

# Predictive value of automated coronary calcium scoring in lung cancer screening with low dose computed tomography

Federica Sabia<sup>1</sup>, Maurizio Balbi<sup>1,2</sup>, Roberta E. Ledda<sup>1,2</sup>, Gianluca Milanese<sup>1,2</sup>, Margherita Ruggirello<sup>3</sup>, Camilla Valsecchi<sup>1</sup>, Alfonso Marchianò<sup>3</sup>, Nicola Sverzellati<sup>2</sup>, Ugo Pastorino<sup>1</sup>

<sup>1</sup>Department of Thoracic Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

<sup>2</sup>Section of Radiology, Department of Medicine and Surgery (DiMeC), University Hospital of Parma, Parma, Italy

<sup>3</sup>Department of Radiology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

Correspondence to:

Ugo Pastorino MD, Division of Thoracic Surgery, Department of Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian 1 – 20133 Milan – Italy. Tel +390223902367. Email: ugo.pastorino@istitutotumori.mi.it

**Competing interests:** No competing interest to declare.

## Abstract

Coronary artery calcium (CAC) is a known risk factor for cardiovascular events, but not yet routinely evaluated in Low Dose Computed Tomography (LDCT) screening. The present analysis compared the accuracy of a new automated CAC quantification versus prior manual quantification on baseline LDCT screening images as predictors of all-cause mortality at 12 years.

The study included 1129 volunteers of the Multicentric Italian Lung Detection (MILD) trial who underwent a baseline LDCT scan from September 2005 to September 2006, already analyzed in a previous paper on CAC scoring. The initial manual CAC (mCAC) had been scored by one operator using a dedicated software, while the new automated CAC (aCAC) score was measured by a fully automated artificial intelligence software. All CAC scores were stratified in four categories: 0, 0.1-19.9, 20-399, and  $\geq 400$ .

The study showed a high correlation between aCAC and mCAC scores, with an Intraclass Correlation Coefficient of 0.887. Of 613 negative mCAC score, 87.6% had aCAC score  $>0$ , and 14.0%  $>20$ . A CAC score  $>20$  revealed a higher risk of 12-year all-cause mortality both with mCAC

and aCAC. Focusing on the 535 individuals with false negative mCAC score, aCAC identified a subset of volunteers with a significantly poorer survival of 86% (aCAC 20-399,  $p=0.0007$ ).

CAC quantification could be accurately and safely performed with a fully automated software on baseline LDCT screening images to predict all-cause mortality risk.

## Introduction

Coronary artery calcium (CAC) is an independent predictor of cardiovascular (CV) events (1). Previous studies in low dose computed tomography (LDCT) screening participants demonstrated the predictive value of CAC score for all-cause and CV mortality (2,3), with a CAC scoring accuracy of LDCT that was similar to the one achieved by electrocardiographic-gated cardiac CT (4,5).

CAC evaluation is not routinely performed in LDCT screening because manual CAC measurements is a highly time-consuming procedure, but currently available artificial intelligence (AI) software allows a fully automated quantification with high accuracy (6).

The aim of the present study was to compare the accuracy of an automated CAC quantification versus manual quantification on baseline LDCT (3) and assess their predictive value for all-cause mortality at 12 years.

## Material and Methods

The study population is represented by the subset of participants of the Multicentric Italian Lung Detection (MILD) trial volunteers (ClinicalTrials.gov Identifier: NCT02837809) who underwent a baseline LDCT scan from September 2005 to September 2006 analysed in a previous paper on manual CAC evaluation (3).

LDCTs were acquired by a 16–detector row CT scanner (Somatom Sensation 16; Siemens Medical Solutions, Forchheim, Germany): the whole chest volume was scanned during one deep inspiratory breath-hold, without the use of contrast medium and with the following scanning parameters: tube voltage, 120 kV; effective tube current, 30 mAs; individual detector collimation, 0.75 mm; gantry rotation time, 0.5 second; and pitch, 1.5. Neither electrocardiographic triggering nor dose-modulation systems were used. Images were reconstructed as follows: one-millimeter-thick sections were reconstructed with an increment of 1 mm (medium-sharp kernel, B50f), and 5-mm-thick sections were reconstructed with an increment of 5 mm (medium-smooth kernel, B30f).

The manual CAC evaluation (hereafter named as mCAC) has been described previously (3); briefly, CT images were transferred to a workstation (Leonardo; Siemens Medical Solutions) and analyzed by one operator with 5 years of experience in cardiac imaging. mCAC was performed on the 5-mm-thick images dataset using a dedicated software (CaScore; Siemens Medical Solutions) (3).

