Supplementary Information

AI-based opportunistic coronary artery calcium assessment on non-gated chest CT: a systematic review and meta-analysis

Yu-Chuan Chen, Yun-Chu Wang, Marly van Assen, and Chih-Yu Yang

Contents

Supplementary Section 1. Search, eligibility, and protocol	2
Supplementary Section 2. Included reports, risk of bias, and applicability	13
Supplementary Section 3. Threshold diagnostic performance and certainty	36
Supplementary Section 4. Agreement of AI-derived CAC categories with ECG-gated CT	46
Supplementary Section 5. Error attribution, image-domain effects, and cross-domain measurement calibration	51
Supplementary Section 6. Subgroups and meta-regression	55
Supplementary Section 7. Prognostic evidence and certainty	60
Supplementary Section 8. Implementation evidence and certainty	65
Supplementary Section 9. Reporting checklists	71
Supplementary Section 10. Reports excluded after full-text assessment and reports not retrieved	78
Supplementary Data 1. Description of the accompanying workbook	90

Search date: 2 October 2026

Supplementary Section 1
Search, eligibility, and protocol

Registered protocol: INPLASY202660006 · DOI 10.37766/inplasy2026.6.0006

Supplementary Note 1. Full search strategies

The complete strategies are reproduced below. No date or language limits were applied. Search B was designed to identify implementation or preventive-care reports that might not mention AI in the title or abstract.

PubMed Search A: Core AI-CAC search

```
(
  (
    ("Vascular Calcification"[Mesh] OR "Calcinosis"[Mesh])
    AND
    (coronary[tiab] OR "Coronary Artery Disease"[Mesh])
  )
  OR "coronary artery calcium"[tiab]
  OR "coronary calcium"[tiab]
  OR "coronary artery calcification"[tiab]
  OR "coronary artery calcifications"[tiab]
  OR "coronary calcification"[tiab]
  OR "coronary calcifications"[tiab]
  OR "calcified coronary plaque"[tiab]
  OR "coronary calcified plaque"[tiab]
  OR Agatston[tiab]
  OR CACS[tiab]
  OR (CAC[tiab] AND (score[tiab] OR scores[tiab] OR scoring[tiab] OR quantification[tiab] OR quantified[tiab] OR category[tiab] OR categories[tiab] OR categorization[tiab] OR
classification[tiab]))
)
AND
(
  "Tomography, X-Ray Computed"[Mesh]
  OR "computed tomography"[tiab]
  OR CT[tiab]
  OR "chest CT"[tiab]
  OR "thoracic CT"[tiab]
  OR "low-dose CT"[tiab]
  OR "low dose CT"[tiab]
  OR LDCT[tiab]
  OR "lung cancer screening"[tiab]
  OR "non-gated"[tiab]
```

```
  OR nongated[tiab]
  OR ungated[tiab]
  OR "non ECG gated"[tiab]
  OR "non-ECG-gated"[tiab]
  OR "non-electrocardiogram-gated"[tiab]
  OR "noncardiac"[tiab]
  OR "non-cardiac"[tiab]
  OR opportunistic[tiab]
  OR incidental[tiab]
  OR "routine CT"[tiab]
  OR "routine computed tomography"[tiab]
  OR "PET/CT"[tiab]
)
AND
(
  "Artificial Intelligence"[Mesh]
  OR "Machine Learning"[Mesh]
  OR "Deep Learning"[Mesh]
  OR "Neural Networks, Computer"[Mesh]
  OR "Image Processing, Computer-Assisted"[Mesh]
  OR "Diagnosis, Computer-Assisted"[Mesh]
  OR "Algorithms"[Mesh]
  OR "artificial intelligence"[tiab]
  OR AI[tiab]
  OR "machine learning"[tiab]
  OR "deep learning"[tiab]
  OR "neural network"[tiab]
  OR "neural networks"[tiab]
  OR "convolutional neural network"[tiab]
  OR "convolutional neural networks"[tiab]
  OR CNN[tiab]
  OR "U-Net"[tiab]
  OR UNet[tiab]
  OR automated[tiab]
  OR automatic[tiab]
  OR "fully automated"[tiab]
  OR "semi-automated"[tiab]
  OR algorithm*[tiab]
  OR software[tiab]
  OR "computer-aided"[tiab]
  OR "computer assisted"[tiab]
```

```
  OR "image processing"[tiab]
  OR HealthCCSng[tiab]
  OR "DL-CAC"[tiab]
  OR "DL-CACS"[tiab]
  OR "AI-CAC"[tiab]
  OR "AVIEW CAC"[tiab]
  OR "ClearRead CT CAC"[tiab]
  OR Nanox[tiab]
  OR "Zebra Medical"[tiab]
  OR Coreline[tiab]
  OR Riverain[tiab]
)
NOT (animals[mh] NOT humans[mh])
```

PubMed Search B: Implementation / preventive-care supplementary search

```
(
  (
    ("Vascular Calcification"[Mesh] OR "Calcinosis"[Mesh])
    AND
    (coronary[tiab] OR "Coronary Artery Disease"[Mesh])
  )
  OR "coronary artery calcium"[tiab]
  OR "coronary calcium"[tiab]
  OR "coronary artery calcification"[tiab]
  OR "coronary artery calcifications"[tiab]
  OR "coronary calcification"[tiab]
  OR "coronary calcifications"[tiab]
  OR "calcified coronary plaque"[tiab]
  OR "coronary calcified plaque"[tiab]
  OR Agatston[tiab]
  OR CACS[tiab]
  OR (CAC[tiab] AND (score[tiab] OR scores[tiab] OR scoring[tiab] OR quantification[tiab] OR quantified[tiab] OR category[tiab] OR categories[tiab] OR categorization[tiab] OR
classification[tiab]))
)
AND
(
  "chest CT"[tiab]
  OR "thoracic CT"[tiab]
  OR "low-dose CT"[tiab]
  OR "low dose CT"[tiab]
  OR LDCT[tiab]
  OR "lung cancer screening"[tiab]
```

```
  OR "non-gated"[tiab]
  OR nongated[tiab]
  OR ungated[tiab]
  OR "non ECG gated"[tiab]
  OR "non-ECG-gated"[tiab]
  OR "non-electrocardiogram-gated"[tiab]
  OR "noncardiac"[tiab]
  OR "non-cardiac"[tiab]
  OR opportunistic[tiab]
  OR incidental[tiab]
  OR "routine CT"[tiab]
  OR "routine computed tomography"[tiab]
)
AND
(
  statin*[tiab]
  OR lipid*[tiab]
  OR "primary prevention"[tiab]
  OR prevent*[tiab]
  OR notification[tiab]
  OR notify[tiab]
  OR notified[tiab]
  OR report*[tiab]
  OR "radiology report"[tiab]
  OR "clinical utility"[tiab]
  OR implement*[tiab]
  OR "risk stratification"[tiab]
  OR "risk reclassification"[tiab]
  OR management[tiab]
  OR referral[tiab]
  OR "downstream testing"[tiab]
  OR follow-up[tiab]
  OR "quality improvement"[tiab]
)
NOT (animals[mh] NOT humans[mh])
```

Web of Science Search A: Core AI-CAC search

```
TS=(
  (
    "coronary artery calcium"
    OR "coronary calcium"
    OR "coronary artery calcif*"
```

```
  OR "coronary calcif*"
  OR "calcified coronary plaque*"
  OR "coronary calcified plaque*"
  OR Agatston
  OR CACS
  OR (CAC NEAR/3 (score* OR scoring OR quantif* OR categor* OR classification))
)
AND
(
  "computed tomography"
  OR CT
  OR "chest CT"
  OR "thoracic CT"
  OR "low-dose CT"
  OR "low dose CT"
  OR LDCT
  OR "lung cancer screen*"
  OR "non-gated"
  OR nongated
  OR ungated
  OR "non ECG gated"
  OR "non-ECG-gated"
  OR "non-electrocardiogram-gated"
  OR noncardiac
  OR "non-cardiac"
  OR opportunistic
  OR incidental
  OR "routine CT"
  OR "routine computed tomography"
  OR "PET/CT"
)
AND
(
  "artificial intelligence"
  OR AI
  OR "machine learning"
  OR "deep learning"
  OR "neural network*"
  OR "convolutional neural network*"
  OR CNN
  OR "U-Net"
```

```
   OR UNet
   OR automated
   OR automatic
   OR "fully automated"
   OR "semi-automated"
   OR algorithm*
   OR software
   OR "computer-aided"
   OR "computer assisted"
   OR "image processing"
   OR HealthCCSng
   OR "DL-CAC"
   OR "DL-CACS"
   OR "AI-CAC"
   OR "AVIEW CAC"
   OR "ClearRead CT CAC"
   OR Nanox
   OR "Zebra Medical"
   OR Coreline
   OR Riverain
  )
)
```

Web of Science Search B: Implementation / preventive-care supplementary search

```
TS=(
 (
  "coronary artery calcium"
  OR "coronary calcium"
  OR "coronary artery calcif*"
  OR "coronary calcif*"
  OR "calcified coronary plaque*"
  OR "coronary calcified plaque*"
  OR Agatston
  OR CACS
  OR (CAC NEAR/3 (score* OR scoring OR quantif* OR categor* OR classification))
 )
 AND
 (
  "chest CT"
  OR "thoracic CT"
  OR "low-dose CT"
  OR "low dose CT"
```

```
  OR LDCT
  OR "lung cancer screen*"
  OR "non-gated"
  OR nongated
  OR ungated
  OR "non ECG gated"
  OR "non-ECG-gated"
  OR "non-electrocardiogram-gated"
  OR noncardiac
  OR "non-cardiac"
  OR opportunistic
  OR incidental
  OR "routine CT"
  OR "routine computed tomography"
 )
 AND
 (
  statin*
  OR lipid*
  OR "primary prevention"
  OR prevent*
  OR notification
  OR notify
  OR notified
  OR report*
  OR "radiology report"
  OR "clinical utility"
  OR implement*
  OR "risk stratification"
  OR "risk reclassification"
  OR management
  OR referral
  OR "downstream testing"
  OR follow-up
  OR "quality improvement"
 )
)
```

CENTRAL Search A: Core AI-CAC search

#1 MeSH descriptor: [Vascular Calcification] explode all trees
#2 MeSH descriptor: [Calcinosis] explode all trees
#3 MeSH descriptor: [Coronary Artery Disease] explode all trees

#4 ((coronary NEAR/3 (calcium OR calcif*)) OR "coronary artery calcium" OR "coronary calcium" OR "coronary artery calcification" OR "coronary artery calcifications" OR "coronary calcification" OR "coronary calcifications" OR "calcified coronary plaque*" OR "coronary calcified plaque*" OR Agatston OR CACS OR (CAC NEAR/3 (score* OR scoring OR quantif* OR categor* OR classification))):ti,ab,kw
#5 ((#1 OR #2) AND #3) OR #4
#6 MeSH descriptor: [Tomography, X-Ray Computed] explode all trees
#7 ("computed tomography" OR CT OR "chest CT" OR "thoracic CT" OR "low-dose CT" OR "low dose CT" OR LDCT OR "lung cancer screen*" OR "non-gated" OR nongated OR ungated OR "non ECG gated" OR "non-ECG-gated" OR "non-electrocardiogram-gated" OR noncardiac OR "non-cardiac" OR opportunistic OR incidental OR "routine CT" OR "routine computed tomography" OR "PET/CT"):ti,ab,kw
#8 #6 OR #7
#9 MeSH descriptor: [Artificial Intelligence] explode all trees
#10 MeSH descriptor: [Machine Learning] explode all trees
#11 MeSH descriptor: [Deep Learning] explode all trees
#12 MeSH descriptor: [Neural Networks, Computer] explode all trees
#13 MeSH descriptor: [Image Processing, Computer-Assisted] explode all trees
#14 MeSH descriptor: [Diagnosis, Computer-Assisted] explode all trees
#15 ("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional neural network*" OR CNN OR "U-Net" OR UNet OR automated OR automatic OR "fully automated" OR "semi-automated" OR algorithm* OR software OR "computer-aided" OR "computer assisted" OR "image processing" OR HealthCCSng OR "DL-CAC" OR "DL-CACS" OR "AI-CAC" OR "AVIEW CAC" OR "ClearRead CT CAC" OR Nanox OR "Zebra Medical" OR Coreline OR Riverain):ti,ab,kw
#16 #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15
#17 #5 AND #8 AND #16

CENTRAL Search B: Implementation / preventive-care supplementary search

#18 (statin* OR lipid* OR "primary prevention" OR prevent* OR notification OR notify OR notified OR report* OR "radiology report" OR "clinical utility" OR implement* OR "risk stratification" OR "risk reclassification" OR management OR referral OR "downstream testing" OR follow-up OR "quality improvement"):ti,ab,kw
#19 #5 AND #8 AND #18

Embase.com Search A

```
(
  'coronary artery calcium':ti,ab,kw
  OR 'coronary calcium':ti,ab,kw
  OR 'coronary artery calcif*':ti,ab,kw
  OR 'coronary calcif*':ti,ab,kw
  OR 'calcified coronary plaque*':ti,ab,kw
  OR 'coronary calcified plaque*':ti,ab,kw
  OR agatston:ti,ab,kw
  OR cacs:ti,ab,kw
  OR (cac NEAR/3 (score* OR scoring OR quantif* OR categor* OR classification)):ti,ab,kw
)
AND
(
  'computed tomography':ti,ab,kw
  OR ct:ti,ab,kw
  OR 'chest ct':ti,ab,kw
```

```
  OR 'thoracic ct':ti,ab,kw
  OR 'low-dose ct':ti,ab,kw
  OR 'low dose ct':ti,ab,kw
  OR ldct:ti,ab,kw
  OR 'lung cancer screen*':ti,ab,kw
  OR 'non-gated':ti,ab,kw
  OR nongated:ti,ab,kw
  OR ungated:ti,ab,kw
  OR 'non ECG gated':ti,ab,kw
  OR 'non-ECG-gated':ti,ab,kw
  OR 'non-electrocardiogram-gated':ti,ab,kw
  OR noncardiac:ti,ab,kw
  OR 'non-cardiac':ti,ab,kw
  OR opportunistic:ti,ab,kw
  OR incidental:ti,ab,kw
  OR 'routine CT':ti,ab,kw
  OR 'routine computed tomography':ti,ab,kw
  OR 'PET/CT':ti,ab,kw
)
AND
(
  'artificial intelligence':ti,ab,kw
  OR ai:ti,ab,kw
  OR 'machine learning':ti,ab,kw
  OR 'deep learning':ti,ab,kw
  OR 'neural network*':ti,ab,kw
  OR 'convolutional neural network*':ti,ab,kw
  OR cnn:ti,ab,kw
  OR 'U-Net':ti,ab,kw
  OR unet:ti,ab,kw
  OR automated:ti,ab,kw
  OR automatic:ti,ab,kw
  OR 'fully automated':ti,ab,kw
  OR 'semi-automated':ti,ab,kw
  OR algorithm*:ti,ab,kw
  OR software:ti,ab,kw
  OR 'computer-aided':ti,ab,kw
  OR 'computer assisted':ti,ab,kw
  OR 'image processing':ti,ab,kw
  OR healthccsng:ti,ab,kw
  OR 'DL-CAC':ti,ab,kw
```

```
  OR 'DL-CACS':ti,ab,kw
  OR 'AI-CAC':ti,ab,kw
  OR 'AVIEW CAC':ti,ab,kw
  OR 'ClearRead CT CAC':ti,ab,kw
  OR nanox:ti,ab,kw
  OR 'Zebra Medical':ti,ab,kw
  OR coreline:ti,ab,kw
  OR riverain:ti,ab,kw
)
```

Supplementary Table 1. Protocol and analytical amendments

Date	Original plan	Amendment	Rationale
25 June 2026	Search bibliographic databases and the additional information sources listed in the registered protocol.	The final identification set comprised PubMed, Embase, Web of Science Core Collection, and CENTRAL; the additional sources listed in the protocol were not run as separate searches.	To report the information sources actually used and align the review record with the reproducible search strategies.
6–7 August 2026	No formal certainty-of-evidence assessment was specified in the registered protocol.	Post hoc GRADE assessments were added for paired-reference diagnostic accuracy, prognostic-factor associations with MACE, and the randomised implementation evidence for six-month statin prescribing.	To provide transparent, outcome-specific certainty assessments across the principal diagnostic-accuracy, prognostic, and implementation evidence bodies.

CENTRAL, Cochrane Central Register of Controlled Trials; GRADE, Grading of Recommendations Assessment, Development and Evaluation; MACE, major adverse cardiovascular events.

Supplementary Section 2. Included reports, risk of bias, and applicability

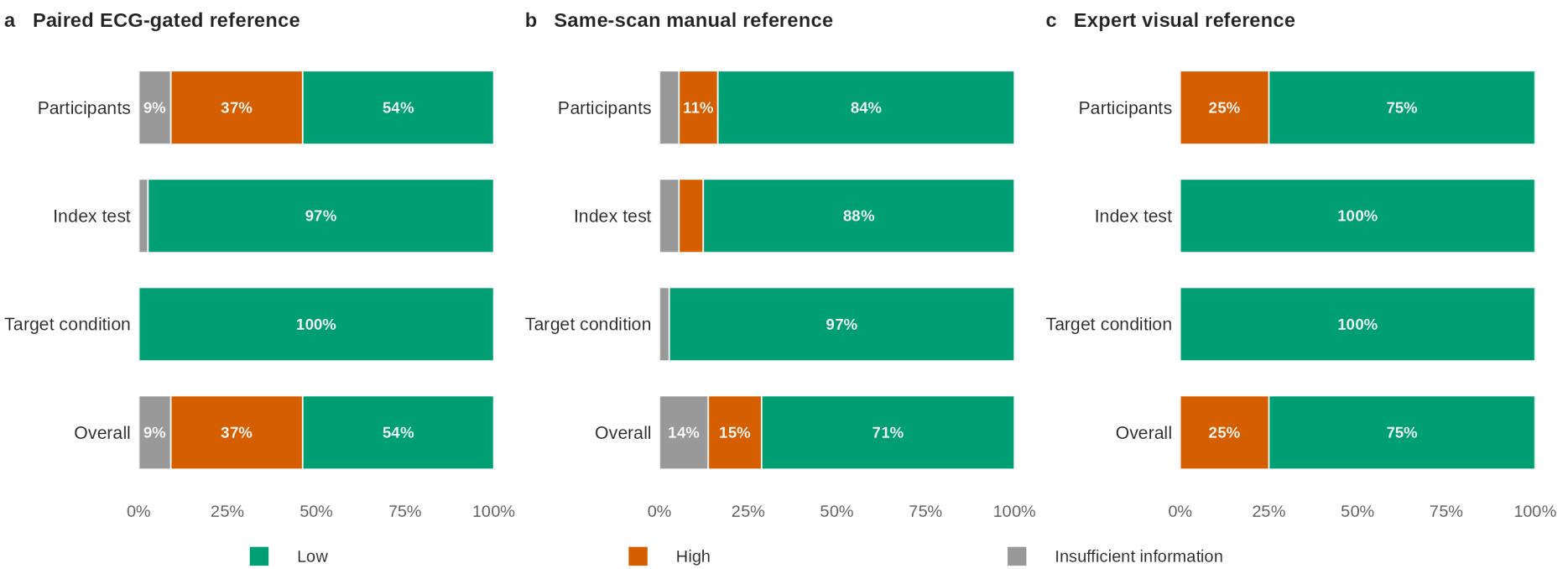
This section presents report characteristics and structured risk-of-bias and applicability summaries. Complete technical acquisition, validation, and report-linkage data are provided in Supplementary Data 1.

Supplementary Table 2. Synthesis questions and ideal test accuracy trials

The ideal trial defines the benchmark for QUADAS-3 risk-of-bias and applicability judgments; it is not an additional eligibility criterion

Synthesis question	Participants and setting	Index test and clinical role	Target condition and reference standard	Thresholds, unit, and ideal analysis
Paired ECG-gated reference	Adults undergoing routine noncardiac, non-gated thoracic or chest CT who also underwent a paired dedicated ECG-gated calcium-scoring CT; no selection by CAC status or test result.	A fixed, fully automated AI-CAC workflow applied to routine non-gated CT, without reference-informed correction; opportunistic cardiovascular risk stratification.	Agatston CAC category defined on a separate, paired ECG-gated calcium-scoring CT using a standardised quantitative reference procedure blinded to the AI result.	CAC >0, ≥100, ≥300, and ≥400 as available; participant-level paired 2×2 analysis including failures and uninterpretable results according to a prespecified rule.
Same-scan manual reference	Adults undergoing routine noncardiac, non-gated thoracic or chest CT; no selection by CAC status, AI success, or manual result.	A fixed, fully automated AI-CAC workflow applied to the same non-gated CT used by the reference reader; technical same-image quantification and categorisation.	Manual or expert quantitative CAC assessment on the same non-gated CT using a prespecified method, performed without knowledge of the AI result.	CAC >0, ≥100, ≥300, and ≥400 as available; participant-level paired 2×2 analysis with all eligible scans and prespecified handling of failures.
Expert visual reference	Adults undergoing routine noncardiac, non-gated thoracic or chest CT in an intended-use population; no result-dependent sampling.	A fixed, fully automated AI-CAC workflow applied to routine non-gated CT; opportunistic detection of visually evident CAC.	A blinded, prespecified expert visual CAC assessment on the same non-gated CT, independent of the AI result.	CAC >0 for the retained diagnostic synthesis; participant-level paired analysis including all eligible and uninterpretable examinations.

