artery. Moreover, coefficients of left main and left anterior descending artery CAC were relatively low, while CAC in the left main is known to be predictive for all-cause mortality (30). The relatively high prevalence of left main and left anterior descending artery calcium in heavy smokers in the NLST cohort possibly mitigated its discriminative value compared with a more heterogeneous cohort regarding age and smoking history similar to the MESA study where participants had less CAC in left main and left anterior descending artery (28,31). Thus, the CAC in left main and left anterior descending artery may not be discriminative in this population of heavy smokers.

This study had some limitations. First, the study cohort consisted of a population of heavy smokers with a limited age range. Thus, the findings might not be directly applicable to the general population. However, a population of heavy smokers is more prone to CVD; thus, identifying subclinical CVD might be beneficial for implementation in lung screening. Second, given that distal and proximal calcium distributions were determined through scoring per slab, TAC in the ascending or descending aorta could not be distinguished. Increased precision of localization might elucidate if ascending or descending TAC is more predictive for CVD mortality. Third, the chest CT scans used here were not electrocardiography synchronized, and therefore they were prone to motion artifacts. This could lead to incorrect prediction of calcium. Nonetheless, it is known that estimates of calcium amount (eg, qualitative visual scoring at low-dose chest CT) is adequate for CVD risk assessment (11,12,32,33). However, visual scoring is manual, and it is subjective. In contrast, the method we proposed here is automatic, fast (0.5 sec/scan), and reproducible. Fourth, the employed multivariable logistic regression model assumes linear relations for mortality prediction, while some variables might have a nonlinear interaction with mortality. Fifth, our method quantifies calcium in left main and left anterior descending artery, left circumflex artery, right coronary artery, aorta, and valves. However, precision of calcium scoring has not been evaluated but only served as an intermediate step toward prediction of mortality. Previous work using a similar approach

Table 3: P Values from DeLong Tests Comparing Receiver Operating Characteristic Curves of Different Models

Model	ParChars+ CAC	CalcScore	ParChars+ CalcScore
ParChars	.11	.049	< .001
ParChars+ CAC	...	.27	.02
CalcScore	...	...	.09

Note.—P values are shown for the comparison of models using participant characteristics (ParChars), participant characteristics and total coronary calcium scores (ParChars+ CAC), proposed image-derived per-slab calcium feature (CalcScore), and participant characteristics and proposed image-derived per-slab calcium feature (ParChars+ CalcScore).

demonstrated accurate coronary calcium scoring, and future work could investigate the precision of this approach for calcium scoring in the other arterial beds. Sixth, the slabs could only be interpreted as proxy measures for proximal and distal calcium score, because the employed method cannot pinpoint the exact distance of a calcification to the arterial root. Finally, the proposed model only considered atherosclerotic calcification as predictors for CVD, while other characteristics (eg, hypertrophy of the heart or aneurysms) might be predictive as well. In future studies, it would be interesting to study a deep learning approach using the full image as an input variable for CVD mortality prediction, similar to the study by Oakden-Rayner et al (34). However, such models are prone to overfitting and might require large sets of data, especially when no prior structural information is available.

We have shown that 5-year CVD mortality can be predicted for lung screening participants in less than half a second, using only site-specific calcium scores automatically derived from lung screening low-dose CT. Hence, the proposed image-based analysis could aid in identification of lung screening participants at risk for CVD mortality, without relying on self-reported participant data.

Acknowledgments: The authors are grateful to the United States National Cancer Institute (NCI) for providing access to NCI's data collected by the National Lung Screening Trial. The statements contained herein are solely ours and do not represent or imply concurrence or endorsement by NCI.

Author contributions: Guarantor of integrity of entire study, B.D.d.V.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, B.D.d.V.; statistical analysis, B.D.d.V.; and manuscript editing, B.D.d.V., I.I.

Disclosures of Conflicts of Interest: B.D.d.V. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author is co-founder of Quantib-U and is currently an advisor and shareholder, current work was largely performed before author's employment there. Other relationships: disclosed no relevant relationships. N.L. disclosed no relevant relationships. P.A.d.J. disclosed no relevant relationships. I.I. Activities related to the present article: author's institution has research grant from Dutch Technology Foundation (No. 12726). Activities not related to the present article: author's institution has research grant from Dutch Technology Foundation with participation of Philips Healthcare and Pie Medical Imaging, research grants from Pie Medical Imaging, research grant from The Netherlands Organisation for Health Research and Development with participation of Pie Medical Imaging; author received money for patent issued by

Pie Medical Imaging. Other relationships: disclosed no relevant relationships.