For the automated evaluation, 1-mm images were transferred to a dedicated graphic station (Alienware Area 51 R6 equipped with Dual NVIDIA GeForce RTX 2080 OC graphics) and analyzed

using a fully automated AI software (AVIEW, Coreline Soft). Automated CAC (aCAC) was measured with a scoring tool based on a 3-dimensional U-net architecture.

The vital status was obtained through the Istituto Nazionale di Statistica (ISTAT, SIATEL 2.0 platform). Participants accumulated person-years of follow-up from the date of baseline until death or the date of the last follow-up as of November 2021. Both mCAC and aCAC scores were stratified in four categories: 0, 0.1-19.9, 20-399, and  $\geq 400$ . Categorical variables were reported as numbers and percentages, whereas continuous variables as medians with interquartile ranges (IQRs); associations were evaluated by the chi-square test for categorical data and by the Mann-Whitney U test for continuous variables. Correlation between mCAC and aCAC categories was estimated by the Intraclass Correlation Coefficient (ICC) with 95% Confidence Interval (CI). Kaplan-Meier curves for 12-year all-cause survival were reported (a) in strata of mCAC score in all participants, (b) in strata of aCAC score in all participants, and (c) in strata of aCAC score (0.1-19.9 vs. 20-399) among negative mCAC score participants. Comparisons were tested by Log-Rank test for trend. Analyses were performed using the Statistical Analysis System Software (Release SAS:9.04; SAS Institute, Cary, North Carolina, USA) and R Statistical software (R Studio).

## Results

Of the initial cohort of 1159 participants, 30 (2.6%) failed the automated AI software evaluation due to LDCT features. Final study population consisted in 1129 participants: median age was 57 years, 68% were males and 65% were current smokers (Table 1).

aCAC score was 0 in 90 participants (8.0%), 0.1-19.9 in 546 (48.4%), 20-399 in 393 (34.8%), and  $\geq 400$  in 100 cases (8.9%); mCAC score was 0 in 613 participants (54.3%), 0.1-19.9 in 160 (14.2%), 20-399 in 279 (24.7%), and  $\geq 400$  in 77 cases (6.8%).

Subjects with aCAC score  $\geq 20$  were older ( $p$ -value  $< 0.0001$ ) and the frequency of aCAC  $\geq 20$  was significantly higher in males than in females (12% of males had aCAC  $\geq 400$  compared to 2% of females). Non-significant differences were observed for smoking status (Table 1).

The comparison of aCAC with mCAC scores, showed a high correlation (ICC = 0.887 (0.874-0.899)). Of 613 negative mCAC score, 87.6% had aCAC score  $> 0$ , and 14.0%  $> 20$ . The 2 cases  $\geq 400$  proved to be failures of AI calculation.

Twelve-year all-cause mortality was 8.9% overall (100/1129), 3.3% with aCAC = 0, 5.1% = 0.1-19.9, 11.5% = 20-399, 24% =  $\geq 400$ , 4.9% with aCAC  $< 20$ , and 14.0% with aCAC  $> 20$ .

Survival curves show that participants with CAC score  $> 20$  had a higher risk of 12-year all-cause mortality, both with mCAC scoring (Figure 1A) and with aCAC scoring (Figure 1B), but the high-risk group was 44% of the cohort with aCAC vs. 32% with mCAC. Volunteers with aCAC score  $< 20$

had a much better survival than subjects with aCAC score  $\geq 20$  (95% vs. 86%, Log-rank p-value $<0.0001$ ) or  $\geq 400$  (95% vs. 76%, Log-Rank p-value $<0.0001$ ). Focused analysis of 535 individuals with negative mCAC and positive aCAC score (Figure 1C), showed that aCAC identified a subset of 86 volunteers with a significantly poorer survival of 86% (aCAC 20-399, p=0.0007).

## Discussion

Our findings showed that CAC quantification can be accurately performed with a fully automated software. The correlation between the two measurements was high (ICC of 0.887), despite the automatic software tended to overestimate CAC (12). In addition, manual assessment of CAC classified 54% of volunteers as negative, while only 8% were classified as negative by the automated score.

The automated AI-driven analysis of coronary artery calcifications of MILD volunteers confirmed our previous observation that CAC is an independent predictor for all-cause mortality (3), and that CAC  $> 400$  is associated with a significantly worse survival at 12 years. These results are in keeping with a case-cohort study from the NELSON trial demonstrating that higher strata of CAC score had a significantly higher risk of coronary events compared with negative (10), and with a meta-analysis of 6 LCS trials showing that subjects with high CAC values had more than two-fold increased relative risk of all-cause mortality (11). The extended follow-up of MILD trial revealed a 3-fold higher mortality risk for aCAC  $>20$  and 5-fold risk for aCAC  $>400$  (p $<0.0001$  and p $<0.0001$  respectively). As reported in Figure 2C, aCAC stratified false negative mCAC volunteers in two categories with significantly different risk profile.