CAC, coronary artery calcium; CT, computed tomography; ECG, electrocardiogram; AI, artificial intelligence. The paired ECG-gated, same-scan manual, and expert visual reference profiles were synthesised separately.



Supplementary Figure 1. QUADAS-3 concerns regarding applicability. Bars show unweighted distributions across selected diagnostic-accuracy estimates within each synthesis question: a, paired ECG-gated reference (78 estimates); b, same-scan manual reference (73 estimates); c, expert visual reference (4 estimates). Applicability was assessed for Participants, Index test, and Target condition, with an overall judgment derived from those three domains. Percentages use selected estimates, rather than studies or participants, as the denominator.

Supplementary Table 3A. Risk-of-bias judgments by report and selected estimates

Identical domain-level patterns within a report and evidence profile are collapsed; assessment scopes list the combined endpoints

A. Diagnostic accuracy — QUADAS-3

Study	Assessment scope	Participants	Index test	Target condition	Analysis	Overall
Paired ECG-gated reference						
Dobrolinska et al, 2023	CAC ≥100, CAC ≥400	×	●	●	●	×
Dobrolinska et al, 2024	CAC >0, CAC ≥100, CAC ≥400	×	●	●	●	×
Eng et al, 2021	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●	●
Gokmen et al, 2026	CAC ≥400	○	●	○	○	○
Hagopian et al, 2025	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●	●
Kang et al, 2023	CAC >0, CAC ≥100, CAC ≥400	×	●	●	●	×
Kim et al, 2025a	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●	●
Kim et al, 2025b	CAC ≥100, CAC >0, CAC ≥300; SNUH cohort (n = 652)	×	●	●	●	×
Kim et al, 2025b	CAC ≥300, CAC ≥100, CAC >0; Stanford-COCA cohort (n = 199)	●	●	○	●	○
Ludwig et al, 2026	CAC ≥400	○	●	○	●	○
McKinney et al, 2025	CAC ≥100, CAC >0, CAC ≥400	×	●	●	●	×
Morf et al, 2022	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●	●
Oh et al, 2024	CAC ≥100, CAC >0, CAC ≥300	●	●	●	●	●
Pieszko et al, 2022	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●	●
Pieszko et al, 2023	CAC >0, CAC ≥100, CAC ≥400	×	●	●	●	×
Xu et al, 2021	CAC ≥100, CAC >0, CAC ≥400	×	○	●	○	×
Xu et al, 2022	CAC >0, CAC ≥300, CAC ≥100	×	●	●	●	×
Xu et al, 2025	CAC ≥100, CAC >0, CAC ≥400	×	○	●	●	×
Yin et al, 2024	CAC >0, CAC ≥100, CAC ≥400	●	○	●	×	×
Yu et al, 2022	CAC >0, CAC ≥100, CAC ≥300	×	●	●	●	×
van Assen et al, 2021	CAC >0, CAC ≥100, CAC ≥400	●	●	×	●	×
Liu Y et al, 2026	CAC ≥100, CAC ≥400	×	●	●	○	×
Same-scan manual reference						
Anifowose et al, 2026	CAC ≥100, CAC ≥400	×	×	●	●	×

Supplementary Table 3A continued

Study	Assessment scope	Participants	Index test	Target condition	Analysis	Overall
Balbi et al, 2023	CAC ≥100, CAC ≥400	●	●	●	●	●
Barrett et al, 2025	CAC >0, CAC ≥100, CAC ≥400	✕	●	○	✕	✕
Cano-Espinosa et al, 2018a	CAC ≥100, CAC ≥400	●	●	●	✕	✕
Cano-Espinosa et al, 2020	CAC ≥100, CAC ≥400	●	●	●	✕	✕
Chamberlin et al, 2021	CAC >0	●	●	●	✕	✕
Choi et al, 2022	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	●	✕
Gernaat et al, 2016	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	●	✕
González et al, 2016	CAC ≥100, CAC ≥400	●	○	●	○	○
Hamelink et al, 2025b	CAC >0, CAC ≥100, CAC ≥300	✕	●	●	●	✕
Huo et al, 2019	CAC >0	●	●	●	○	○
Išgum et al, 2010	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	●	✕
Išgum et al, 2012	CAC ≥100, CAC >0, CAC ≥400	●	✕	●	●	✕
Išgum et al, 2016	CAC ≥100, CAC ≥400	●	●	●	●	●
Kim et al, 2025b	CAC ≥300, CAC ≥100, CAC >0	●	●	●	●	●
Lessmann et al, 2016	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●	●
Lessmann et al, 2018	CAC ≥100	●	●	●	✕	✕
McKinney et al, 2025	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	●	✕
Miller et al, 2023	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	●	✕
Mineo et al, 2025	CAC ≥100, CAC ≥300, CAC >0	●	●	●	●	●
Rathore et al, 2025	CAC ≥100, CAC ≥400	●	○	●	●	○
Shadmi et al, 2018	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●	●
Suh et al, 2023	CAC ≥100, CAC >0, CAC ≥400	●	○	●	●	○
Takx et al, 2014	CAC ≥100, CAC >0, CAC ≥400	●	○	●	✕	✕
Zeleznik et al, 2021	CAC >0, CAC ≥300	●	●	●	●	●
de Vos et al, 2019	CAC ≥100, CAC ≥400	●	●	●	●	●
van Velzen et al, 2020	CAC >0; baseline DL calcium scoring network	●	●	●	●	●
van Velzen et al, 2020	CAC >0; evaluation cohort n = 274	●	●	✕	○	✕
Hussain et al, 2026	CAC >0, CAC ≥100	○	○	●	○	○
Ochs et al, 2026	CAC >0 from Fig. 8 category definitions	●	●	●	○	○
Ochs et al, 2026	CAC ≥100 from Fig. 8 category definitions, CAC ≥300 from Fig. 8 category definitions	●	●	○	○	○

Supplementary Table 3A continued

Study	Assessment scope	Participants	Index test	Target condition	Analysis	Overall
Shanbhag et al, 2026	Identification of any CAC	●	●	○	○	○
Hussain et al, 2026	CAC ≥400	○	○	●	●	○
Expert visual reference						
Alnasser et al, 2026	CAC >0	●	○	○	●	○
Bonatti et al, 2026	CAC >0	✕	●	●	●	✕
Chamberlin et al, 2023	CAC >0	●	●	●	✕	✕
Shear et al, 2025	CAC >0	●	●	●	●	●
Smith et al, 2023	CAC >0	●	●	●	✕	✕

B. Prognostic evidence — QUIPS

Study	Endpoint(s)	Participation	Attrition	Prognostic factor	Outcome measurement	Confounding	Analysis/reporting	Overall
Yang et al, 2026	Study-defined MACE	▲	▲	▲	▲	▲	▲	▲
Williams et al, 2024	Study-defined MACE	●	▲	●	●	●	●	▲
Pieszko et al, 2023	Study-defined MACE	✕	▲	●	●	✕	●	✕
Dobrolinska et al, 2024	Study-defined MACE	●	▲	▲	▲	✕	●	✕
Lee et al, 2025	Study-defined MACE	●	▲	●	▲	●	●	▲
Miller et al, 2023	Study-defined MACE	▲	●	●	▲	●	●	▲
Zhou et al, 2026	Study-defined MACE	●	▲	▲	▲	●	●	▲
Noureldine et al, 2026	all-cause mortality; myocardial infarction	●	▲	▲	●	▲	▲	▲
Shanbhag et al, 2026	Death or myocardial infarction	●	▲	●	●	●	●	▲
Cheng et al, 2026	All-cause mortality; Cardiovascular mortality (ICD-10 I00–I99) Categorical CAC contrasts	●	●	●	●	●	●	●
Cheng et al, 2026	All-cause mortality; Cardiovascular mortality (ICD-10 I00–I99) CAC per doubling	●	●	●	●	●	▲	▲

C. Implementation trials — RoB 2

Study	Outcome	Randomization	Deviations	Missing outcome data	Outcome measurement	Selective reporting	Overall
Dudum et al, 2026 (NOTIFY-PICTURE)	Statin prescription within 6 months	◆	●	●	◆	◆	◆
Sandhu et al, 2023 (NOTIFY-1)	Statin prescription within 6 months	●	●	●	●	●	●

Symbols: ●, Low; ▲, Moderate (QUIPS); ◆, Some concerns (RoB 2); ✕, High; ○, Insufficient information (QUADAS-3). Symbols and colours are distinct to preserve interpretation in greyscale. Overall judgments follow the applicable tool. Complete estimate-level decisions are provided in Supplementary Data 1, sheet S2. Li et al, 2023 contributed a complete 4×4 cross-classification table to the alternative-scheme cross-classification analysis (0, 1–99, 100–399, and ≥400 AU), but did not contribute a selected diagnostic-accuracy estimate and was therefore not included in the QUADAS-3 assessment.

Supplementary Table 3B. QUADAS-3 concerns regarding applicability by report and selected estimates

Identical applicability profiles within a report and reference profile are collapsed; assessment scopes list the combined thresholds

Diagnostic-accuracy evidence — QUADAS-3

Study	Assessment scope	Participants	Index test	Target condition	Overall
Paired ECG-gated reference					
Dobrolinska et al, 2023	CAC ≥100, CAC ≥400	✕	●	●	✕
Dobrolinska et al, 2024	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	✕
Eng et al, 2021	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●
Gokmen et al, 2026	CAC ≥400	○	●	●	○
Hagopian et al, 2025	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●
Kang et al, 2023	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●
Kim et al, 2025a	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Kim et al, 2025b	CAC ≥100, CAC >0, CAC ≥300; SNUH cohort (n = 652)	●	●	●	●
Kim et al, 2025b	CAC ≥300, CAC ≥100, CAC >0; Stanford-COCA cohort (n = 199)	○	●	●	○
Ludwig et al, 2026	CAC ≥400	✕	●	●	✕
McKinney et al, 2025	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Morf et al, 2022	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	✕
Oh et al, 2024	CAC ≥100, CAC >0, CAC ≥300	●	●	●	●
Pieszko et al, 2022	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	✕
Pieszko et al, 2023	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	✕
Xu et al, 2021	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	✕
Xu et al, 2022	CAC >0, CAC ≥300, CAC ≥100	●	●	●	●
Xu et al, 2025	CAC ≥100, CAC >0, CAC ≥400	✕	●	●	✕
Yin et al, 2024	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	✕
Yu et al, 2022	CAC >0, CAC ≥100, CAC ≥300	✕	●	●	✕
van Assen et al, 2021	CAC >0, CAC ≥100, CAC ≥400	○	●	●	○
Liu Y et al, 2026	CAC ≥100, CAC ≥400	✕	○	●	✕
Same-scan manual reference					
Anifowose et al, 2026	CAC ≥100, CAC ≥400	✕	✕	●	✕
Balbi et al, 2023	CAC ≥100, CAC ≥400	●	●	●	●
Barrett et al, 2025	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	✕
Cano-Espinosa et al, 2018a	CAC ≥100, CAC ≥400	●	●	●	●
Cano-Espinosa et al, 2020	CAC ≥100, CAC ≥400	●	●	●	●
Chamberlin et al, 2021	CAC >0	●	●	●	●
Choi et al, 2022	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Gernaat et al, 2016	CAC ≥100, CAC >0, CAC ≥400	●	✕	●	✕
González et al, 2016	CAC ≥100, CAC ≥400	●	●	●	●
Hamelink et al, 2025b	CAC >0, CAC ≥100, CAC ≥300	●	●	●	●
Huo et al, 2019	CAC >0	○	●	●	○
Išgum et al, 2010	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Išgum et al, 2012	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Išgum et al, 2016	CAC ≥100, CAC ≥400	●	●	●	●
Kim et al, 2025b	CAC ≥300, CAC ≥100, CAC >0	○	●	●	○
Lessmann et al, 2016	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Lessmann et al, 2018	CAC ≥100	●	●	●	●

Supplementary Table 3B continued

Study	Assessment scope	Participants	Index test	Target condition	Overall
McKinney et al, 2025	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Miller et al, 2023	CAC >0, CAC ≥100, CAC ≥400	✕	●	●	✕
Mineo et al, 2025	CAC ≥100, CAC ≥300, CAC >0	●	●	●	●
Rathore et al, 2025	CAC ≥100, CAC ≥400	●	○	●	○
Shadmi et al, 2018	CAC >0, CAC ≥100, CAC ≥400	●	●	●	●
Suh et al, 2023	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Takx et al, 2014	CAC ≥100, CAC >0, CAC ≥400	●	●	●	●
Zeleznik et al, 2021	CAC >0, CAC ≥300	●	●	●	●
de Vos et al, 2019	CAC ≥100, CAC ≥400	●	●	●	●
van Velzen et al, 2020	CAC >0	●	●	●	●
Hussain et al, 2026	CAC >0, CAC ≥100, CAC ≥400	●	○	●	○
Ochs et al, 2026	CAC >0 from Fig. 8 category definitions	●	●	●	●
Ochs et al, 2026	CAC ≥100 from Fig. 8 category definitions, CAC ≥300 from Fig. 8 category definitions	●	●	○	○
Shanbhag et al, 2026	Identification of any CAC	✕	●	●	✕
Expert visual reference					
Alnasser et al, 2026	CAC >0	✕	●	●	✕
Bonatti et al, 2026	CAC >0	●	●	●	●
Chamberlin et al, 2023	CAC >0	●	●	●	●
Shear et al, 2025	CAC >0	●	●	●	●
Smith et al, 2023	CAC >0	●	●	●	●

Symbols: ●, Low concern; ✕, High concern; ○, Insufficient information. Applicability was judged against the relevant synthesis question and its ideal test accuracy trial. Complete estimate-level decisions and rationales are provided in Supplementary Data 1, sheet S2.

Supplementary Table 4. Summary of reasons for non-Low risk-of-bias judgments

The table summarises the recurrent reasons for domain-level judgments above Low across the three assessment tools.

Tool and domain	Non-Low judgments, n	Main reason(s)
QUADAS-3: Participants	High 19; Insufficient information 3	Selected or enriched samples, exclusions of difficult scans or participants, and incompletely described recruitment or flow.
QUADAS-3: Index test	High 2; Insufficient information 9	Evaluation-informed configuration, uncertain model or workflow independence, or insufficient detail about blinding and separation from development data.
QUADAS-3: Target condition	High 2; Insufficient information 7	Uncertain reference-label provenance or blinding, missing reference assessments, or intervals between scans that could change category assignment.
QUADAS-3: Analysis	High 9; Insufficient information 9	Exclusion of failed, unassessable, or missing observations; incomplete flow reporting; or analysis denominators that could not be reconciled.
QUIPS: Participation	High 1; Moderate 2	Selected cohorts or incompletely established representativeness of the prognostic population.
QUIPS: Attrition	Moderate 8	Incomplete follow-up or insufficient reporting of missing outcome data and their handling.
QUIPS: Prognostic factor	Moderate 4	Uncertain reproducibility or validation of the AI-derived CAC measurement in the analysed population.
QUIPS: Outcome measurement	Moderate 5	Potentially incomplete or non-blinded ascertainment and heterogeneous definitions of major adverse cardiovascular events.
QUIPS: Confounding	High 2; Moderate 2	Incomplete adjustment for cardiovascular risk factors, comorbidity, or treatment.
QUIPS: Analysis/reporting	Moderate 3	Incomplete model reporting or concern about selective analytical reporting.
RoB 2: Randomization	Some concerns 1	Randomisation or allocation-concealment procedures were not described sufficiently to exclude concern.
RoB 2: Outcome measurement	Some concerns 1	Outcome assessment could have been influenced by knowledge of intervention assignment.
RoB 2: Selective reporting	Some concerns 1	The available protocol or analysis information was insufficient to exclude selective reporting.

Counts are domain-level judgments; one estimate or report may contribute to more than one domain. Complete estimate-level decisions and rationales are provided in Supplementary Data 1, sheet S2, and complete signalling-question responses in Supplementary Data 1, sheet S4.

Supplementary Table 5A. Characteristics of all included reports

All 110 included reports are shown separately and grouped by their principal evidence contribution