References

- Cardiovascular diseases (CVDs) - Fact Sheet. World Health Organization. <http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>. Published June 7, 2018. Accessed May 1, 2020.
- National Lung Screening Trial Research Team; Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365(5):395–409.
- Jacobs PC, Prokop M, van der Graaf Y, et al. Comparing coronary artery calcium and thoracic aorta calcium for prediction of all-cause mortality and cardiovascular events on low-dose non-gated computed tomography in a high-risk population of heavy smokers. *Atherosclerosis* 2010;209(2):455–462.
- Budoff MJ, Hokanson JE, Nasir K, et al. Progression of coronary artery calcium predicts all-cause mortality. *JACC Cardiovasc Imaging* 2010;3(12):1229–1236.
- Kälsch H, Lehmann N, Berg MH, et al. Coronary artery calcification outperforms thoracic aortic calcification for the prediction of myocardial infarction and all-cause mortality: the Heinz Nixdorf Recall Study. *Eur J Prev Cardiol* 2014;21(9):1163–1170.
- Wang AYM, Wang M, Woo J, et al. Cardiac valve calcification as an important predictor for all-cause mortality and cardiovascular mortality in long-term peritoneal dialysis patients: a prospective study. *J Am Soc Nephrol* 2003;14(1):159–168.
- Budoff MJ, Nasir K, Katz R, et al. Thoracic aortic calcification and coronary heart disease events: the multi-ethnic study of atherosclerosis (MESA). *Atherosclerosis* 2011;215(1):196–202.
- Budoff MJ, Nasir K, McClelland RL, et al. Coronary calcium predicts events better with absolute calcium scores than age-sex-race/ethnicity percentiles: MESA (Multi-Ethnic Study of Atherosclerosis). *J Am Coll Cardiol* 2009;53(4):345–352.
- Shemesh J. Coronary artery calcification in clinical practice: what we have learned and why should it routinely be reported on chest CT? *Ann Transl Med* 2016;4(8):159.
- Mets OM, Vliegenthart R, Gondrie MJ, et al. Lung cancer screening CT-based prediction of cardiovascular events. *JACC Cardiovasc Imaging* 2013;6(8):899–907.
- Chiles C, Duan F, Gladish GW, et al. Association of Coronary Artery Calcification and Mortality in the National Lung Screening Trial: A Comparison of Three Scoring Methods. *Radiology* 2015;276(1):82–90.
- Watts JR Jr, Sonavane SK, Snell-Bergeon J, Nath H. Visual scoring of coronary artery calcification in lung cancer screening computed tomography: association with all-cause and cardiovascular mortality risk. *Coron Artery Dis* 2015;26(2):157–162.
- Jairam PM, Gondrie MJA, Grobbee DE, et al. Incidental imaging findings from routine chest CT used to identify subjects at high risk of future cardiovascular events. *Radiology* 2014;272(3):700–708.
- Willemink MJ, Takx RAP, Išgum I, et al. Prognostic value of heart valve calcifications for cardiovascular events in a lung cancer screening population. *Int J Cardiovasc Imaging* 2015;31(6):1243–1249.
- Takx RAP, Išgum I, Willemink MJ, et al. Quantification of coronary artery calcium in nongated CT to predict cardiovascular events in male lung cancer screening participants: results of the NELSON study. *J Cardiovasc Comput Tomogr* 2015;9(1):50–57.
- Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;32(30):5381–5397.
- Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97(18):1837–1847.
- Rutten A, Išgum I, Prokop M. Calcium scoring with prospectively ECG-triggered CT: using overlapping datasets generated with MPR decreases inter-scan variability. *Eur J Radiol* 2011;80(1):83–88.
- de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Išgum I. Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. *IEEE Trans Med Imaging* 2019;38(9):2127–2138.
- Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Series B Stat Methodol* 1996;58(1):267–288.
- Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33(1):1–22.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837–845.
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12(1):77.
- de Vos BD, de Jong PA, Wolterink JM, et al. Automatic machine learning based prediction of cardiovascular events in lung cancer screening data. In: Hadjiiski LM, Tourassi GD, eds. *Proceedings of SPIE: medical imaging 2015—computer-aided diagnosis*. Vol 9414. Bellingham, Wash: International Society for Optics and Photonics, 2015; 94140D.
- Thomas IC, McClelland RL, Michos ED, et al. Density of calcium in the ascending thoracic aorta and risk of incident cardiovascular disease events. *Atherosclerosis* 2017;265:190–196.
- Kälsch H, Lehmann N, Mahabadi AA, et al. Beyond Framingham risk factors and coronary calcification: does aortic valve calcification improve risk prediction? The Heinz Nixdorf Recall Study. *Heart* 2014;100(12):930–937.
- van Hamersvelt RW, Willemink MJ, Takx RA, et al. Cardiac valve calcifications on low-dose unenhanced ungated chest computed tomography: inter-observer and inter-examination reliability, agreement and variability. *Eur Radiol* 2014;24(7):1557–1564.
- Blaha MJ, Budoff MJ, Tota-Maharaj R, et al. Improving the CAC Score by Addition of Regional Measures of Calcium Distribution: Multi-Ethnic Study of Atherosclerosis. *JACC Cardiovasc Imaging* 2016;9(12):1407–1416.
- Grunfeld C, Scherzer R, Varosy PD, Ambarish G, Nasir K, Budoff M. Relation of coronary artery plaque location to extent of coronary artery disease studied by computed tomographic angiography. *J Cardiovasc Comput Tomogr* 2010;4(1):19–26.
- Williams M, Shaw LJ, Raggi P, et al. Prognostic value of number and site of calcified coronary lesions compared with the total score. *JACC Cardiovasc Imaging* 2008;1(1):61–69.
- Brown ER, Kronmal RA, Bluemke DA, et al. Coronary calcium coverage score: determination, correlates, and predictive accuracy in the Multi-Ethnic Study of Atherosclerosis. *Radiology* 2008;247(3):669–675.
- Azour L, Kadoch MA, Ward TJ, Eber CD, Jacobi AH. Estimation of cardiovascular risk on routine chest CT: Ordinal coronary artery calcium scoring as an accurate predictor of Agatston score ranges. *J Cardiovasc Comput Tomogr* 2017;11(1):8–15.
- Shemesh J, Henschke CI, Shaham D, et al. Ordinal scoring of coronary artery calcifications on low-dose CT scans of the chest is predictive of death from cardiovascular disease. *Radiology* 2010;257(2):541–548.
- Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ. Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Sci Rep* 2017;7(1):1648.