AI has been increasingly used in diagnostic radiology and different reports have compared fully or mainly automated CAC scoring vs. manual quantification, showing promising results (6, 13-15). A recent study demonstrated that a deep learning-based automated CAC assessment could be reliably used for CV risk stratification on non-gated chest CT, with both 1- and 3-mm reconstruction (16). Major advantages of an automated CAC quantification include a potential increase in measurement reproducibility, while reducing the time-consuming process of manual coronary arteries segmentation.

This study has few limitations. First, the retrospective design is prone to confounding factors, such as selection of patients. Second, images reconstruction thickness might have affected the categorical agreement. Indeed, CAC quantification on 1-mm LDCT is known to overscore calcium as compared to 5-mm LDCT, suggesting that part of the CAC overestimation from AI analysis might be due to different slice thickness rather than to a true software misclassification.

In conclusion, a fully automated quantification of the CAC score by means of an AI software could be safely performed on chest LDCTs for the purpose of mortality risk stratification within lung cancer screening activities.

### Funding sources

The MILD trial was supported by grants from the Italian Ministry of Health (RF 2004), the Italian Association for Cancer Research (AIRC 2004 IG 1227 and AIRC 5xmille IG 12162), Fondazione Cariplo (2004-1560), and the National Cancer Institute (EDRN UO1 CA166905). The sponsors had no role in conducting and interpreting the study.

### References

1. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol. *J Am Coll Cardiol*. 2019;73:e285–e350.
2. Gendarme S, Goussault H, Assié JB, Taleb C, Chouaïd C, Landre T. Impact on All-Cause and Cardiovascular Mortality Rates of Coronary Artery Calcifications Detected during Organized, Low-Dose, Computed-Tomography Screening for Lung Cancer: Systematic Literature Review and Meta-Analysis. *Cancers (Basel)*. 2021 Mar 28;13(7):1553. doi: 10.3390/cancers13071553. PMID: 33800614; PMCID: PMC8036563.
3. Sverzellati N, Cademartiri F, Bravi F, Martini C, Gira FA, Maffei E, et al. Relationship and prognostic value of modified coronary artery calcium score, FEV1, and emphysema in lung cancer screening population: the MILD trial. *Radiology*. 2012 Feb;262(2):460-7. doi: 10.1148/radiol.11110364.
4. Kim SM, Chung MJ, Lee KS, Choe YH, Yi CA, Choe BK. Coronary calcium screening using low-dose lung cancer screening: effectiveness of MDCT with retrospective reconstruction. *AJR Am J Roentgenol*. 2008 Apr;190(4):917-22. doi: 10.2214/AJR.07.2979. PMID: 18356437.
5. Arcadi T, Maffei E, Sverzellati N, Mantini C, Guaricci AI, Tedeschi C, Martini C, La Grutta L, Cademartiri F. Coronary artery calcium score on low-dose computed tomography for lung cancer screening. *World J Radiol*. 2014 Jun 28;6(6):381-7. doi: 10.4329/wjr.v6.i6.381. PMID: 24976939; PMCID: PMC4072823.