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Paired ECG-gated reference											
Chen et al, 2026 (ref 4)	13	China	Observational study; routine chest CT	1,213	Mean 62.4±12.5 years; Median 63.0 (IQR 55.0–71.0) years	51.2% (621/1213)	Siemens, GE, United Imaging	Non-gated chest CT; 100–120 kVp; 0.625–1.25 mm and 5 mm; smooth, sharp, and ultra-sharp kernels	Paired ECG-gated CAC	CACScoreDoc AI-based calcium scoring system (commercial) Output: Agatston score; CAC volume; CAC mass; Severity categories	Diagnostic accuracy; Agreement or cross-classification
Cho et al, 2025 (ref 5)	14	Republic of Korea	Observational cohort	289	Median 66 (IQR 59–73) years	42% (120/289)	GE	PET/CT attenuation-correction CT; Low-dose chest CT; 120 kVp; 3.75 mm; Standard	Paired ECG-gated CAC	AVIEW-CAC version 1.1.42.6 (commercial) Output: Agatston score; CAC volume; Severity categories	Agreement or cross-classification; Prognosis
Dobrolinska et al, 2023 (ref 6)	15	The Netherlands	Observational study; lung screening	213	Mean 64±9 years	47.9% (102/213)	Siemens	Low-dose chest CT; Non-contrast; 130 kVp; 3.0 mm; B31s	Paired ECG-gated CAC	syngo.via CT CaScoring VB50 automatic calcium scoring (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Dobrolinska et al, 2024 (ref 7)	16	Netherlands	Observational cohort; lung screening	572	Mean 58±9 years	46% (263/572)	Philips	PET/CT attenuation-correction CT; Low-dose chest CT; Non-contrast; 120 kVp; 5.0 mm; filtered back projection, kernel B; 5 mm increment	Paired ECG-gated CAC	Automatic CAC scoring from LDCT using deep-learning CAC-map/pseudomass method; Visual LDCT CAC scoring comparator (non-AI) (academic) Output: CAC pseudomass; Severity categories; CAC presence	Diagnostic accuracy; Prognosis (MACE)
Eng et al, 2021 (ref 8)	17	Multicentre; United States and Latin America/Brazil represented by participating sites/affiliations	Observational validation study; multicentre; routine chest CT	391	NR	41.7% (163/391)	Siemens, GE	Non-gated chest CT; Non-contrast; 5 mm	Paired ECG-gated CAC	Non-gated chest CT coronary artery calcium scoring deep learning model with XGBoost post-processing (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Gokmen et al, 2026 (ref 9)	18	United States	Observational validation study	205	NR	NR	NR	Non-gated chest CT	Paired ECG-gated CAC	CARD-ViT; AI-CAC (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Hagopian et al, 2025 (ref 10)	19	United States	Retrospective cohort; lung screening	10,085	Mean 65.9±9.2 years	5.8% (46/795)	Siemens, GE, Philips, Toshiba, Canon, Hitachi	Non-contrast LDCT; 2.5–5 mm; variable tube voltage (120 and non-120 kVp)	Paired ECG-gated CAC	AI-CAC (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Homayounieh et al, 2021 (ref 11)	20	USA	Retrospective cohort; lung screening	106	Mean 64±7 years	35.8% (38/106)	Siemens, GE, Philips	Low-dose chest CT; Non-contrast; 120 kVp; 1–1.25 mm	Paired ECG-gated CAC	FRONTIER radiomics prototype (research prototype) Output: CAC classification	Diagnostic accuracy
Jung et al, 2025 (ref 12)	21	South Korea	Observational validation study; lung screening	183	Mean 61±7 years	13.1% (24/183)	Siemens, GE	LDCT; 2.5–5 mm; mean tube voltage 127 kVp	Paired ECG-gated CAC	Conditional diffusion LDCT-to-CSCT translation model with random pointing and intermediate sampling (academic) Output: CAC volume	Agreement or cross-classification; Implementation
Kang et al, 2023 (ref 13)	22	South Korea	Observational validation study; lung screening	567	Mean 58.3±11.1 years	39.5% (224/567)	GE, Philips	LDCT; 120 kVp; 1.0 and 2.5 mm; Philips sharp C or GE Standard kernels; iterative reconstruction	Paired ECG-gated CAC	AVIEW CAC applied to LDCT2.5-mm; AVIEW CAC applied to LDCT1.0-mm (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Kim et al, 2025a (ref 14)	23	Republic of Korea	Retrospective cohort; multicentre; lung screening	225	Mean 57.3±9.2 years	27.6% (62/225)	Siemens, GE, Toshiba	Non-contrast LDCT; 120 kVp; native 1.0/1.25 mm, sharp kernel; additional kernel-conversion and/or 3-mm reconstructions	Paired ECG-gated CAC	AVIEW CAC version 1.1.43; original and kernel/thickness-converted images (commercial) Output: Agatston score; CAC volume; CAC density; Severity categories	Diagnostic accuracy
Kim et al, 2025b (ref 15)	24	South Korea and United States	Observational validation study; lung screening	1,077	Mean 60.8 years	20.6% (134/652)	Siemens, GE, Philips, Toshiba	LDCT: SNUH 1.0–1.25 mm, 120 kVp or Sn100; Stanford 0.7–5.0 mm, 100/120/140 kVp; standard and sharp kernels	Paired ECG-gated CAC	DL-CACS using ClariCardio with ClariCT.AI denoising for LDCT (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy
Li et al, 2023 (ref 16)	25	USA dataset (COCA from Stanford Hospital); author affiliations in China	Observational validation study; routine chest CT	659	NR	NR	NR	Non-gated chest CT	Paired ECG-gated CAC	Semantic-Prompt Scoring Siamese (SPSS) network (academic) Output: Agatston score; Severity categories	Four-category CAC cross-classification analysis
Lin et al, 2024 (ref 17)	26	Taiwan	Observational validation study; routine chest CT	76	Mean 66.32±10.81 years	30.3% (23/76)	Toshiba	Non-gated chest CT; Non-contrast; 120 kVp; 3 mm; FC02 soft kernel / FC56 sharp kernel	Paired ECG-gated CAC	AVIEW CAC applied to standard ECG-gated CCT (CSAICCT); AVIEW CAC applied to non-gated ChCT soft-kernel FC02 (CSChCTsoft); AVIEW CAC applied to non-gated ChCT sharp-kernel FC56 (CSChCTsharp) (commercial) Output: Agatston score; Severity categories	Agreement or cross-classification
Ludwig et al, 2026 (ref 18)	27	United States	Retrospective cohort; lung screening	620	Mean 64.3±10 years	NR	Siemens, GE, Philips (combined datasets)	Non-gated subset; acquisition parameters not reported separately	Paired ECG-gated CAC	DL-CAC (commercial) Output: Agatston score	Diagnostic accuracy
Marjanski et al, 2024 (ref 19)	28	Poland	Observational cohort; oncology/PET-CT	193	Mean 67.4±6.9 years	NR	Siemens	PET/CT attenuation-correction CT; Low-dose chest CT; Non-contrast; 120 kVp; 2 mm	Paired ECG-gated CAC	Syngo.via Siemens Healthcare VB60S_HF1 Ca-Score function (commercial) Output: Agatston score	Prognosis
McKinney et al, 2025 (ref 20)	29	United States	Retrospective cohort; routine chest CT	491	Mean 65.6±10.1 years	55.4% (272/491)	Siemens, GE, Philips, Toshiba	Non-contrast chest CT; >0.5 to ≤3.0 mm; soft or standard kernels	Paired ECG-gated CAC	Riverain Technologies ClearRead | CT CAC (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Morf et al, 2022 (ref 21)	30	Switzerland	Observational validation study; oncology/PET-CT	100	Mean 66±11.4 years	34% (34/100)	GE	PET/CT attenuation-correction CT; Non-contrast	Paired ECG-gated CAC	AVIEW CAC (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy
Oh et al, 2024 (ref 22)	31	Republic of Korea	Observational validation study; oncology/PET-CT	210	Mean 55.4±10.5 years	46.7% (98/210)	Siemens	PET/CT attenuation-correction CT; Non-contrast; 120 kVp; 3.0 mm; B30f	Paired ECG-gated CAC	AVIEW CAC (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy
Pieszko et al, 2022 (ref 23)	32	USA and Canada	Observational validation study	507	NR	NR	NR	PET/CT attenuation-correction CT; Low-dose chest CT	Paired ECG-gated CAC	convolutional long-short term memory deep neural network (conv-LSTM); U-Net reference model (academic) Output: CAC score; Severity categories	Diagnostic accuracy
Pieszko et al, 2023 (ref 24)	33	United States; Canada	Observational validation study; oncology/PET-CT	4,331	Median 71 (IQR 64–79) years	42% (1,828/4331)	Siemens	PET/CT attenuation-correction CT; Low-dose chest CT; 100 kVp; 3 mm	Paired ECG-gated CAC	ConvLSTM CAC segmentation and scoring model (academic) Output: Agatston-like score; Severity categories	Diagnostic accuracy; Prognosis (MACE)
Sartoretti et al, 2022 (ref 25)	34	Switzerland	Retrospective cohort; oncology/PET-CT	100	Mean 70.5±8 years	30% (30/100)	GE	PET/CT attenuation-correction CT; Non-contrast; 120 or 140 kVp; 1.25 mm; standard soft tissue kernel	Paired ECG-gated CAC	AVIEW CAC (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy; Agreement or cross-classification; Implementation
van Assen et al, 2021 (ref 26)	35	USA	Retrospective cohort; routine chest CT	263	NR	NR	Siemens	Non-gated chest CT; Non-contrast; 100–120 kVp; 1–3 mm	Paired ECG-gated CAC	AI-Rad Companion Chest CT (commercial) Output: CAC volume; Severity categories	Diagnostic accuracy
van Velzen et al, 2023 (ref 27)	36	Netherlands	Retrospective cohort; oncology/PET-CT	692	Mean 58±10 years	46.8%	Philips	PET/CT attenuation-correction CT; Non-contrast; 120 kVp; 5.0 mm	Paired ECG-gated CAC	Automatic CAC quantification and Visauto risk categorisation (academic) Output: CAC pseudomass; Severity categories; CAC presence	Diagnostic accuracy; Agreement or cross-classification
Xu et al, 2021 (ref 28)	37	China	Observational validation study; routine chest CT	901	Mean 62.9±22.9 years	42.5% (383/901)	Siemens, GE, Toshiba	Non-contrast chest CT; 1.0/1.25 mm; Bf40, FC07, or Standard kernels; FBP or iterative reconstruction	Paired ECG-gated CAC	CACScoreDoc (commercial) Output: Agatston score; CAC volume; CAC mass	Diagnostic accuracy
Xu et al, 2022 (ref 29)	38	China	Observational validation study; routine chest CT	523	Mean 61.8±12.1 years	46.6% (189/406)	Siemens, GE, Philips, Toshiba	Non-gated chest CT; Non-contrast; 120 kVp; 1.0 mm / 3.0 mm	Paired ECG-gated CAC	Two-stage segmentation deep learning algorithm, group A 1-mm model; Two-stage segmentation deep learning algorithm, group B 3-mm model (academic) Output: Agatston score; Severity categories; CAC presence	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Xu et al, 2025 (ref 30)	39	China	Retrospective cohort; multicentre; routine chest CT	2,241	Mean 61.18±11.29 years; Median 61 (IQR [54, 68]) years	43.91% (984/2241)	Siemens, GE, Philips, Canon, United Imaging	Non-gated chest CT; 120 kVp; 0.625–1.0 mm; standard soft-tissue kernel	Paired ECG-gated CAC	uAI Discover-Nongated Coronary Artery Calcium Score (commercial) Output: Agatston-like score; Severity categories	Diagnostic accuracy
Yin et al, 2024 (ref 31)	40	China	Observational study; lung screening	112	Mean 58.49±9.52 years	33.9% (38/112)	Siemens SOMATOM Force	Non-contrast LDCT; modulated 70–120 kVp; 1, 3, and 5 mm	Paired ECG-gated CAC	CACScoreDoc version 6.11 (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy
Yu et al, 2022 (ref 32)	41	China	Observational validation study; lung screening	405	Mean 59.6±11.8 years	42% (170/405)	Siemens, Toshiba	Non-gated chest CT; Non-contrast; 120 kVp; 1.5–2.0 mm; B30f/B35f or FC18 standard soft convolution kernels	Paired ECG-gated CAC	CACScoreDoc (commercial) Output: Agatston score; CAC volume; CAC mass; Severity categories	Diagnostic accuracy
Zhang et al, 2025 (ref 33)	42	China	Observational validation study; routine chest CT	290	Median 65 (IQR 58–72) years	38.3% (111/290)	Siemens, Philips	Non-gated chest CT; 120 kVp; 1 mm; FBP, ADMIRE 4, or iDose4 reconstruction	Paired ECG-gated CAC	CACScoreDoc v1.0 (commercial) Output: Agatston score; CAC volume; CAC mass	Agreement or cross-classification
Liu Y et al, 2026 (ref 106)	112	China	Retrospective paired chest/gated CT	1,334	Mean 63.75 (SD 9.94)	43.5% (580/1334)	Scope 16-row spiral CT; 60-row spiral CT; SOMATOM Sensation 64-slice spiral CT; SOMATOM Force dual-source CT	Non-contrast chest CT; 120–130 kVp; 1–1.5 mm; B80s, B35f, and BI35 kernels (Philips kernel NR)	Paired ECG-gated CAC	United Imaging intelligent post-processing software, with two-reader blinded review and consensus; version NR; Agatston score; calcification volume (mm^3); equivalent mass (mg); risk category	Diagnostic accuracy; Agreement
Liu YK et al, 2026 (ref 110)	116	China	Retrospective multicentre cross-sectional study; separate CAC validation	4,589; validation 853	CAC validation cohorts: NR; main association cohort: median 55 (IQR 45–63)	Validation NR; main cohort 41.9% (1924/4589)	NR	Main cohort: 100–120 kVp; 1–1.5 mm; standard body reconstruction. CAC validation: 0.625–1.25 mm; other parameters NR	Paired ECG-gated CAC	Custom two-stage 3D CNN for CAC; version NR; total and branch Agatston scores; CAC severity categories	Agreement
Same-scan manual reference											
Anifowose et al, 2026 (ref 34)	43	United Kingdom; United States public datasets	Observational validation study; oncology/PET-CT	323	NR	NR	Siemens	PET/CT attenuation-correction CT; Non-contrast; 100–140 kVp	Same-scan manual CAC	OncCalc (academic) Output: Agatston score	Diagnostic accuracy
Balbi et al, 2023 (ref 35)	44	Italy	Prospective cohort; lung screening	4,098	Median 60 (IQR 55–64) years	39.2% (1,607/4098)	Siemens	Low-dose chest CT; 120 kVp; 1.0 mm; medium-sharp kernel B50f	Same-scan manual CAC	AVIEW (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Barrett et al, 2025 (ref 36)	45	United States; Canada; Finland	Prospective cohort; multicentre; oncology/PET-CT	498	Median 63 (IQR 57–69) years	28.9% (144/498)	multicentre; not reported by site	PET/CT attenuation-correction CT; Low-dose chest CT; Non-contrast	Same-scan manual CAC	Previously validated deep-learning CAC model for PET/CT attenuation-correction maps (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Cano-Espinosa et al, 2018a (ref 37)	46	United States	Observational validation study; multicentre	5,973 scans	NR	NR	NR	Non-gated chest CT; Non-contrast	Same-scan manual CAC	Direct Agatston-score regression 3D CNN (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Cano-Espinosa et al, 2018b (ref 38)	47	NR	Observational validation study	6,663 scans	NR	NR	NR	Non-gated chest CT; Non-contrast	Same-scan manual CAC	3D CNN direct Agatston regression using mean logarithmic absolute error (loss function 4) (academic) Output: Agatston score; Severity categories	Eligible diagnostic evidence; linked participant cohort represented once in synthesis
Cano-Espinosa et al, 2020 (ref 39)	48	United States	Observational validation study	6,651	NR	NR	GE, Philips	Non-gated chest CT; Non-contrast	Same-scan manual CAC	RL-U-Net; RL-SE-Net (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Chamberlin et al, 2021 (ref 40)	49	USA	Retrospective cohort; lung screening	117	Mean 66.8±5.74 years	52.1% (61/117)	Siemens	Non-contrast LDCT; 110/120 kVp; 1-mm acquisition; CAC reference on 1.5-mm soft-tissue reformats	Same-scan manual CAC	AI-RAD Companion (VA10A) chest CT calcium detection model / AI-CACV (commercial) Output: CAC volume	Diagnostic accuracy
Ching et al, 2026 (ref 41)	50	USA	Observational study; routine chest CT	203 scans	NR	NR	NR	Non-gated chest CT	Same-scan manual CAC	YOLOv8m-Seg (open source research benchmark) Output: CAC score	Diagnostic accuracy; Agreement or cross-classification
Choi et al, 2022 (ref 42)	51	South Korea	Retrospective cohort; oncology/PET-CT	913	Mean 68.2±10.6 years; Median 68.3 years	30.7% (280/913)	NR	Non-contrast chest CT; 100/120/140 kVp; <2.5–5 mm (predominantly 3 mm); soft-tissue kernel	Same-scan manual CAC	AVIEW CAC (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy
de Vos et al, 2019 (ref 43)	52	Netherlands and United States	Observational validation study; routine chest CT	2,590 scans	NR	39%	NR	Non-contrast LDCT; 120/140 kVp; native 1–3 mm, resampled to 3 mm for scoring	Same-scan manual CAC	Direct automatic coronary calcium scoring with atlas-registration ConvNet and calcium-scoring ConvNet (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Gernaat et al, 2016 (ref 44)	53	Netherlands	Observational validation study; oncology/PET-CT	561	Median 61 (IQR 54–68) years	NR	Philips	Low-dose chest CT; Non-contrast; 120 kVp; 3 mm	Same-scan manual CAC	Išgum et al. automatic CAC scoring algorithm (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Gernaat et al, 2018 (ref 45)	54	Netherlands; Singapore	Observational cohort; oncology/PET-CT	2,288	NR	100% (2,288/2288)	Philips	Non-gated chest CT; Non-contrast; 120 kVp; 3 mm / 5 mm	Same-scan manual CAC	Deep learning automatic calcium scoring method based on Lessmann et al. (2018) (academic/research) Output: Agatston score; Severity categories	Agreement or cross-classification
González et al, 2016 (ref 46)	55	USA	Observational validation study; routine chest CT	1,500	NR	NR	NR	Non-gated chest CT	Same-scan manual CAC	Automated Agatston score computation method (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Hamelink et al, 2025b (ref 47)	56	Netherlands	Population-based ImaLife low-dose CT cohort; selected validation subset	684	Mean 58.9±4.8 years	51.2% (350/684)	Siemens	Non-contrast LDCT; 120 kVp; 3 mm; Qr36 kernel; FBP	Same-scan manual CAC	AI-Rad Companion chest CT prototype software (commercial) Output: Agatston score; CAC volume; Severity categories; CAC presence	Diagnostic accuracy
Huo et al, 2019 (ref 48)	57	USA	Observational validation study; routine chest CT	5,075 scans	NR	NR	NR	Non-gated chest CT; Non-contrast	Same-scan manual CAC	AID-Net; AID-Net (lambda=0; strategy-2 3-class intermediate output) (academic) Output: CAC presence	Diagnostic accuracy
Imuro et al, 2025 (ref 49)	58	United States	Observational validation study; routine chest CT	203 scans	NR	NR	NR	Non-gated chest CT	Same-scan manual CAC	Coned ResNeXt + PyRadiomics Random Forest CAC risk classifier (academic) Output: Severity categories	Diagnostic accuracy; Agreement or cross-classification
Išgum et al, 2010 (ref 50)	59	Netherlands and Belgium	Observational validation study; lung screening	121	NR	NR	Philips	Non-gated chest CT; Non-contrast; 120–140 kVp; 3.1 mm; Philips B moderately soft kernel	Same-scan manual CAC	Atlas-based automatic coronary calcium scoring system (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Išgum et al, 2012 (ref 51)	60	Netherlands/ Belgium	Observational validation study; multicentre; lung screening	568	Mean 59.9±5.8 years	NR	Philips	Non-contrast chest CT; 120–140 kVp; 1-mm images subsampled to 3.1 mm; moderately soft Philips B kernel	Same-scan manual CAC	Automatic coronary calcium scoring with probabilistic coronary calcium map and classifier combination (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
Išgum et al, 2016 (ref 52)	61	United States of America	Observational study; oncology/PET-CT	134	NR	36.6% (49/134)	Siemens	CT attenuation correction; 3 mm; 100 kVp reported in abstract	Same-scan manual CAC	Previously developed automatic CAC scoring algorithm for CSCT, retrained for CTAC (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Lessmann et al, 2016 (ref 53)	62	Netherlands	Observational validation study; lung screening	1,028	NR	NR	Siemens, Philips	Non-contrast LDCT; native 0.75/1.0 mm; averaging of consecutive slices for scoring; 120/140 kVp; medium-soft kernels	Same-scan manual CAC	Deep CNN automatic coronary calcium scoring system (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
Lessmann et al, 2018 (ref 54)	63	United States (NLST imaging data); author institutions in the Netherlands	Observational validation study; lung screening	1,459	NR	NR	four major CT scanner vendors	Non-contrast LDCT; 120/140 kVp; native 1–3 mm, reconstructed to 3 mm; soft and medium/sharp kernels	Same-scan manual CAC	Two-stage CNN calcium detector trained on soft reconstructions only; Two-stage CNN calcium detector trained on soft and sharp reconstructions (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
Miller et al, 2023 (ref 55)	64	United States and Canada	Observational validation study	8,879	Median 69 (IQR 58–74) years	49.5% (1,124/2271)	GE	Non-contrast CT attenuation correction; 110–130 kVp across datasets; 2.5–5 mm	Same-scan manual CAC	Deep learning CAC model using cascaded convLSTM networks (academic) Output: Agatston-like score; Severity categories	Diagnostic accuracy; Prognosis (MACE)
Mineo et al, 2025 (ref 56)	65	Brazil; external validation cohorts in Iran and United States	Observational validation study; routine chest CT	2,424	Mean 58±14 years	49% (1,038/2132)	NR	Non-gated chest CT; 120 kVp; 1 mm	Same-scan manual CAC	Two-step 2D U-Net deep learning pipeline for automated CAC quantification (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
Nguyen et al, 2024 (ref 57)	66	NR	Observational validation study; multicentre; oncology/PET-CT	801 scans	NR	NR	NR	PET/CT attenuation-correction CT; Non-contrast; 3.0 mm	Same-scan manual CAC	3D full-resolution multiclass nnU-Net with TotalSegmentator-derived heart/aorta/lung masks and linear Agatston correction (academic) Output: Agatston score	Agreement or cross-classification; non-gated-only performance not separable
Rathore et al, 2025 (ref 58)	67	United States (NLST imaging data); United Kingdom/United States author institutions	Observational validation study; lung screening	1,442	Mean 60±10 years	39%	NR	Non-contrast LDCT; 120 kVp; reconstruction interval 1.0–2.5 mm; soft-tissue reconstruction	Same-scan manual CAC	KALCIUM / KardioNet (commercial) Output: Agatston score; Severity categories; CAC presence	Diagnostic accuracy
Shadmi et al, 2018 (ref 59)	68	Israel; NLST/NIH cohort also analysed (country not explicitly stated in article)	Observational validation study; routine chest CT	NR	NR	NR	NR	Non-gated chest CT; Non-contrast	Same-scan manual CAC	FC-DenseNet coronary calcium scoring system; U-Net coronary calcium scoring system (commercial) Output: Agatston score; Severity categories	Diagnostic accuracy

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Shen et al, 2023 (ref 60)	69	China	Retrospective cohort; multicentre; routine chest CT	1,468	NR	46.5% (683/1468)	NR	Non-gated chest CT; Non-contrast	Same-scan manual CAC	AI-CACS software (CACScoreDoc version 1.8.2) (commercial) Output: CAC score; Severity categories	Agreement or cross-classification; Prognosis
Suh et al, 2023 (ref 61)	70	South Korea	Observational validation study; routine chest CT	452	Mean 55.6±9.8 years	15% (68/452)	Siemens, GE, Philips	Non-contrast LDCT; 120/140 kVp; 1.25–5 mm; predominantly sharp kernels (91.2%)	Same-scan manual CAC	AVIEW CAC (commercial) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
Takx et al, 2014 (ref 62)	71	Netherlands (UMC Utrecht single-site scan set; parent NELSON trial was Dutch-Belgian)	Observational validation study; lung screening	1,793	Median 60.1 (IQR 56.7–64.3) years	5.4% (97/1793)	Philips	Non-contrast chest CT; 120/140 kVp according to body weight; 3.1 mm	Same-scan manual CAC	Automatic coronary calcium scoring in low-dose chest CT using coronary calcium atlas and machine learning (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
Tang et al, 2017 (ref 63)	72	NR	Observational validation study	182	NR	NR	multiple vendors	Non-gated chest CT; Non-contrast; 1–3 mm	Same-scan manual CAC	Proposed fully automatic calcium detection and Agatston scoring framework (academic/research) Output: Agatston score; Severity categories	Diagnostic accuracy; Agreement or cross-classification; routine non-gated status not explicit
van Velzen et al, 2020 (ref 64)	73	The Netherlands; United States	Observational validation study; lung screening	7,240	NR	NR	Siemens, Philips	Non-contrast CT from multiple settings; 100–140 kVp; 0.9–3.0 mm across datasets	Same-scan manual CAC	baseline DL calcium scoring network; data-specific DL calcium scoring network; combined DL calcium scoring network (academic) Output: Agatston score; CAC volume; Severity categories	Diagnostic accuracy
van Velzen et al, 2022 (ref 65)	74	Netherlands; United States (NLST training scans)	Observational validation study; oncology/PET-CT	2,902	NR	NR	Philips	Non-contrast LDCT; 120/140 kVp; 1–3 mm across datasets	Same-scan manual CAC	Proposed Calcium CycleGAN CAC-map method; Automatic clinical calcium scoring comparator (ClinicalA) (academic) Output: Agatston score; CAC pseudomass; Severity categories	Diagnostic accuracy; Agreement or cross-classification; Implementation
Wang et al, 2026 (ref 66)	75	China; United States; Switzerland and other European centres	Retrospective cohort; lung screening	2,914	NR	NR	Siemens, GE, Philips, Toshiba	Non-gated chest CT; 1.0, 1.5, 2.0, or 5.0 mm; medium-smooth and medium-sharp kernels	Same-scan manual CAC	Unnamed validated deep-learning opportunistic CAC scoring algorithm Output: Agatston score; Severity categories	Diagnostic accuracy; Agreement or cross-classification

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Zeleznik et al, 2021 (ref 67)	7	United States / North America	Retrospective cohort; lung screening	20,084	Mean 61.5±5.1 years	40.9% (6,110/14959)	variety of CT scanners	Non-contrast chest CT; soft kernel preferred when multiple reconstructions were available	Same-scan manual CAC	Deep learning system for automated coronary artery calcium scoring (academic) Output: Agatston score; Severity categories	Diagnostic accuracy
Shanbhag et al, 2026 (ref 107)	113	USA; Canada	Multicentre CTAC development and validation	8,335	External CTAC: median 64 (IQR 56–73)	49.8% (2,974/5969); external cohort	NR	Non-contrast CT attenuation correction; 100–120 kVp; 3–5 mm; kernels NR	Same-scan expert CAC; clinical follow-up	KAD-CAC; version NR; per-patient Agatston; per-vessel Agatston; CAC categories	Agreement; Prognosis
Ochs et al, 2026 (ref 108)	114	Switzerland	Retrospective lung cancer screening	323	Median 61	44.3% (143/323)	Siemens NAEOTOM Alpha (171), SOMATOM Edge Plus (130), SOMATOM Force (22)	Non-contrast LDCT; 100–120 kVp; 2 mm; medium-smooth Br36/Br38 kernels	Same-scan semi-automated Agatston scoring	ClearRead CT CAC; version release 1.2.1.90; Agatston score; plaque-burden category; DICOM report/screenshots	Diagnostic accuracy; Agreement; Workflow
Hussain et al, 2026 (ref 113)	119	Pakistan	Prospective diagnostic and workflow study	210	Mean 61.8 (SD 10.4)	39.0% (82/210)	NR	Non-contrast routine chest CT; acquisition and reconstruction parameters NR	Same-scan manual Agatston scoring	Unspecified convolutional neural network CAC scoring algorithm; version NR; whole-heart total Agatston score; risk category; processing time	Diagnostic accuracy; Agreement; Workflow
Expert visual reference											
Alnasser et al, 2026 (ref 68)	76	England, United Kingdom	Retrospective cohort	642	Mean 64±13 years	63% (298/473)	Siemens, GE, Canon	Non-gated chest CT; Mixed contrast status; 120 kVp	Expert visual CAC	Fully automated two-stage 3D DL coronary artery segmentation and calcification detection model (academic) Output: CAC volume; CAC presence	Diagnostic accuracy
Bonatti et al, 2026 (ref 69)	77	Italy	Observational study; routine chest CT	90	Mean 65±17 years	46.7% (42/90)	Siemens	Non-contrast chest CT; 110/120 kVp; AI input 3-mm BL57 lung-kernel images; ADMIRE 3	Expert visual CAC	xAid Chest / xAID (commercial) Output: Agatston score; CAC presence	Diagnostic accuracy
Chamberlin et al, 2023 (ref 70)	78	USA	Observational study; oncology/PET-CT	189	Median 66 (IQR 18) years	54.5% (103/189)	Siemens	PET/CT attenuation-correction CT; Low-dose chest CT; Non-contrast; 120 kVp; 2 mm; BR38	Expert visual CAC	AI-Rad Companion (commercial) Output: CAC volume; Severity categories; CAC presence	Diagnostic accuracy
Kerndt et al, 2023 (ref 71)	79	United States	Observational validation study; routine chest CT	527	Mean 69.9±9.29 years	48.45% (125/258)	Siemens, GE, Philips	Non-gated chest CT; Non-contrast; 0.1–3.1 mm	Expert visual CAC	NANOX AI / Nanox CAC scoring algorithm (commercial; FDA-approved/FDA-cleared as reported) Output: Agatston-equivalent score; Severity categories	Agreement or cross-classification

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Park et al, 2024 (ref 72)	80	Republic of Korea (South Korea)	Observational cohort; lung screening	1,002	Mean 62.4±5.4 years	0.8% (8/1002)	Siemens, GE	LDCT; 120 kVp; 2.5/3.0 mm; GE Standard or Siemens Bf37/Br40 kernels	Expert visual CAC	aview CAC; syngo Calcium Scoring (commercial; U.S. FDA approved according to article; commercial; Not U.S. FDA approved according to article statement that only aview CAC was FDA approved) Output: Agatston score; Severity categories; CAC presence	Agreement or cross-classification; Prognosis
Shear et al, 2025 (ref 73)	81	United Kingdom	Retrospective cohort; routine chest CT	291	Mean 64±19 years	51% (147/291)	Siemens, GE, Philips	Non-contrast chest CT; acquisition and reconstruction settings varied across scans	Expert visual CAC	4DMedical CAC software / Imbio CAC (commercial) Output: Agatston score; CAC volume; CAC mass; Severity categories; CAC presence	Diagnostic accuracy
Smith et al, 2023 (ref 74)	82	United States of America	Retrospective cohort; lung screening	120	Mean 62±15.7 years	60% (72/120)	Siemens	Low-dose chest CT; Non-contrast; 110 kVp; 1.25 mm; B50s	Expert visual CAC	AI-RAD Companion (VA10A) (commercial) Output: CAC volume; CAC presence	Descriptive diagnostic accuracy
Wu et al, 2025 (ref 75)	83	Taiwan	Observational validation study; oncology/PET-CT	677	NR	NR	GE, Philips	PET/CT attenuation-correction CT; Low-dose chest CT; Non-contrast; 120 kVp; 3.75 mm	Expert visual CAC	DLA-3D / DLA34-3D (academic) Output: CAC presence	Diagnostic accuracy; Implementation
Xie et al, 2017 (ref 76)	84	USA	Observational study; lung screening	1,359	NR	NR	Siemens	Low-dose chest CT; 110 or 130 kVp; 2.0 mm; B70s	Expert visual CAC	Fully automated CAC identification and artery-labelling algorithm (Academic) Output: Agatston score; CAC volume; CAC mass	Diagnostic accuracy; Agreement or cross-classification
Prognostic evidence											
Atkins et al, 2022 (ref 77)	85	United States	Retrospective cohort; lung screening	428	NR	48.6% (208/428)	GE	Non-gated chest CT; Non-contrast; ≤3.0 mm	Clinical outcome ascertainment	DL-CAC algorithm (academic) Output: Agatston-like score; Severity categories	Prognosis
Barda et al, 2022 (ref 78)	86	Israel	Retrospective cohort; routine chest CT	14,135	Mean 61.1±9.9 years	41.3% (5,836/14135)	NR	Non-gated chest CT; Non-contrast	Clinical outcome ascertainment	Zebra Medical Vision coronary calcium quantification software / variant of Shadmi et al. automated CAC scoring method (commercial; CE and FDA-approved product in the USA and Europe) Output: Agatston score	Prognosis

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
de Vos et al, 2015 (ref 79)	87	Netherlands	Observational study; lung screening	3,457	NR	0% (0/3457)	Siemens, Philips	Non-contrast LDCT; 120/140 kVp by body weight; 3.1 mm; moderately smooth B30f or Philips B kernel	Clinical outcome ascertainment	Previously developed automatic aortic and coronary calcification detection algorithms plus linear SVM CVE classifier (academic) Output: CAC volume	Prognosis
de Vos et al, 2021 (ref 80)	88	United States	Nested case-control study; multicentre; lung screening	5,564	Median 61 years	37.9% (2,109/5564)	multiple vendors	Low-dose chest CT; Non-contrast; 120–140 kVp; 3.0 mm	Clinical outcome ascertainment	CalcScore / automatic deep learning calcium scoring (academic) Output: Agatston score	Prognosis; Implementation
Dekker et al, 2021 (ref 81)	89	Netherlands	Observational cohort; oncology/PET-CT	747	Mean 67±10 years	49.5% (370/747)	Siemens	PET/CT attenuation-correction CT; Low-dose chest CT; 120 kVp	Clinical outcome ascertainment	validated deep learning method for automated CAC scoring on LDACT (unclear; developed/validated academically, but article notes the method is available with a purchased licence) Output: Agatston score; Severity categories	Prognosis
Kim EJ et al, 2025 (ref 82)	90	Republic of Korea	Observational cohort; lung screening	193	Mean 61.6±5.2 years	0% (0/193)	Siemens, GE	LDCT; 120 kVp; 2.5/3.0 mm; Standard, Bf37, or Br40 kernels	Clinical outcome ascertainment	AVIEW CAC (commercial) Output: Agatston score; Severity categories	Prognosis
Kim et al, 2022 (ref 83)	91	Republic of Korea	Observational cohort; oncology/PET-CT	1,111	Median 50 years	NR	NR	PET/CT attenuation-correction CT; Non-contrast	Clinical outcome ascertainment	AVIEW-CAC v1.1.38.6 (commercial software) Output: Agatston score; CAC volume; CAC mass	Prognosis
Kim et al, 2026 (ref 84)	92	Republic of Korea	Observational cohort; lung screening	1,109	Mean 60.83±8.56 years; Median 60.3 (IQR 55.0–66.9) years	13% (144/1109)	Siemens, GE	Non-contrast chest CT; 100/120 kVp; 1–5 mm; standard and sharp kernels	Clinical outcome ascertainment	ClariCardio version 1.0 (commercial) Output: Agatston score; CAC volume; CAC density	Prognosis
Kim S et al, 2025 (ref 85)	93	South Korea	Retrospective cohort; oncology/PET-CT	3,153	NR	36.5% (1,151/3153)	NR	Non-gated chest CT; Non-contrast	Clinical outcome ascertainment	AVIEW CAC (commercial) Output: Agatston score; Severity categories	Prognosis
Lee et al, 2025 (ref 86)	94	Korea	Retrospective cohort; lung screening	7,404	Median 53.9 (IQR 48.1–60.2) years	20.7% (1,529/7404)	NR	Non-contrast chest CT; 100/120 kVp; 1–1.25 mm or 3–5 mm by centre	Same-scan manual CAC	ClariCardio (commercial) Output: Agatston score; Severity categories	Prognosis (MACE)
Lessmann et al, 2019 (ref 87)	95	United States	Nested case-control study; lung screening	5,718	Mean 61.1±7.8 years	37.9% (2,165/5718)	NR	Low-dose chest CT; Non-contrast; 120 kVp; 3.0 mm	Clinical outcome ascertainment	Automatic in-house calcium scoring software (academic/in-house) Output: Modified Agatston score; CAC volume	Prognosis

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Mets et al, 2013 (ref 88)	96	Netherlands	Observational validation study; lung screening	3,648	NR	2.4% (89/3648)	Siemens, Philips	LDCT; 120/140 kVp by body weight; 3.1-mm scoring images; smooth B30f or Philips B kernel	Clinical outcome ascertainment	Dedicated noncommercial in-house coronary calcium quantification algorithm (academic/noncommercial) Output: Agatston score; CAC volume	Prognosis
Sato et al, 2024 (ref 89)	97	Japan	Observational cohort; routine chest CT	548	Median 72.0 (IQR 61.0–79.0) years	22.6% (124/548)	Canon	Non-gated chest CT; Non-contrast; 120 kVp; 0.5 mm; Precise IQ Engine reconstruction method	Clinical outcome ascertainment	SYNAPSE VINCENT Workspace (commercial) Output: Agatston score	Prognosis
Stemmer et al, 2020 (ref 90)	98	United States	Retrospective cohort; lung screening	12,332	Mean 61.8±5.1 years	40.7% (5,024/12332)	NR	Low-dose chest CT; Non-contrast	Clinical outcome ascertainment	CCS-Alg (commercial) Output: CAC score; Severity categories	Prognosis
Takx RA et al, 2015 (ref 91)	99	Netherlands and Belgium; analysed CT scans from Groningen and Utrecht screening sites	Observational cohort; lung screening	3,111	Median 59.3 (IQR 56.0–63.3) years	0% (0/3111)	Siemens, Philips	Non-contrast LDCT; 120/140 kVp by body weight; 3.1 mm	Clinical outcome ascertainment	Automatic coronary calcium scoring algorithm for low-dose chest CT (previously described by Išgum et al.) (academic) Output: Modified Agatston score; CAC volume	Prognosis
Takx RAP et al, 2015 (ref 92)	100	Netherlands and Belgium	Observational cohort; lung screening	3,080	Median 59.2 (IQR 55.9–63.3) years	0% (0/3080)	Siemens, Philips	Non-contrast chest CT; 120/140 kVp by body weight	Clinical outcome ascertainment	Dedicated in-house automatic coronary calcium scoring software (academic/in-house) Output: CAC volume	Prognosis
Weber et al, 2026 (ref 93)	101	USA	Retrospective cohort; routine chest CT	2,546	Median 59.2 (IQR 52.6–64.5) years	66.5% (1,694/2546)	NR	Non-gated chest CT; Non-contrast	Clinical outcome ascertainment	Nanox.AI HealthCCSng (commercial; Food and Drug Administration-approved) Output: Agatston score; Severity categories	Prognosis
Williams et al, 2024 (ref 94)	102	United States and Canada (data sources); United Kingdom affiliation for first author	Observational validation study; multicentre; clinical cardiac cohort	11,063	Mean 64±12 years	50% (2,974/5969)	NR	PET/CT attenuation-correction CT; Non-contrast; 100 or 120 kVp; 3, 5, or 2.5 mm	Expert-derived per-vessel Agatston CAC score on the same non-gated attenuation-correction CT	convLSTM vessel-specific deep learning CAC model (academic) Output: Agatston score; Severity categories	Prognosis (MACE)
Xing et al, 2026 (ref 95)	103	China	Retrospective cohort; routine chest CT	324	NR	38% (123/324)	Philips, Neusoft	Non-gated chest CT; Non-contrast; 120 kVp; 1 mm	Clinical outcome ascertainment	Shukun Technology artificial intelligence analysis system (commercial company developed) Output: Agatston score; CAC volume; CAC mass; Severity categories	Prognosis

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Yang et al, 2026 (ref 96)	104	China	Observational cohort; routine chest CT	509	Median 64.0 (IQR 57.00–70.50) years	37.7% (192/509)	Siemens, GE, Philips, Canon, United Imaging	Non-gated chest CT; 120 kVp; 0.625, 0.75, or 1.00 mm; standard soft tissue kernel	Clinical outcome ascertainment	DL-CACS (academic) Output: Agatston score; Severity categories	Prognosis (MACE)
Zhou et al, 2026 (ref 97)	105	International multicentre study; sites in United States, Canada, Mexico, Italy, and Switzerland	Retrospective cohort; multicentre; oncology/PET-CT	4,552	Mean 65±12 years	48.6% (2,214/4552)	Siemens, GE, Philips	Non-contrast CT attenuation correction; 2–5 mm; typically 100–140 kVp	Clinical outcome ascertainment	Automated vessel-specific CAC segmentation plus proximal CAC identification pipeline (academic) Output: Agatston score; CAC density	Prognosis (MACE)
Noureldine et al, 2026 (ref 109)	115	USA	Retrospective multicentre immune-checkpoint-inhibitor cohort	1,873	Mean 66	43%	NR	Routine non-gated CT; acquisition and reconstruction parameters NR	Clinical follow-up	Previously validated AI opportunistic CAC model of Hagopian et al (2025 NEJM AI); version NR; CAC Agatston category/score	Prognosis
Yi et al, 2026 (ref 111)	117	USA; Canada; Italy; Mexico	Retrospective multicentre myocardial perfusion registry	28,509	Overall report: median 66 (IQR 58–74); external CAC-only subgroup: NR	44%; overall report	NR	Non-gated CT attenuation correction; site-specific 80–140 kVp and 2–5 mm protocols; B30s, I26f, and B19s where reported	Clinical follow-up	Previously validated automated CAC scoring model; version NR; CAC score (AU)	Prognosis
Cheng et al, 2026 (ref 112)	118	China	Retrospective CKD development and external validation	2,877	Derivation cohort: median 66 (IQR 58–72)	38.1% (918/2411); development cohort	NR	Routine non-gated chest CT; predominantly 120 kVp and 5 mm; kernels NR	Clinical follow-up	Commercial CAC-scoring module of DeepChest; version NR; CAC score; CAC categories	Prognosis
Implementation trials or pathways											
Bouladian et al, 2025 (NOTIFY-EXTEND; ref 98)	106	United States	Observational study; routine chest CT	163	NR	NR	NR	Non-gated chest CT; Non-contrast	Clinical implementation workflow	deep learning algorithm Output: CAC presence	Implementation
Chernina et al, 2023 (ref 99)	107	Russian Federation	Retrospective cohort; routine chest CT	160	Mean 49.5 years	53%	Philips	Non-gated chest CT; Non-contrast; 1.0 mm; Two series: soft-tissue kernel and pulmonary kernel; iterative model reconstruction	Expert radiologist confirmation and clinical implementation workflow	Agatston-IRA within the multi-IRA combined AI service (commercial software product; Approved for prospective use in 153 Moscow healthcare organisations; formal regulatory clearance was not stated) Output: Agatston score; CAC presence	Diagnostic accuracy; Implementation

Supplementary Table 5A continued

Report	Main-text ref.	Country	Setting and design	N	Age	Female	CT imaging		Reference standard	AI method, output, and status	Outcomes contributed
							Scanner vendor	Acquisition and reconstruction			
Dudum et al, 2026 (NOTIFY-PICTURE; ref 100)	12	United States	Randomised trial; routine chest CT	202	Mean 67.9±9.9 years	57.4% (116/202)	NR	Non-gated chest CT; Non-contrast	Randomised comparator	AI algorithm for CAC quantification/screening; referenced prior method was deep learning automated CAC scoring (FDA-cleared) Output: CAC score	Implementation outcome
Sandhu et al, 2023 (NOTIFY-1; ref 101)	11	United States	Randomised trial; routine chest CT	173	Median 70.8 (IQR 64.0–75.8) years	52.6% (91/173)	NR	Non-gated chest CT; Non-contrast	Expert visual CAC assessment	DL-CAC algorithm (academic) Output: Agatston score; Severity categories; CAC presence	Implementation outcome
Other eligible evidence											
Bunk et al, 2026 (ref 102)	108	Netherlands	Observational study; lung screening	1,436	Mean 59.7±5.7 years	13.7% (197/1436)	Siemens, Philips	Non-contrast LDCT; 120/140 kVp; 1-mm images; medium-soft Philips B or Siemens B30f kernels	AI-derived CACV on the baseline scan used as the comparator for AI-derived CACV on the repeat scan	AI-Rad Companion Chest CT, Cloud product version VA40 (commercial product) Output: CAC volume; Severity categories	Agreement or cross-classification
Hamelink et al, 2025a (ref 103)	109	Netherlands	Population-based ImaLife follow-up LDCT cohort; scan–rescan repeatability study	632	Mean 63.3±11 years	43.4% (274/632)	Siemens	Non-contrast LDCT; 120 kVp; 1-mm thickness, 0.7-mm increment; Br40 soft-tissue kernel	Repeat non-gated CT; no independent reference	AI-Rad Companion chest CT research application, prototype version 01/2023 (Commercial) Output: CAC volume; Severity categories	Scan–rescan repeatability
Ledda et al, 2025 (ref 104)	110	Italy	Observational study; lung screening	361	NR	NR	Siemens	LDCT/ultra-low-dose CT; 120 kVp or tin-filtered 100–150 kVp; 1 mm; Br49/Qr49 kernels with ADMIRE	Paired LDCT automated Agatston CAC categories from the same AI software	AVIEW (Commercial) Output: Agatston score; Severity categories	Diagnostic accuracy; Agreement or cross-classification
Yacoub et al, 2022 (ref 105)	111	United States	Consecutive clinically indicated noncontrast chest CT; retrospective diagnostic-performance comparison	100	Mean 64.2±14.8 years	43% (43/100)	Siemens	Routine noncontrast chest CT; Non-contrast; 120 kVp; 1.0 mm; standard soft tissue reconstruction algorithm	Same-scan blinded binary consensus assessment by two cardiothoracic radiologists	AI-Rad Companion (commercial; FDA cleared) Output: CAC volume; CAC presence	Descriptive diagnostic accuracy at study-defined calcium-volume ≥10 mm^3; not pooled

Each row represents one included report; Main-text ref. is the reference number of the report in the Article, and (ref n) is its number in the reference list of this Supplementary Information. N, age, and sex describe the report population, as reported; analysis-specific denominators are provided with the corresponding results. For Hagopian et al, 2025, the female proportion describes the paired-gated subset (46/795), whereas report N includes all reported cohorts. CT imaging columns describe the non-gated input unless otherwise specified; available acquisition and reconstruction parameters are shown. Complete technical characteristics, including validation and failed-scan fields, are provided in Supplementary Data 1, sheet S1. Age is reported as mean±SD or median (IQR), as available. ASCVD, atherosclerotic cardiovascular disease; CAC, coronary artery calcium; CT, computed tomography; FBP, filtered back projection; IQR, interquartile range; LDCT, low-dose CT; NR, not reported; SD, standard deviation.