6. Vonder M, Zheng S, Dorrius MD, van der Aalst CM, de Koning HJ, Yi J, et al. Deep Learning for Automatic Calcium Scoring in Population-Based Cardiovascular Screening. *JACC Cardiovasc Imaging*. 2022 Feb;15(2):366-367. doi: 10.1016/j.jcmg.2021.07.012.
7. Yip R, Jirapatnakul A, Hu M, Chen X, Han D, Ma T, et al. Added benefits of early detection of other diseases on low-dose CT screening. *Transl Lung Cancer Res*. 2021 Feb;10(2):1141-1153. doi: 10.21037/tlcr-20-746
8. Pinsky PF, Lynch DA, Gierada DS. Incidental Findings on Low-Dose CT Scan Lung Cancer Screenings and Deaths From Respiratory Diseases. *Chest*. 2022 Apr;161(4):1092-1100. doi: 10.1016/j.chest.2021.11.015.
9. Mets OM, de Jong PA, Prokop M. Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. *JAMA*. 2012 Oct 10;308(14):1433-4. doi: 10.1001/jama.2012.12656.
10. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer. *AJR Am J Roentgenol*. 2012 Mar;198(3):505-11. doi: 10.2214/AJR.10.5577.
11. Gendarme S, Goussault H, Assié JB, Taleb C, Chouaïd C, Landre T. Impact on All-Cause and Cardiovascular Mortality Rates of Coronary Artery Calcifications Detected during Organized, Low-Dose, Computed-Tomography Screening for Lung Cancer: Systematic Literature Review and Meta-Analysis. *Cancers (Basel)*. 2021 Mar 28;13(7):1553. doi: 10.3390/cancers13071553.
12. Christensen JL, Sharma E, Gorvitovskaia AY, Watts JP Jr, Assali M, Neverson J, Wu WC, Choudhary G, Morrison AR. Impact of Slice Thickness on the Predictive Value of Lung Cancer Screening Computed Tomography in the Evaluation of Coronary Artery Calcification. *J Am Heart Assoc*. 2019 Jan 8;8(1):e010110. doi: 10.1161/JAHA.118.010110. PMID: 30620261; PMCID: PMC6405734.
13. Wang W, Wang H, Chen Q, Zhou Z, Wang R, Wang H, Zhang N, Chen Y, Sun Z, Xu L. Coronary artery calcium score quantification using a deep-learning algorithm. *Clin Radiol*. 2020 Mar;75(3):237.e11-237.e16. doi: 10.1016/j.crad.2019.10.012. Epub 2019 Nov 11. PMID: 31718789.

14. de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Isgum I. Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. *IEEE Trans Med Imaging*. 2019 Sep;38(9):2127-2138. doi: 10.1109/TMI.2019.2899534. Epub 2019 Feb 18. PMID: 30794169.
15. Gautam N, Saluja P, Malkawi A, Rabbat MG, Al-Mallah MH, Pontone G, Zhang Y, Lee BC, Al'Aref SJ. Current and Future Applications of Artificial Intelligence in Coronary Artery Disease. *Healthcare (Basel)*. 2022 Jan 26;10(2):232. doi: 10.3390/healthcare10020232. PMID: 35206847; PMCID: PMC8872080.
16. Xu C, Guo H, Xu M, Duan M, Wang M, Liu P, Luo X, Jin Z, Liu H, Wang Y. Automatic coronary artery calcium scoring on routine chest computed tomography (CT): comparison of a deep learning algorithm and a dedicated calcium scoring CT. *Quant Imaging Med Surg*. 2022 May;12(5):2684-2695. doi: 10.21037/qims-21-1017. PMID: 35502379; PMCID: PMC9014138.

**Table 1.** Characteristics and outcomes of study population according to Automated CAC Scores.

		Total	Automated CAC Score				p-value
			0	0.1-19.9	20-399	400+	
		1129	90 (8.0%)	546 (48.4%)	393 (34.8%)	100 (8.9%)	
<b>Age</b>	<b>Median (IQR)</b>	57 (53-61)	54.5 (51-59)	55 (52-59)	59 (54-63)	61 (56-66.5)	<0.0001
<b>Sex</b>	<b>Male</b>	770 (68.2%)	44 (5.7%)	312 (40.5%)	321 (41.7%)	93 (12.1%)	<0.0001
	<b>Female</b>	359 (31.8%)	46 (12.8%)	234 (65.2%)	72 (20.1%)	7 (2.0%)	
<b>Smoking Status</b>	<b>Ex Smoker</b>	396 (35.1%)	36 (9.1%)	178 (45.0%)	139 (35.15)	43 (10.9%)	0.1578
	<b>Current Smoker</b>	733 (64.9%)	54 (7.4%)	368 (50.2%)	254 (34.7%)	57 (7.8%)	
<b>mCAC Score</b>	<b>0</b>	613 (54.3%)	76 (12.4%)	449 (73.3%)	86 (14.0%)	2 (0.3%)	<0.0001
	<b>0.1-19.9</b>	160 (14.2%)	4 (2.5%)	61 (38.1%)	95 (59.4%)	0	
	<b>20-399</b>	279 (24.7%)	10 (3.6%)	34 (12.25)	195 (69.9%)	40 (14.3%)	
	<b>400+</b>	77 (6.8%)	0	2 (2.6%)	17 (22.1%)	58 (75.3%)	
<b>12-year all-cause mortality</b>		100 (8.9%)	3 (3.3%)	28 (5.1%)	45 (11.5%)	24 (24.0%)	<0.0001



**Figure 1.** Kaplan-Meier survival curves for all-cause mortality: A) stratified by mCAC score in all 1129 participants; b) stratified by aCAC score in all 1129 participants; and C) stratified by aCAC score in participants with negative mCAC score and positive automated aCAC score.