Supplementary Table 5B. Summary of report linkage and potential cohort overlap

The summary identifies the report-linkage categories used to prevent duplicate participant contributions within a synthesis.

Overlap status	Reports, n	Linked reports or basis	Synthesis handling
Definite	9	Nine reports in four linked groups: NOTIFY-1/NOTIFY-EXTEND; Dobrolinska 2024/van Velzen 2023; Kim 2025b/Li 2023/Gokmen 2026; and Williams 2024/Shanbhag 2026.	Definite or probable overlap was represented once within each synthesis.
Probable	4	Four Cano-Espinosa/González reports formed one linked group; possible extensions are documented in Supplementary Data 1, sheet S12.	Definite or probable overlap was represented once within each synthesis.
Possible	15	Fifteen reports had possible linkage that could not be resolved from report-level information.	Reports remained eligible as distinct reporting units; possible linkage remained identified.
No evidence of overlap	43	Forty-three reports had no identified participant overlap.	Reports remained eligible as distinct reporting units; possible linkage remained identified.

Detailed report-level cohort records, pairwise overlap decisions, algorithm lineage, and estimate-selection decisions are provided in Supplementary Data 1, sheet S12.

Supplementary Section 3

Threshold diagnostic performance and certainty

Supplementary Table 6. Complete threshold-specific pooled results, small-study effects, and sensitivity analyses

Panel A. Threshold-specific pooled sensitivity and specificity

Threshold-specific pools are grouped by reference standard. Different thresholds and reference standards were not combined.

Reference standard	Threshold	k reports	k estimates	N	Sensitivity	Sensitivity 95% CI	Specificity	Specificity 95% CI	Model	Continuity correction
Paired ECG-gated reference	CAC >0	17	24	9,393	87.1%	83.2–90.2%	92.3%	87.2–95.4%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 1, specificity 4 estimates
Paired ECG-gated reference	CAC ≥100	19	26	10,940	83.1%	78.5–86.8%	96.3%	94.2–97.7%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 0, specificity 3 estimates
Paired ECG-gated reference	CAC ≥300	4	6	1,727	80.5%	61.5–91.5%	98.9%	96.2–99.7%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 1, specificity 2 estimates
Paired ECG-gated reference	CAC ≥400	17	22	9,543	80.1%	74.3–84.9%	97.7%	96.2–98.7%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 1, specificity 5 estimates
Same-scan manual reference	CAC >0	20	22	10,878	93.8%	91.0–95.7%	88.7%	83.4–92.4%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 2, specificity 1 estimates
Same-scan manual reference	CAC ≥100	24	25	14,371	93.3%	91.1–95.0%	97.5%	96.0–98.4%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 1, specificity 3 estimates
Same-scan manual reference	CAC ≥300	5	6	1,925	90.5%	82.0–95.2%	98.5%	97.1–99.3%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 2, specificity 3 estimates
Same-scan manual reference	CAC ≥400	20	20	12,742	90.3%	86.7–93.0%	98.7%	97.7–99.3%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 3, specificity 5 estimates
Expert visual reference	CAC >0	4	4	1,043	95.8%	93.9–97.1%	90.3%	68.0–97.6%	Bivariate random effects (Reitsma; REML)	0.5 for zero cells; sensitivity 0, specificity 0 estimates

Values are pooled estimates with 95% confidence intervals. N is the summed sample size across retained estimates; k reports is the number of contributing reports. The 0.5 continuity correction was applied only to individual estimates containing a zero cell and separately for sensitivity and specificity. AU=Agatston units; CAC, coronary artery calcium; REML, restricted maximum likelihood.

Panel B. Deeks tests in the paired ECG-gated-reference analyses

Deeks weighted regression was applied separately at each threshold after aggregation of dependent estimates within reports.

Threshold	k reports	Deeks P value	Interpretation
CAC >0	17	0.139	No evidence of funnel asymmetry
CAC ≥100	19	0.067	Funnel asymmetry suggested
CAC ≥300	4	0.062	Interpret cautiously; fewer than 10 reports
CAC ≥400	17	0.054	Funnel asymmetry suggested

Panel C. Sensitivity analyses

Cohort-family analyses were repeated at CAC >0, ≥100, ≥300, and ≥400 after linking reports by named dataset, institution, enrolment period, and sample size. Additional analyses at CAC >0 and CAC ≥100 examined overall Low QUADAS-3 risk of bias, within-report aggregation, confirmed non-contrast CT, and two extreme allocations of reconstructable failed or unassessable scans.

Threshold	Analysis	k reports	k model inputs	N	Reports removed	Sensitivity (95% CI)	Specificity (95% CI)
CAC >0	Cohort-family restriction	17	24	9,393	0	87.1% (83.2–90.2)	92.3% (87.2–95.4)
CAC >0	Overall low risk of bias only	6	11	2,228	11 reports; 13 estimates	87.5% (81.6–91.7)	84.8% (75.8–90.9)
CAC >0	Within-report aggregation	17	17	9,393	0 reports; 0 estimates	86.5% (82.5–89.7)	92.8% (86.3–96.3)
CAC >0	Confirmed non-contrast CT only	15	22	8,486	2 reports; 2 estimates	86.7% (82.4–90.1)	92.4% (87.7–95.4)
CAC >0	Reconstructable failed scans as false negatives	17	24	9,409	0 reports; 0 estimates	86.8% (82.7–90.1)	92.3% (87.2–95.4)
CAC >0	Reconstructable failed scans as false positives	17	24	9,409	0 reports; 0 estimates	87.1% (83.2–90.2)	91.2% (86.1–94.6)
CAC ≥100	Main analysis	19	26	10,940	—	83.1% (78.5–86.8)	96.3% (94.2–97.7)
CAC ≥100	Cohort-family restriction	19	26	10,940	0	83.1% (78.5–86.8)	96.3% (94.2–97.7)
CAC ≥100	Overall low risk of bias only	6	11	2,228	13 reports; 15 estimates	78.5% (71.4–84.1)	94.8% (90.6–97.2)
CAC ≥100	Within-report aggregation	19	19	10,940	0 reports; 0 estimates	83.1% (78.5–87.0)	96.8% (94.4–98.2)
CAC ≥100	Confirmed non-contrast CT only	17	24	10,033	2 reports; 2 estimates	82.2% (77.4–86.1)	96.5% (94.4–97.9)
CAC ≥100	Reconstructable failed scans as false negatives	19	26	10,956	0 reports; 0 estimates	82.3% (77.7–86.2)	96.3% (94.2–97.7)
CAC ≥100	Reconstructable failed scans as false positives	19	26	10,956	0 reports; 0 estimates	83.1% (78.6–86.9)	95.9% (93.6–97.3)
CAC ≥300	Cohort-family restriction	4	6	1,727	0	80.5% (61.5–91.5)	98.9% (96.2–99.7)
CAC ≥400	Cohort-family restriction	17	22	9,543	0	80.1% (74.3–84.9)	97.7% (96.2–98.7)

No definite or probable overlapping cohort was represented twice within any paired-gated threshold, so the cohort-family refits retained the selected inputs and estimates. All analyses used the same bivariate random-effects model and continuity-correction rule. At CAC >0 and CAC ≥100, the failed-scan scenarios assigned all 16 reconstructable failures as either false negatives or false positives; failures in three reports could not be mapped to the selected paired subset. In the CAC ≥100 low-risk-of-bias model, the estimated between-study sensitivity–false-positive-rate correlation reached the boundary (1.00). AU=Agatston units; CI, confidence interval; REML, restricted maximum likelihood.

Panel D. Likelihood ratios for paired ECG-gated-reference analyses

Likelihood ratios were derived from each threshold-specific bivariate model and its joint fixed-effect coefficient covariance.

Threshold	k reports	k estimates	N	LR+ (95% CI)	LR− (95% CI)
CAC >0	17	24	9,393	11.24 (6.81–18.91)	0.14 (0.11–0.18)
CAC ≥100	19	26	10,940	22.38 (14.44–34.81)	0.18 (0.14–0.22)
CAC ≥300	4	6	1,727	74.21 (23.57–217.89)	0.20 (0.09–0.39)
CAC ≥400	17	22	9,543	35.36 (21.13–59.18)	0.20 (0.16–0.26)

LR+, positive likelihood ratio; LR−, negative likelihood ratio.

Supplementary Table 6 continued

Panel E. Predictive values across assumed target prevalences

Values apply the pooled paired ECG-gated-reference sensitivity and specificity at CAC >0, CAC ≥100, CAC ≥300, and CAC ≥400 to four assumed target prevalences.

CAC threshold	Measure, % (95% CI)	Assumed prevalence of CAC meeting the specified threshold			
		10%	15%	25%	40%
CAC >0	PPV	55.5 (43.1–67.8)	66.5 (54.6–76.9)	78.9 (69.4–86.3)	88.2 (82.0–92.6)
	NPV	98.5 (98.0–98.8)	97.6 (96.9–98.1)	95.5 (94.3–96.6)	91.5 (89.2–93.3)
	Target-positive after a negative result	1.5 (1.2–2.0)	2.4 (1.9–3.1)	4.5 (3.4–5.7)	8.5 (6.7–10.8)
CAC ≥100	PPV	71.3 (61.6–79.5)	79.8 (71.8–86.0)	88.2 (82.8–92.1)	93.7 (90.6–95.9)
	NPV	98.1 (97.6–98.5)	97.0 (96.2–97.6)	94.5 (93.1–95.6)	89.5 (87.1–91.6)
	Target-positive after a negative result	1.9 (1.5–2.4)	3.0 (2.4–3.8)	5.5 (4.4–6.9)	10.5 (8.4–12.9)
CAC ≥300	PPV	89.2 (72.4–96.0)	92.9 (80.6–97.5)	96.1 (88.7–98.6)	98.0 (94.0–99.3)
	NPV	97.9 (95.9–99.0)	96.6 (93.6–98.5)	93.8 (88.6–97.2)	88.4 (79.5–94.5)
	Target-positive after a negative result	2.1 (1.0–4.1)	3.4 (1.5–6.4)	6.2 (2.8–11.4)	11.6 (5.5–20.5)
CAC ≥400	PPV	79.7 (70.1–86.8)	86.2 (78.9–91.3)	92.2 (87.6–95.2)	95.9 (93.4–97.5)
	NPV	97.8 (97.2–98.3)	96.5 (95.6–97.3)	93.7 (92.0–95.1)	88.1 (85.1–90.6)
	Target-positive after a negative result	2.2 (1.7–2.8)	3.5 (2.7–4.4)	6.3 (4.9–8.0)	11.9 (9.4–14.9)

Prevalences are illustrative assumptions applied separately to each threshold, not observed prevalences in one cohort. PPV and NPV are derived from the paired ECG-gated diagnostic estimates. Target-positive after a negative result = 100% − NPV.

CI, confidence interval. CAC thresholds are in Agatston units.

NPV, negative predictive value; PPV, positive predictive value.

Supplementary Table 7. GRADE-DTA certainty of evidence for automated CAC classification on non-gated CT against paired ECG-gated calcium-scoring CT

Certainty is rated separately for sensitivity and specificity. The table is noncontextualised: it rates accuracy evidence, not the downstream clinical effectiveness of testing.

Panel A. Evidence profile

Threshold	Outcome	Evidence base	Pooled accuracy (95% CI)	Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Certainty
CAC >0	Sensitivity	17 studies; 24 estimates; N = 9,393	87.1% (83.2–90.2)	Serious[a]	Not serious[b]	Serious[c]	Not serious	Not strongly suspected[j]	⨁⨁◯◯ Low
CAC >0	Specificity	17 studies; 24 estimates; N = 9,393	92.3% (87.2–95.4)	Serious[a]	Not serious[b]	Very serious[d]	Not serious	Not strongly suspected[j]	⨁◯◯◯ Very low
CAC ≥100	Sensitivity	19 studies; 26 estimates; N = 10,940	83.1% (78.5–86.8)	Serious[a]	Not serious[b]	Serious[e]	Not serious	Not strongly suspected[j]	⨁⨁◯◯ Low
CAC ≥100	Specificity	19 studies; 26 estimates; N = 10,940	96.3% (94.2–97.7)	Serious[a]	Not serious[b]	Serious[f]	Not serious	Not strongly suspected[j]	⨁⨁◯◯ Low
CAC ≥300	Sensitivity	4 studies; 6 estimates; N = 1,727	80.5% (61.5–91.5)	Serious[a]	Not serious[b]	Serious[g]	Serious[i]	Not strongly suspected[j]	⨁◯◯◯ Very low
CAC ≥300	Specificity	4 studies; 6 estimates; N = 1,727	98.9% (96.2–99.7)	Serious[a]	Not serious[b]	Serious[h]	Not serious	Not strongly suspected[j]	⨁⨁◯◯ Low

Panel B. Illustrative absolute classifications per 1,000 people

Threshold	Outcome	Illustrative probability	Correct results per 1,000 (95% CI)	Incorrect results per 1,000 (95% CI)	Certainty
CAC >0	Sensitivity	60% reference positive	TP 522 (499–541)	FN 78 (59–101)	⨁⨁◯◯ Low
CAC >0	Specificity	60% reference positive	TN 369 (349–382)	FP 31 (18–51)	⨁◯◯◯ Very low
CAC ≥100	Sensitivity	40% reference positive	TP 332 (314–347)	FN 68 (53–86)	⨁⨁◯◯ Low
CAC ≥100	Specificity	40% reference positive	TN 578 (565–586)	FP 22 (14–35)	⨁⨁◯◯ Low
CAC ≥300	Sensitivity	15% reference positive	TP 121 (92–137)	FN 29 (13–58)	⨁◯◯◯ Very low
CAC ≥300	Specificity	15% reference positive	TN 841 (818–847)	FP 9 (3–32)	⨁⨁◯◯ Low

a. Risk of bias was downgraded one level for every outcome. High overall risk was present in 12/24 CAC >0 estimates, 14/26 CAC ≥100 estimates, and 4/6 CAC ≥300 estimates; each threshold had one additional estimate with insufficient information. The dominant concern was participant selection or exclusion of difficult scans, high-burden cases, motion, or hardware, with occasional analysis-domain limitations. The evidence was not downgraded two levels because low-risk estimates remained and no single bias mechanism dominated the pooled results.

b. Indirectness was not downgraded. All three evidence bodies directly evaluated automated CAC classification on non-gated CT against paired ECG-gated calcium-scoring CT at the thresholds of interest. Selected clinical or protocol-defined cohorts created estimate-level applicability concerns, but the contributing studies collectively covered complementary routine-care, screening, and clinically referred populations; these concerns were therefore not mechanically counted again as body-level indirectness.

c. CAC >0 sensitivity was inconsistent (serious): report estimates ranged from 62.7% to 100.0%; the 95% prediction interval was 62.6%–96.4% (descriptive I^2 92.8%).

d. CAC >0 specificity was very seriously inconsistent: report estimates ranged from 56.7% to 100.0%; the 95% prediction interval was 50.2%–99.3% (descriptive I^2 93.8%).

e. CAC ≥100 sensitivity was inconsistent (serious): report estimates ranged from 50.0% to 94.5%; the 95% prediction interval was 56.4%–94.9% (descriptive I^2 90.5%).

f. CAC ≥100 specificity was inconsistent (serious): report estimates ranged from 83.6% to 100.0%; the 95% prediction interval was 76.1%–99.5% (descriptive I^2 91.6%).

g. CAC ≥300 sensitivity was inconsistent (serious): only four reports contributed, report estimates ranged from 41.7% to 100.0%, and the 95% prediction interval was 29.3%–97.6% (descriptive I^2 83.3%). The limited number of reports and wide interval were also reflected in the separate imprecision judgment and were not used to impose a second inconsistency downgrade.

h. CAC ≥300 specificity was inconsistent (serious): report estimates ranged from 91.1% to 100.0%, and the 95% prediction interval was 81.0%–99.9% (descriptive I^2 78.7%).

i. Imprecision was serious only for CAC ≥300 sensitivity: the pooled 95% confidence interval was 61.5%–91.5%, only four reports contributed, and the estimate was based on 328 reference-positive participants. The narrower crossings of exploratory benchmarks for CAC >0 specificity (87.2%–95.4%) and CAC ≥100 sensitivity (78.5%–86.8%) did not by themselves warrant downgrading because those benchmarks were not protocol-prespecified clinical decision thresholds.

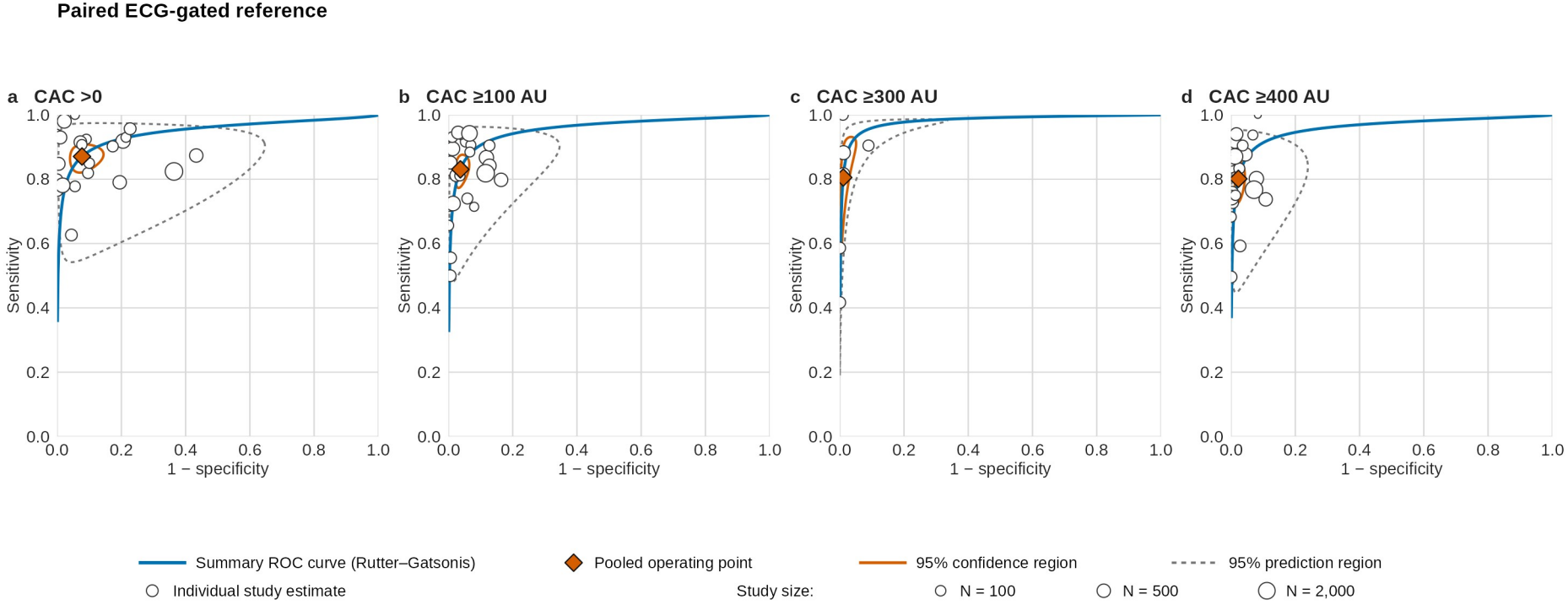
j. Publication bias was not rated as strongly suspected. Paired-reference Deeks regression suggested small-study asymmetry at CAC ≥100 (k = 19, P = 0.067), but this single statistical signal was insufficient on its own to establish selective non-publication. No asymmetry was detected at CAC >0 (k = 17, P = 0.139), and CAC ≥300 had too few reports for reliable regression-based inference (k = 4).

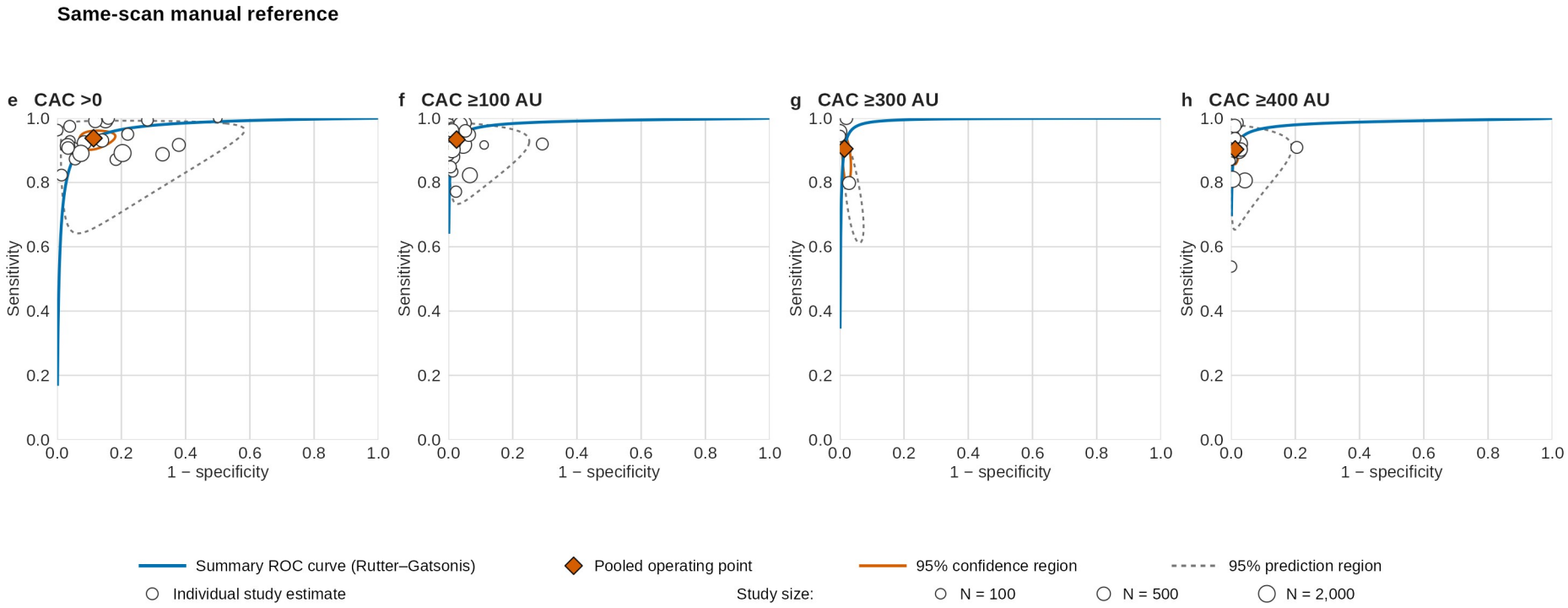
GRADE approach. Certainty in the noncontextualised diagnostic-accuracy evidence was assessed post hoc using GRADE. The target question was whether a fixed, fully automated coronary artery calcium (CAC) workflow applied to routine non-gated CT classified CAC at >0, ≥100, and ≥300 AU relative to a separate paired ECG-gated calcium-scoring CT. Certainty was rated separately for sensitivity and specificity because they inform different correct and incorrect classifications. CAC ≥400, same-scan manual references, and expert visual references were outside this certainty assessment.[123–125]

Eligible diagnostic-accuracy designs began at high certainty. Risk of bias was informed by estimate-level QUADAS-3 judgments. Indirectness was judged against the corresponding ideal paired-acquisition study in Supplementary Table 2. Inconsistency considered the distribution of sensitivity or specificity, confidence-interval overlap, prediction intervals, and model-based heterogeneity. Because no clinically validated minimum accuracy threshold was prespecified, imprecision was judged from confidence-interval width, clinical implications, and reference-positive or reference-negative information. Publication bias incorporated the search and retrieval record, contributing-report distribution, and paired-reference Deeks regression, treating asymmetry as supportive rather than determinative.

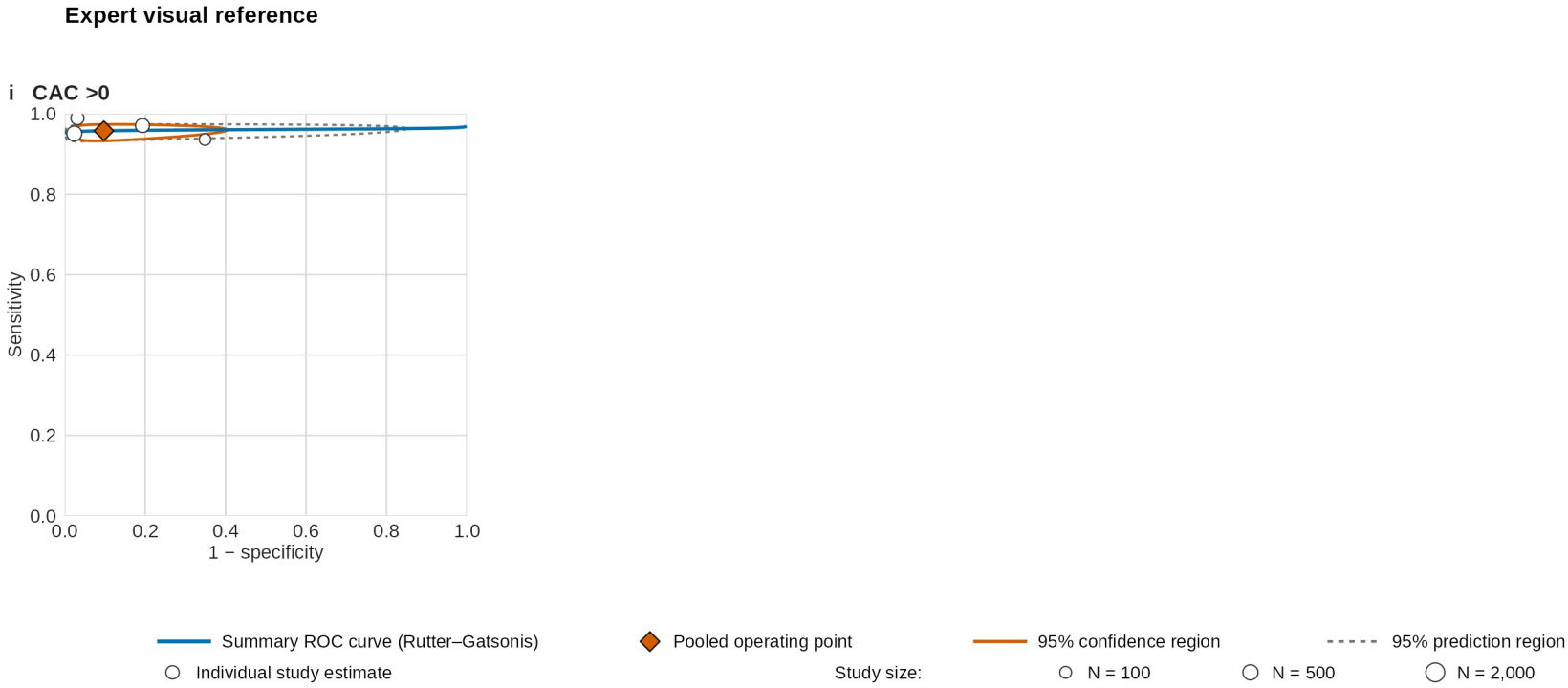
Illustrative reference-positive probabilities (60%, 40%, and 15%) are rounded medians of the contributing estimate-level proportions (62.7%, 40.4%, and 13.7%, respectively); they are not estimates of prevalence in a target screening population. Correct and incorrect classifications are calculated from the pooled accuracy and its 95% confidence interval.
AU=Agatston units; CAC, coronary artery calcium; CI, confidence interval; CT, computed tomography; ECG, electrocardiogram; FN, false negative; FP, false positive; GRADE, Grading of Recommendations Assessment, Development and Evaluation; II, insufficient information; TN, true negative; TP, true positive.

Supplementary Figure 2. Threshold-specific summary receiver operating characteristic curves. Panels a–d show paired ECG-gated reference analyses at CAC >0, ≥100 AU, ≥300 AU, and ≥400 AU; panels e–h show the corresponding same-scan manual reference analyses; panel i shows expert visual assessment of CAC >0. Blue lines are Rutter–Gatsonis SROC curves from separate bivariate random-effects models. Open circles are individual estimates, with the size aesthetic mapped to the square root of N; reference keys show N = 100, N = 500, and N = 2,000. Orange diamonds are pooled operating points. Orange solid contours show 95% confidence regions around the pooled operating points; grey dashed contours show 95% prediction regions for the expected location of a new study. The key is repeated on each page, and no curve combines thresholds.

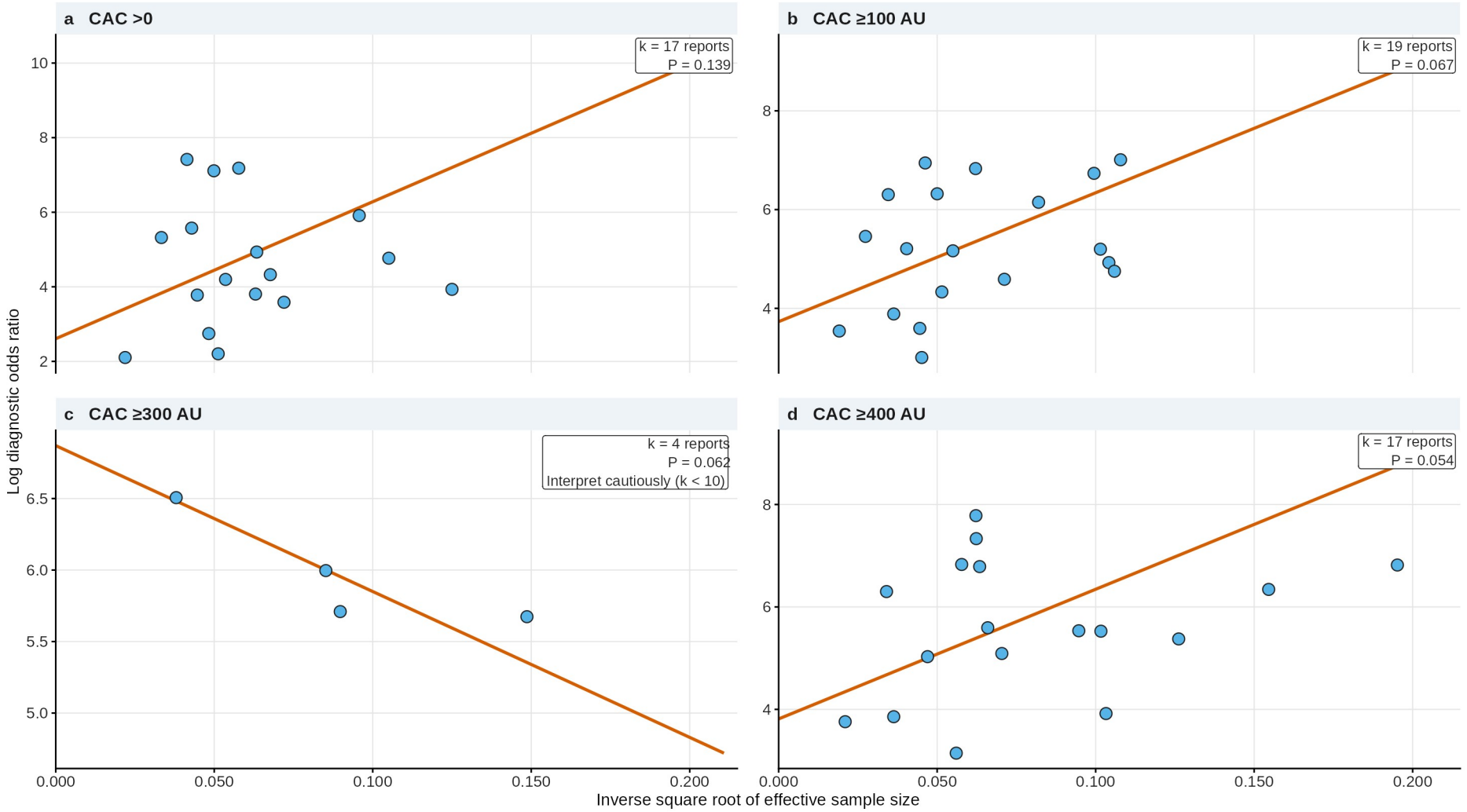



Same-scan manual reference
e CAC >0
f CAC ≥100 AU
g CAC ≥300 AU
h CAC ≥400 AU
Sensitivity
1 − specificity
0.0
0.2
0.4
0.6
0.8
1.0
Summary ROC curve (Rutter–Gatsonis)
Pooled operating point
95% confidence region
95% prediction region
Individual study estimate
Study size:
N = 100
N = 500
N = 2,000

Supplementary Figure 2 continued



Supplementary Figure 3. Deeks funnel plots for paired ECG-gated-reference analyses. Panels a–d show CAC >0, ≥100 AU, ≥300 AU, and ≥400 AU, respectively. Each point represents one report after retained non-overlapping cohorts were combined within threshold. Orange lines show Deeks weighted linear regressions of log diagnostic odds ratio on inverse square-root effective sample size, weighted by effective sample size. Two-sided P < 0.10 indicated funnel asymmetry. The CAC ≥300 analysis contained only four reports and is interpreted cautiously; asymmetry can reflect clinical or methodological heterogeneity and does not by itself establish publication bias.



Supplementary Section 4

Agreement of AI-derived CAC categories with ECG-gated CT

Supplementary Table 8. Primary four-category cross-classification analysis

Input 4×4 cross-classification tables are provided in Supplementary Data 1, sheet S11C. Panels A and B report the pooled posterior cross-classification table and the probabilities of underestimation and overestimation across category boundaries.

Supplementary Table 8, Panel A. Pooled posterior cross-classification table

Each cell is a posterior row-conditional probability. Rows are paired ECG-gated reference categories; columns are automated categories from non-gated CT.

Gated reference	Automated non-gated: 0	Automated non-gated: 1–99	Automated non-gated: 100–299	Automated non-gated: ≥300
0	Mean 95.5% Median 96.0% 95% CrI 89.5–98.4%	Mean 3.9% Median 3.4% 95% CrI 1.1–9.6%	Mean 0.3% Median 0.2% 95% CrI 0.0–1.1%	Mean 0.3% Median 0.2% 95% CrI 0.0–1.1%
1–99	Mean 30.7% Median 29.7% 95% CrI 13.8–54.1%	Mean 62.5% Median 63.3% 95% CrI 40.7–78.5%	Mean 5.8% Median 5.1% 95% CrI 1.7–14.0%	Mean 0.9% Median 0.8% 95% CrI 0.2–2.8%
100–299	Mean 2.7% Median 2.0% 95% CrI 0.4–9.5%	Mean 37.6% Median 36.6% 95% CrI 13.6–68.1%	Mean 51.9% Median 52.5% 95% CrI 26.1–74.3%	Mean 7.7% Median 6.3% 95% CrI 1.4–22.5%
≥300	Mean 0.7% Median 0.4% 95% CrI 0.0–2.8%	Mean 3.5% Median 2.7% 95% CrI 0.6–10.8%	Mean 17.6% Median 16.2% 95% CrI 5.1–39.0%	Mean 78.3% Median 79.7% 95% CrI 57.0–91.1%

Each cell reports posterior mean, median, and 95% CrI. Light grey denotes correct classification, pale orange denotes underestimation, and unshaded cells denote overestimation. Classification direction is also defined by the row and column labels and does not depend on colour. CrI, credible interval.

Supplementary Table 8, Panel B. Probabilities of underestimation and overestimation across category boundaries

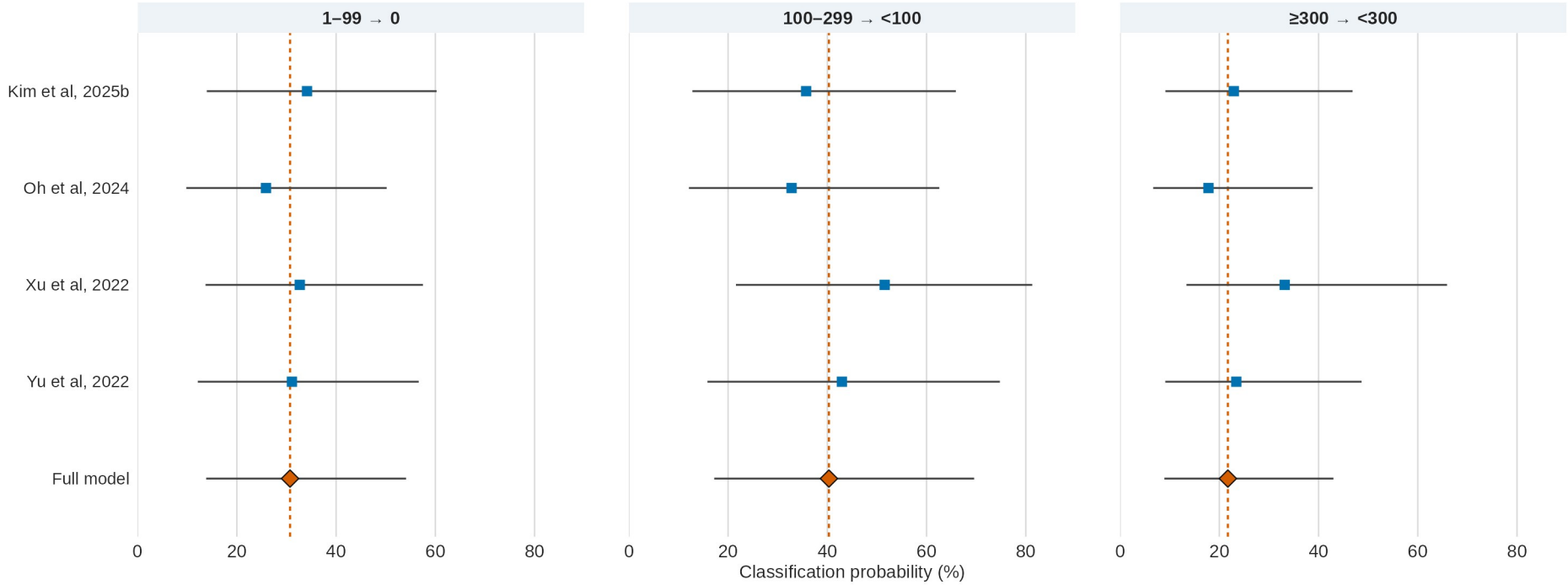
The panel reports the probabilities of underestimation and overestimation across the specified CAC category boundaries, calculated from the posterior cross-classification table.

Category boundary	Direction	Mean	Median	95% CrI
0 → >0	Overestimation	4.5%	4.0%	1.6–10.5%
0 → ≥100	Overestimation	0.6%	0.5%	0.1–1.7%
1–99 → ≥100	Overestimation	6.8%	6.1%	2.3–15.3%
100–299 → <100	Underestimation	40.3%	39.4%	17.2–69.6%
≥300 → <100	Underestimation	4.1%	3.4%	1.0–11.8%
≥300 → <300	Underestimation	21.7%	20.3%	8.9–43.0%

These probabilities may overlap and should not be summed. Individual classification probabilities are reported in the table above. AU=Agatston units; CrI, credible interval.

Convergence criteria and diagnostics. Model adequacy required maximum R-hat <1.01, minimum bulk and tail effective sample sizes ≥400, no divergent transitions, no maximum-treedepth events, and minimum E-BFMI ≥0.30. Across 21 full and leave-one-report-out fits, maximum R-hat was 1.008, minimum bulk ESS was 1,162, minimum tail ESS was 1,697, and minimum E-BFMI was 0.666; all fits met the convergence criteria.

Supplementary Figure 4. Primary-scheme leave-one-report-out analysis. The primary scheme comprised CAC 0, 1–99, 100–299, and ≥300 AU; panels show the classification probabilities for 1–99 → 0, 100–299 → <100, and ≥300 → <300 AU. Blue squares and horizontal lines are posterior means and 95% CrIs after omission of all cross-classification tables associated with the named report; related reports sharing participants were omitted together. The orange diamond is the full-model estimate; the vertical dotted line marks its posterior mean. CrI, credible interval.



Supplementary Table 9. Alternative four-category cross-classification analysis

Input 4×4 cross-classification tables are provided in Supplementary Data 1, sheet S11D. Panels A and B report the pooled posterior cross-classification table and the probabilities of underestimation and overestimation across category boundaries.

Supplementary Table 9, Panel A. Pooled posterior cross-classification table

Each cell is a posterior row-conditional probability. Rows are paired ECG-gated reference categories; columns are automated categories from non-gated CT.

Gated reference	Automated non-gated: 0	Automated non-gated: 1–99	Automated non-gated: 100–399	Automated non-gated: ≥400
0	Mean 91.8% Median 92.1% 95% CrI 86.3–95.6%	Mean 6.6% Median 6.3% 95% CrI 3.0–12.1%	Mean 1.1% Median 1.0% 95% CrI 0.3–2.3%	Mean 0.6% Median 0.5% 95% CrI 0.2–1.3%
1–99	Mean 30.7% Median 30.5% 95% CrI 20.7–41.9%	Mean 61.3% Median 61.4% 95% CrI 50.9–70.8%	Mean 6.8% Median 6.6% 95% CrI 3.6–11.3%	Mean 1.2% Median 1.2% 95% CrI 0.5–2.4%
100–399	Mean 2.9% Median 2.8% 95% CrI 1.5–5.0%	Mean 28.9% Median 28.8% 95% CrI 20.2–38.5%	Mean 61.7% Median 61.7% 95% CrI 53.3–70.2%	Mean 6.5% Median 6.3% 95% CrI 3.7–10.3%
≥400	Mean 0.8% Median 0.7% 95% CrI 0.4–1.4%	Mean 3.0% Median 3.0% 95% CrI 1.8–4.6%	Mean 15.7% Median 15.5% 95% CrI 11.2–20.9%	Mean 80.5% Median 80.6% 95% CrI 75.3–85.0%

Each cell reports posterior mean, median, and 95% CrI. Light grey denotes correct classification, pale orange denotes underestimation, and unshaded cells denote overestimation. Classification direction is also defined by the row and column labels and does not depend on colour. CrI, credible interval.

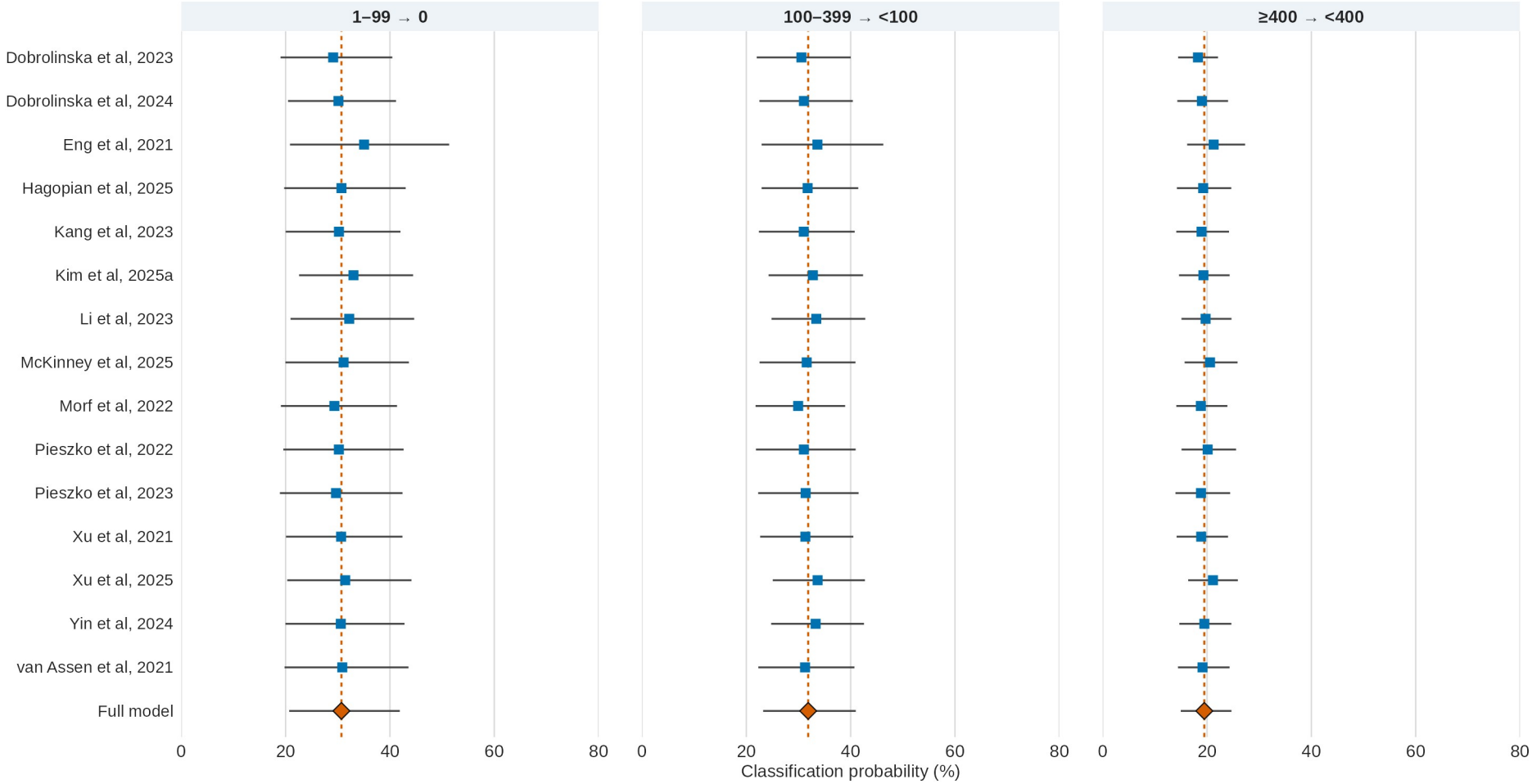
Supplementary Table 9, Panel B. Probabilities of underestimation and overestimation across category boundaries

The panel reports the probabilities of underestimation and overestimation across the specified CAC category boundaries, calculated from the posterior cross-classification table.

Category boundary	Direction	Mean	Median	95% CrI
0 → >0	Overestimation	8.2%	7.9%	4.4–13.7%
0 → ≥100	Overestimation	1.6%	1.6%	0.7–3.1%
1–99 → ≥100	Overestimation	8.0%	7.9%	4.6–12.5%
100–399 → <100	Underestimation	31.9%	31.8%	23.2–41.0%
≥400 → <100	Underestimation	3.8%	3.8%	2.5–5.5%
≥400 → <400	Underestimation	19.5%	19.4%	15.0–24.7%

These probabilities may overlap and should not be summed. Individual classification probabilities are reported in the table above. AU=Agatston units; CrI, credible interval.

Supplementary Figure 5. Alternative-scheme leave-one-report-out analysis. The alternative scheme comprised CAC 0, 1–99, 100–399, and ≥400 AU; panels show the classification probabilities for 1–99 → 0, 100–399 → <100, and ≥400 → <400 AU. Plotting conventions match Supplementary Fig. 4. The alternative set of reports was fitted separately and was not combined with the primary scheme.



Supplementary Table 10. Overall CAC severity-category agreement

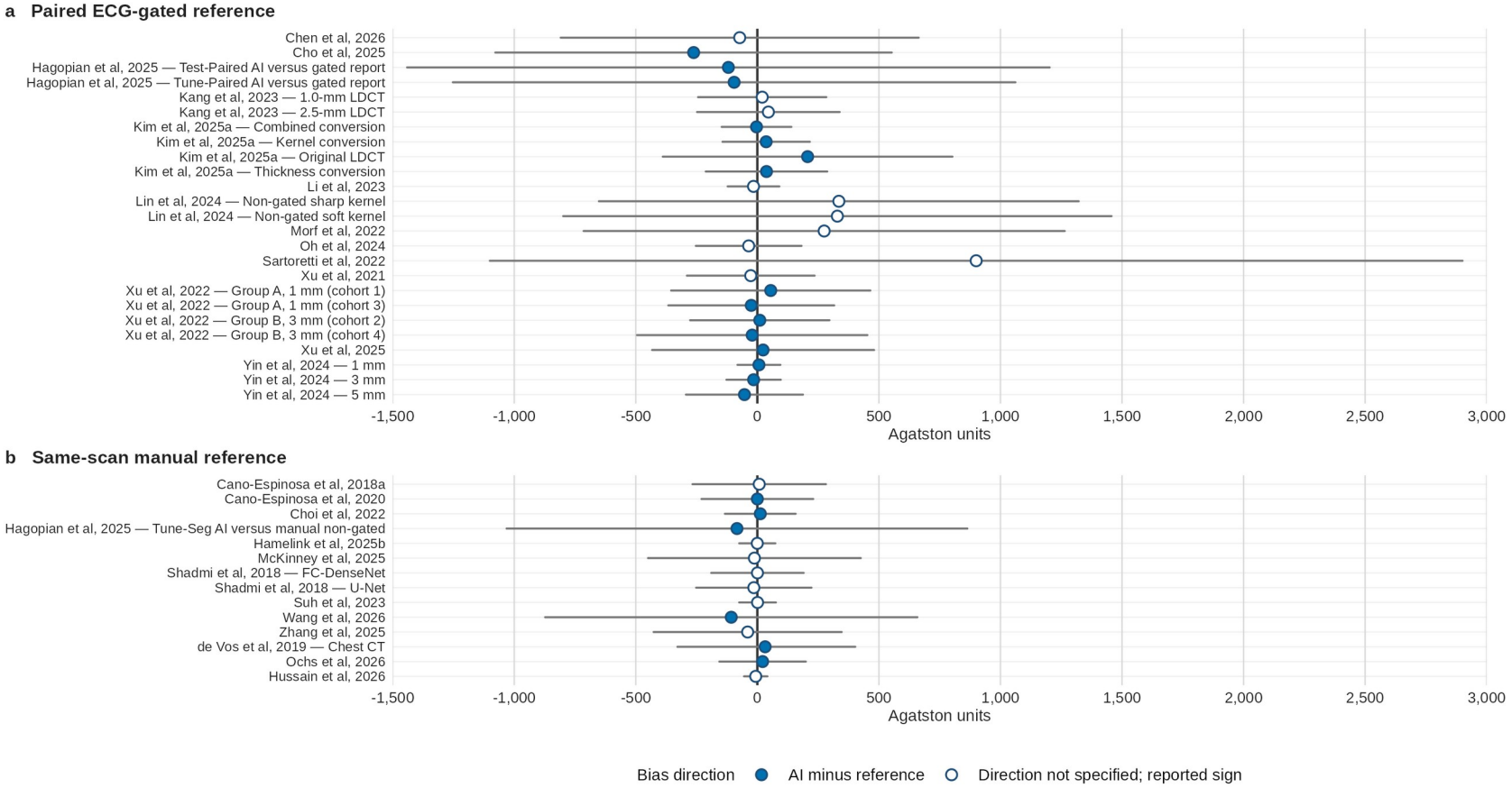
Reference standard	Classification scheme	Reports	Participants	Pooled linear-weighted κ (95% CI)	I^2
Paired ECG-gated CT	300-AU	4	1,727	0.78 (0.60–0.96)	96.3%
Paired ECG-gated CT	400-AU	15	7,935	0.75 (0.68–0.81)	96.9%
Manual scoring on the same non-gated CT	300-AU	4	1,416	0.90 (0.84–0.97)	80.7%
Manual scoring on the same non-gated CT	400-AU	14	9,654	0.83 (0.78–0.88)	96.9%

The 300-AU scheme comprised CAC 0, 1–99, 100–299, and ≥300 AU; the 400-AU scheme comprised CAC 0, 1–99, 100–399, and ≥400 AU.

CAC, coronary artery calcium. CT, computed tomography. ECG, electrocardiogram.

Supplementary Section 5

Error attribution, image-domain effects, and cross-domain measurement calibration



Supplementary Figure 6. Agatston-unit Bland–Altman bias and 95% limits of agreement. Points show mean bias and horizontal lines show 95% limits of agreement for 39 comparisons from 26 reports. Panel a compares AI scores on non-gated CT with paired ECG-gated CT (25 comparisons from 14 reports); panel b compares AI with manual scoring on the same non-gated images (14 comparisons from 13 reports). One report contributes to both panels. Filled points use an explicit or harmonised AI-minus-reference direction; open points retain the source-reported sign where direction was not specified. Comparisons may share participants and were not pooled. Other comparison types and measurement scales are retained in Supplementary Data 1, sheet S5.

Supplementary Table 11. Human and automated assessment of the same non-gated image

Six reports compared human and AI assessment of the same non-gated images against a common paired ECG-gated reference. None provided a formal paired test of human-versus-AI superiority. Results were summarised descriptively.

Study	N	Non-gated input	Human method	AI method	Common gated reference	Category scheme	Human result	AI result	Error direction	Interpretation and limitations
Dobrolinska et al, 2023	213	PET/LDCT attenuation-correction CT	Manual Agatston scoring and visual six-category assessment	syngo.via CT CaScoring VB50 automatic Agatston score	Manual Agatston score on dedicated CSCT	0 / 1–10 / 11–100 / 101–400 / 401–1000 / >1000	Manual: linear κ 0.58 (95% CI 0.52–0.63), exact 35.2%, within-one 94.4%. Visual: κ 0.82 (0.77–0.87), exact 74.2%.	Linear κ 0.50 (95% CI 0.44–0.56), exact 29.6%, within-one 83.6% recalculated.	Manual and AI Agatston scoring both underestimated; visual assessment was descriptively closer to reference.	AI sensitivity denominators and within-one-category agreement were inconsistent in the report; recalculated values are identified. Separate matrices do not recover paired human–AI discordance.
McKinney et al, 2025	331 paired-reference subset	The same non-gated chest CT series	Three-radiologist consensus Agatston category	Riverain ClearRead CT CAC v1.1.1.37	Prior paired gated-CT clinical Agatston report	0 / 1–99 / 100–399 / ≥400	Quadratic-weighted κ 0.907 (95% CI 0.883–0.930), exact 79.5%, Spearman ρ 0.941.	Quadratic-weighted κ 0.906 (95% CI 0.882–0.927), exact 79.2%, Spearman ρ 0.942.	Near-identical descriptive performance; no material directional difference was demonstrated.	Separate matrices do not recover paired human–AI discordance. Quadratic-weighted κ differs from the linear κ used in other reports.
Oh et al, 2024	210	PET/CT attenuation-correction CT	Two experienced readers using Agatston scoring	AVIEW CAC v1.1.40	Reader-derived Agatston score on ECG-gated CaCT	0 / 1–99 / 100–299 / ≥300	Weighted κ 0.662, exact 77.6%, Spearman ρ 0.837, ICC 0.911, bias −27.1 AU.	Weighted κ 0.615, exact 73.8%, Spearman ρ 0.768, ICC 0.842, bias −35.7 AU.	Both underestimated; AI had the more negative bias and a modest descriptive disadvantage.	Bowker tests assessed symmetry within each matrix, not human-versus-AI superiority; the three-way patient-level classification was unavailable.
Yu et al, 2022	405	The same non-gated chest CT	Blinded vessel-specific CAC identification and vessel count	CACScoreDoc	Two-radiologist manual gated cardiac-CT vessel assessment	Matched vessel-count categories N0–N4	Linear κ 0.821 (95% CI 0.784–0.858), exact 79.0%, sensitivity 90.2%, specificity 94.1%.	Linear κ 0.75 (95% CI 0.70–0.79), exact 76.0%, sensitivity 84.8%, specificity 99.5%.	Both missed vessel-positive cases; AI was less sensitive but more specific.	The endpoint was vessel count, not Agatston severity. Aggregate matrices did not recover paired human–AI discordance.
Xu et al, 2022	144 per test group (1 mm and 3 mm)	Paired non-gated chest CT reconstructed at 1 mm or 3 mm	Radiologist manual Agatston score	Two-stage deep-learning Agatston scorer	Clinical manual score from paired dedicated gated CT	0 / 1–99 / 100–299 / ≥300; human categories not published	Bland–Altman bias: −21.6 AU (1 mm) and −49.2 AU (3 mm).	Bland–Altman bias: −55.0 AU (1 mm) and −10.0 AU (3 mm); AI category matrices were reported.	Predominantly downward bias, with magnitude dependent on thickness and scorer.	Human agreement was available only from Bland–Altman plots; human category matrices and threshold accuracy were unavailable.
Morf et al, 2022	15 exploratory supplement subset	The same ungated PET/CT CT	Manual PET/CT Agatston score	AVIEW CAC	Two-physician consensus on gated SPECT-MPI CT	0 / 1–100 / 101–400 / >400	Linear κ 0.660, exact 66.7%, MAE 144.0 AU, bias −83.5 AU.	Linear κ 0.526, exact 46.7%, MAE 168.1 AU, bias −167.7 AU.	Both underestimated; AI showed the larger downward error.	Exploratory subset of 15 patients. The paired comparison was documented in the supplementary data despite a conflicting limitation statement in the article.

CI, confidence interval; CSCT, calcium-scoring CT; MAE, mean absolute error.

Supplementary Table 12. Same algorithm, same patients, alternate inputs

Comparisons are ordered from the most controlled input contrast to designs in which preprocessing or time also changed. Results are reported on their original scales. AI, artificial intelligence; CAC, coronary artery calcium; CCC, concordance correlation coefficient; CSCT, calcium-scoring CT; LDCT, low-dose CT; ULDCT, ultra-low-dose CT.

Study	N	What was held constant	What changed	Comparator	Direct paired statistic	Key result	Limitation
1. Same raw data with different reconstruction							
Chen et al, 2026 — three reconstructions	1,213	Same patients, raw non-gated CT, CACScoreDoc, and common gated reference.	Slice thickness and reconstruction kernel/window series	Paired ECG-gated manual CACS	No— all 3,639 series pooled	Pooled rho 0.865; kappa 0.641; reported exact accuracy 77.5%, but matrix diagonal is 75.3%.	Results pooled 3,639 reconstructions from 1,213 patients; reconstruction-specific estimates were unavailable.
Kang et al, 2023 — CSCT and 1.0/2.5-mm LDCT	567	Within N = 453 LDCT subset: same raw LDCT, AVIEW CAC, patients, and gated reference.	Acquisition/gating plus LDCT slice thickness	Manual CSCT Agatston score	Arm estimates vs common reference; no direct LDCT1-vs-LDCT2.5 test	Corrected Table 2: ICC 0.937 for 2.5 mm and 0.955 for 1.0 mm; linear kappa 0.776 and 0.809.	Only 453 of 567 patients had the 1.0-mm reconstruction; the software version was not reported.
Yin et al, 2024 — 1/3/5-mm reconstructions	112	Same scan, CACScoreDoc v6.11, patients, and common gated reference; no human correction.	Slice thickness	Paired manual ECG-gated CACS and direct 1-vs-3-mm comparison	Yes for 1 vs 3 mm; arm estimates for all three	Vs gated reference: rho 0.973/0.941/0.834 and exact 91.0%/84.8%/62.5% for 1/3/5 mm.	A direct paired comparison was reported for 1 versus 3 mm; other reconstruction results were comparisons against gated CT.
Lin et al, 2024 — gated/soft/sharp kernels	76	For soft/sharp arms: same chest-CT raw data, AVIEW CAC, patients, and gated reference.	Gating/acquisition and reconstruction kernel	Manual Agatston on paired gated CCT	Arm estimates vs common reference; no soft-vs-sharp paired test	ICC 1.000 (gated AI), 0.929 (soft), 0.948 (sharp); kappa 1.000, 0.556, 0.537.	No direct paired test of soft versus sharp reconstruction was reported.
Wang et al, 2026 — 1.5 versus 5.0 mm	346	Same raw projections, patients, and explicitly identical model version.	Slice thickness and kernel	Thin-slice AI treated as paired comparator	Direct paired AI-to-AI statistic reported.	Pearson 0.968; Spearman 0.962; linear kappa 0.770; exact 238/346 (68.8%).	AI measurements on the thin-slice reconstruction were the comparator; this assesses reconstruction agreement rather than accuracy against gated CT. Slice thickness and kernel changed together.
Wang et al, 2026 — 1.0 versus 5.0 mm	420	Same raw projections, patients, and explicitly identical model version.	Slice thickness and kernel	Thin-slice AI treated as paired comparator	Direct paired AI-to-AI statistic reported.	Pearson 0.929; Spearman 0.895; linear kappa 0.705; exact 271/420 (64.5%).	AI measurements on the thin-slice reconstruction were the comparator; this assesses reconstruction agreement rather than accuracy against gated CT. Slice thickness and kernel changed together.
Wang et al, 2026 — 2.0 versus 5.0 mm	852	Same raw projections, B50f kernel, patients, and explicitly identical model version.	Slice thickness only	Thin-slice AI treated as paired comparator	Direct paired AI-to-AI statistic reported.	Pearson 0.972; Spearman 0.960; linear kappa 0.816; exact agreement 653/852 in Table 2 versus 654/852 in Table 5.	AI measurements on the thin-slice reconstruction were the comparator; this assesses reconstruction agreement rather than accuracy against gated CT.

Supplementary Table 12 continued

Study	N	What was held constant	What changed	Comparator	Direct paired statistic	Key result	Limitation
2. Fixed downstream scorer after controlled image transformation							
Kim et al, 2025a — four LDCT variants	225	Same LDCT, patients, AVIEW CAC v1.1.43 downstream scorer, and gated reference.	Deep-learning kernel conversion and/or 1–1.25 to 3-mm thickness conversion	Manual Agatston on paired gated CSCT	Arm estimates vs common reference; no direct AI-vs-AI hypothesis test	CCC improved from 0.269 (original) to 0.881 (combined); bias from 206.7 to -3.45; kappa from 0.115 to 0.792.	The combined conversion changed both kernel and thickness; the reported combined improvement does not separate their contributions.
3. Same-breath-hold or same-day dose/respiration experiment							
Ledda et al, 2025 — LDCT versus ULDCT	327	Same patients, one breath-hold, and AVIEW pipeline.	Radiation dose/spectral shaping and ULDCT kernel	AI-derived LDCT score used as paired comparator	Direct paired AI-to-AI statistic reported.	Pearson r 0.96; ICC 0.86; weighted kappa 0.86; exact category agreement 84.0%.	Dose and reconstruction differed, and AI-derived LDCT scores were the comparator rather than gated CT.
van Velzen et al, 2022 — ClinicalA	1,543	Same patients, same day, and the same ClinicalA comparator.	Respiratory instruction/acquisition image	Direct interscan reproducibility	Direct paired AI-to-AI statistic reported.	All 1,543 pairs: CAC-mass ICC 0.93, Agatston ICC 0.91, kappa 0.81, exact category agreement 92%.	The comparison assesses interscan reproducibility; model checkpoint identifiers were not reported.
van Velzen et al, 2022 — proposed model	1,543	Same patients, same day, and the same trained proposed model.	Respiratory instruction/acquisition image	Direct interscan reproducibility	Direct paired AI-to-AI statistic reported.	All 1,543 pairs: CAC-mass ICC 0.97, adjusted-Agatston ICC 0.96, kappa 0.84, exact category agreement 90%.	The comparison assesses interscan reproducibility; model checkpoint identifiers were not reported.
4. Paired gated and non-gated acquisitions							
Hamelink et al, 2025b — cardiac versus chest CT	684	Same patients and AI-Rad Companion prototype version 08/2022; final reconstruction matched.	Gating, dose/current, pitch, and scan coverage; final reconstructions matched at 3 mm/Qr36	Manual CACS on each corresponding modality	Parallel within-patient arms; no AI-vs-AI paired test	Cardiac vs chest: sensitivity 98.1% vs 92.3%; ICC 0.997 vs 0.992; kappa 0.952 vs 0.901.	Several acquisition parameters changed between cardiac and chest CT; no direct paired AI-to-AI test was reported.
Suh et al, 2023 — gated versus non-gated	452	Same patients and named AVIEW CAC software; direct AI-to-AI comparison.	Gating, acquisition protocol, thickness/kernel	Direct AI-on-AI plus modality-matched manual CACS	Direct paired AI-to-AI statistic reported.	Direct auto CSCT-vs-LDCT: ICC 0.938, bias 26.2 ±189.1, volume ICC 0.949, linear kappa 0.703, exact 67.3%.	Gating, acquisition, and reconstruction differed; the exact software version was not reported.
5. Preprocessing- or time-confounded comparison							
Kim et al, 2025b — gated versus non-gated	652	Same patients and ClariCardio 2D U-Net core scorer.	Gating/acquisition and modality-specific denoising	Manual CACS on gated calcium CT	Parallel arms vs same gated manual reference	Gated vs non-gated: ICC 0.997 vs 0.968; exact category accuracy 96.8% vs 86.0%; kappa 0.97 vs 0.88.	Only LDCT received additional denoising, so acquisition and preprocessing effects cannot be separated.
Morf et al, 2022 — gated versus ungated	15	Same 15 patients and unchanged AVIEW algorithm; no study training.	Gating/acquisition/ modality	Direct AI-on-AI; manual gated CACS supplied as context	Calculated directly from published patient-level supplement data.	Direct AI gated-vs-ungated: Pearson 0.874, CCC 0.685, linear kappa 0.528, exact 53.3% (N = 15).	Exploratory subset of 15 patients; scans could be up to 6 months apart.

Supplementary Section 6

Subgroups and meta-regression

Supplementary Table 13. Prespecified subgroup and meta-regression analyses

Panel A summarises the multiplicity-adjusted global test pattern; Panel B reports the fitted global covariate tests; Panel C describes reconstruction thickness, product concentration, and commercial status by publication period for the CAC ≥100 evidence base. Component coefficients and absolute subgroup estimates for the FDR-significant tests are shown in Supplementary Figs. 7 and 8. Complete machine-readable results, including non-fitted rows and reasons, are supplied in Supplementary Data 1, sheet S6.

Panel A. FDR q-value grid

Cells show Benjamini–Hochberg FDR q values for the 33 fitted global likelihood-ratio tests. Darker blue denotes smaller q; bold values indicate q < 0.05. Complete likelihood-ratio statistics, degrees of freedom, P values, fitted levels, and non-fitted domains are provided in Supplementary Data 1, sheet S6. No CAC ≥300 covariate met the minimum information requirement for fitting.

Covariate	CAC >0	CAC ≥100	CAC ≥400
Commercial status	**0.013**	**<0.0001**	**0.0006**
Country or region	0.105	0.061	**0.0050**
CT clinical setting	0.669	0.070	0.089
Model architecture	0.468	0.357	**0.012**
Paired-scan timing	0.668	0.438	0.654
Publication year	0.105	0.732	0.654
Publication-year band	**0.025**	0.383	0.357
Reconstruction kernel	Not tested	0.419	Not tested
Scanner vendor	0.291	0.357	0.123
Slice thickness	0.468	0.357	Not tested
Training ground-truth pairing	0.452	0.419	0.654
Validation design	0.668	0.419	0.654

Colour key: dark blue, q < 0.001; medium blue, q = 0.001–<0.01; light blue, q = 0.01–<0.05; unshaded, q ≥ 0.05; grey, not tested.

Supplementary Table 13 continued

Panel B. Fitted global covariate tests

The 33 fitted likelihood-ratio tests are shown. All 56 prespecified covariate–threshold rows, including 23 descriptive or non-fitted rows and reasons, are provided in Supplementary Data 1, sheet S6.

Covariate	CAC >0				CAC ≥100				CAC ≥400			
	Reports / estimates N	LRT χ^2 (df)	P	FDR q	Reports / estimates N	LRT χ^2 (df)	P	FDR q	Reports / estimates N	LRT χ^2 (df)	P	FDR q
Commercial status	17 / 24 9,393	13.508 (2)	0.0012	**0.013**	18 / 25 10,609	29.524 (2)	<0.0001	**<0.0001**	17 / 22 9,543	19.587 (2)	<0.0001	**0.0006**
Country or region	17 / 24 9,393	7.104 (2)	0.029	0.105	19 / 26 10,940	13.224 (4)	0.010	0.061	17 / 22 9,543	18.478 (4)	0.0010	**0.0050**
CT clinical setting	17 / 24 9,393	2.363 (4)	0.669	0.669	19 / 26 10,940	11.993 (4)	0.017	0.070	17 / 22 9,543	6.671 (2)	0.036	0.089
Model architecture	16 / 23 9,062	2.206 (2)	0.332	0.468	17 / 24 9,275	3.762 (2)	0.152	0.357	15 / 20 7,878	15.514 (4)	0.0037	**0.012**
Paired-scan timing	16 / 22 8,687	1.078 (2)	0.583	0.668	18 / 24 10,234	1.826 (2)	0.401	0.438	15 / 20 8,831	0.977 (2)	0.613	0.654
Publication year	17 / 24 9,393	13.324 (6)	0.038	0.105	19 / 26 10,940	5.235 (8)	0.732	0.732	17 / 22 9,543	5.372 (6)	0.497	0.654
Publication-year band	17 / 24 9,393	15.076 (4)	0.0045	**0.025**	19 / 26 10,940	5.694 (4)	0.223	0.383	17 / 22 9,543	5.806 (4)	0.214	0.357
Reconstruction kernel	—	—	—	—	10 / 11 5,613	2.178 (2)	0.337	0.419	—	—	—	—
Scanner vendor	15 / 22 8,555	4.044 (2)	0.132	0.291	17 / 24 10,102	3.507 (2)	0.173	0.357	14 / 19 8,500	5.579 (2)	0.061	0.123
Slice thickness	10 / 11 6,470	2.156 (2)	0.340	0.468	11 / 12 6,683	3.448 (2)	0.178	0.357	—			
Training ground-truth pairing	17 / 24 9,393	2.801 (2)	0.246	0.452	18 / 25 9,606	2.320 (2)	0.313	0.419	16 / 21 8,209	1.278 (2)	0.528	0.654
Validation design	16 / 23 9,281	0.997 (2)	0.608	0.668	17 / 24 9,494	2.103 (2)	0.349	0.419	15 / 20 8,097	0.849 (2)	0.654	0.654

Fitted levels

Covariate	CAC >0	CAC ≥100	CAC ≥400
Commercial status	Commercial; Noncommercial		
Country or region	East Asia; North America	East Asia; Europe; North America	North America; East Asia; Europe
CT clinical setting	General chest CT; Lung-screening LDCT; Oncology/PET-CT attenuation-correction or cancer-related CT		General chest CT; Oncology/PET-CT attenuation-correction or cancer-related CT
Model architecture	U-Net family; Other CNN family		U-Net family; Other CNN family; Transformer-based
Paired-scan timing	Longer or mixed interval; All pairs within 1 month		
Publication year	2025; 2021; 2022; 2024	2025; 2021; 2022; 2023; 2024	2025; 2021; 2023; 2026
Publication-year band	2025–2026; 2019–2022; 2023–2024		
Reconstruction kernel	Not tested	Mixed or variable kernel; Soft or standard kernel	Not tested
Scanner vendor	Multivendor/mixed; Siemens only		
Slice thickness	Intermediate (>1.25 to 3 mm); Thin (≤1.25 mm)		Intermediate (>1.25 to 3 mm)
Training ground-truth pairing	Paired ECG-gated training ground truth; Non-paired training ground truth		
Validation design	External validation; Internal validation		

Benjamini–Hochberg adjustment is applied within each threshold; bold q values are <0.05. df, degrees of freedom. Model-specific reference levels are provided in Supplementary Data 1, sheet S6.

[1] Slice thickness at CAC ≥400 did not meet the fitting criteria. No covariate was fitted at CAC ≥300 (four reports). Descriptive and other non-fitted rows remain in Supplementary Data 1, sheet S6.

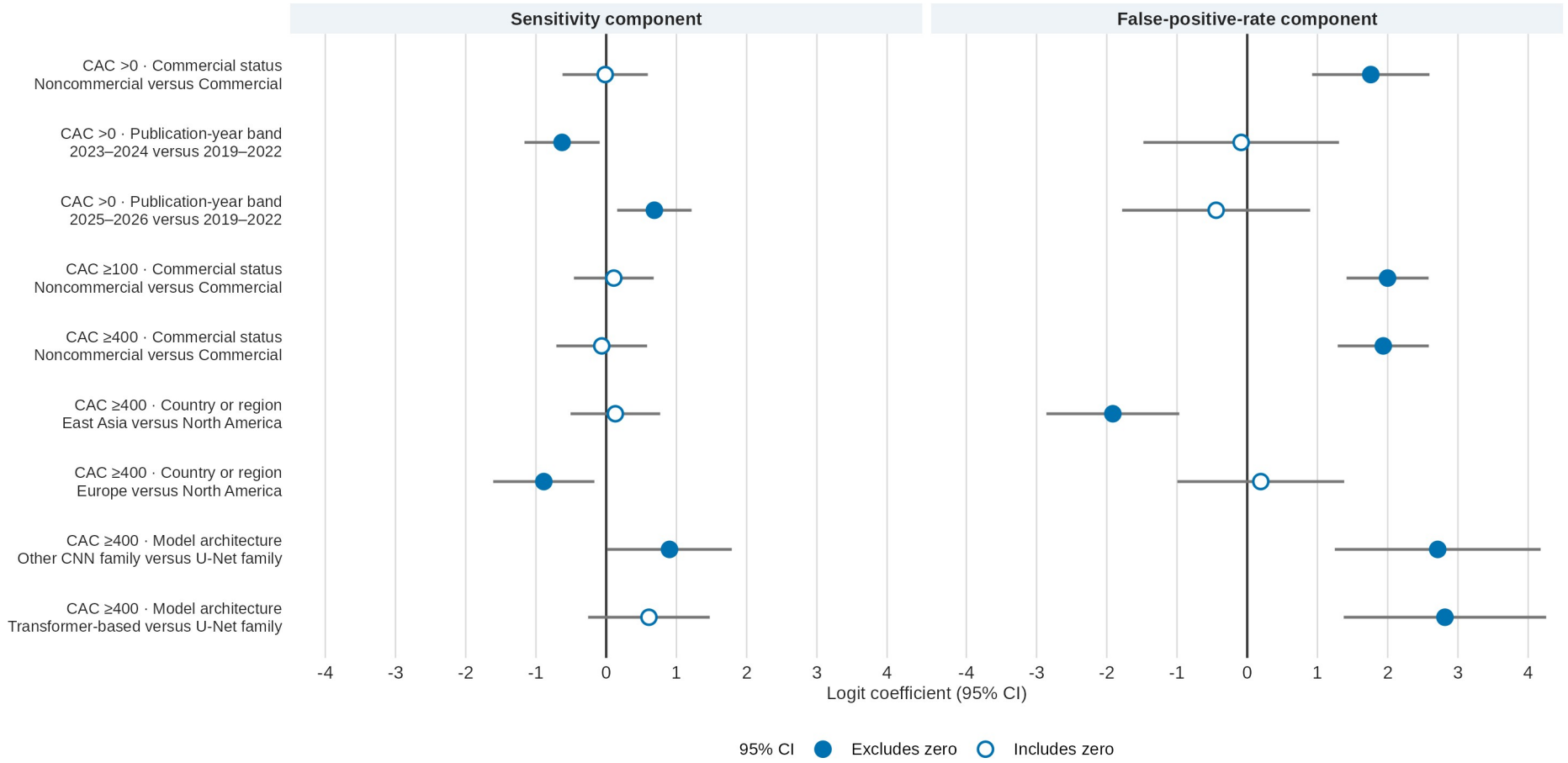
FDR, false discovery rate; LRT, likelihood-ratio test; ML, maximum likelihood.

Panel C. Descriptive context for the CAC ≥100 analysis

These summaries describe the evidence structure and do not constitute additional covariate tests.

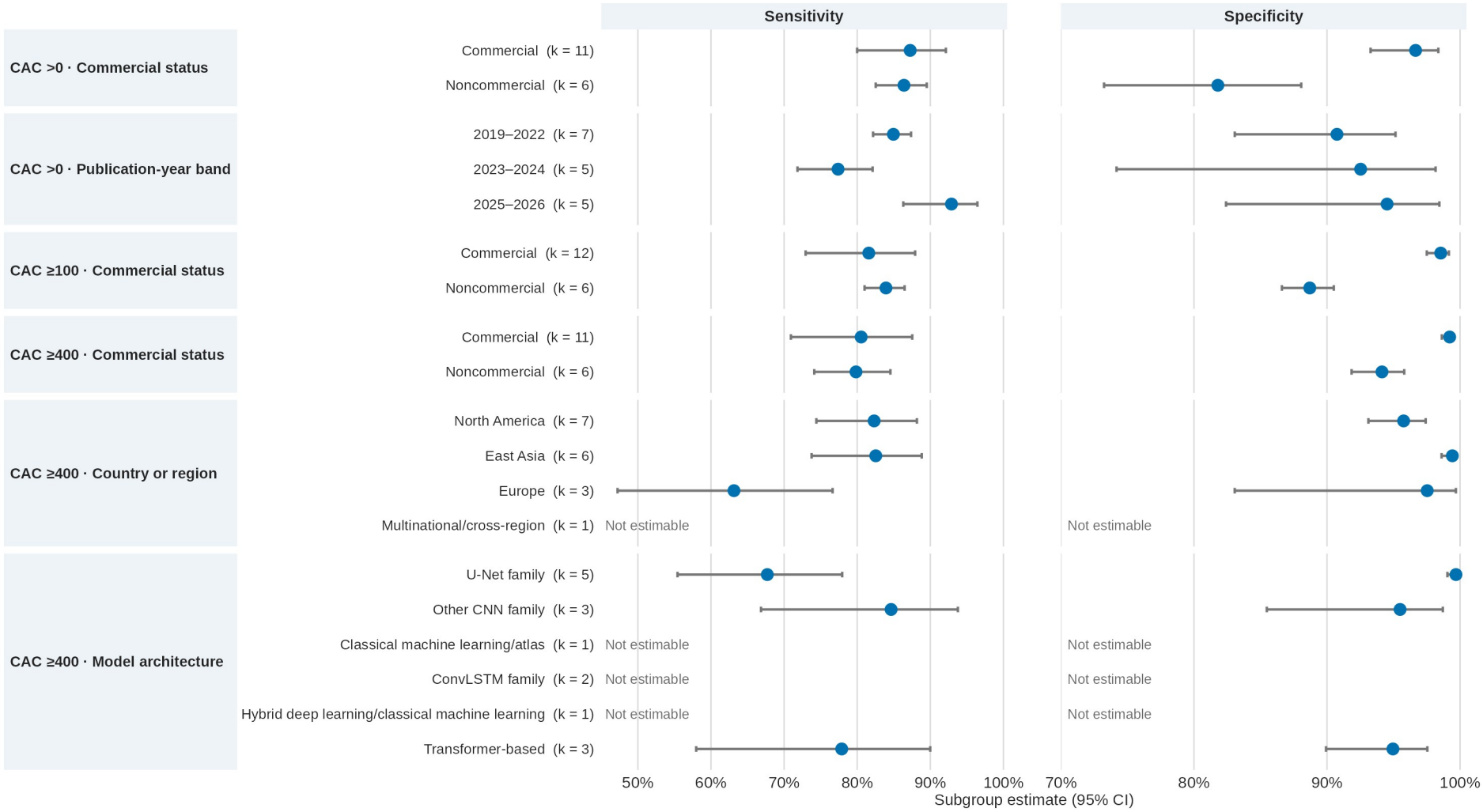
Feature	Observation
Evidence base	26 estimates from 19 reports (N = 10,940); 13 product or model families.
Reconstruction thickness	≤1.25 mm: 5 estimates/4 reports; >1.25–3 mm: 7 estimates/7 reports; >3 mm: 3 estimates/2 reports; cross-band: 4 estimates/4 reports; not reported: 7 estimates/4 reports. The between-report global covariate test was not significant after FDR adjustment (q = 0.357).
Product concentration	The largest family by model inputs was Eng DL/XGBoost model (6/26 estimates; 23.1%); the largest participant share was Cedars-Sinai ConvLSTM (3,244/10940; 29.7%).
Commercial status by publication period	Reports in 2019–22/2023–24/2025–26: Commercial: 4/4/4; Noncommercial: 3/2/1; Unclear: 0/0/1.

Supplementary Figure 7. Component coefficients within FDR-significant meta-regression tests



Points show logit-scale component coefficients and horizontal lines show 95% confidence intervals. The vertical line marks the null value of zero. Every component contrast nested within the six FDR-significant global meta-regression tests is shown in both panels, irrespective of its component Wald P value. Filled points have confidence intervals excluding zero; open points do not. Positive false-positive-rate coefficients indicate lower specificity relative to the stated reference level. Component Wald tests were not separately FDR-adjusted.

Supplementary Figure 8. Subgroup sensitivity and specificity for FDR-significant covariates



Points show absolute subgroup sensitivity and specificity and horizontal lines show 95% confidence intervals from level-specific Reitsma bivariate logit random-effects models fitted by restricted maximum likelihood. All 19 subgroup levels belonging to the six FDR-significant global tests are retained; four sparse levels without an estimable pooled result are explicitly labelled Not estimable. Values are descriptive subgroup estimates rather than direct paired effects. Complete counts, prediction intervals, and non-estimable levels are provided in Supplementary Data 1, sheet S6.

Supplementary Section 7

Prognostic evidence and certainty

Supplementary Table 14. Prognostic report characteristics

Characteristics are shown for the six reports contributing adjusted hazard ratios to the primary category-based MACE synthesis and seven reports providing additional, non-exchangeable prognostic evidence. Treatment adjustment refers to covariates in the selected prognostic model, not merely baseline treatment reporting.

Report	Cohort	N	Setting	AI method	CAC categories	Outcome definition	Follow-up	Events	Adjustment covariates	Statin/treatment adjustment
Miller et al, 2023	External testing cohort	2,271	SPECT/CT myocardial perfusion imaging with low-dose CT attenuation correction; known coronary artery disease excluded	Cascaded ConvLSTM CAC model	0; 1–100; 101–400; >400 AU	Death, revascularisation, non-fatal myocardial infarction, or unstable-angina admission	2.8 years	320	Age; sex; BMI; hypertension; diabetes; dyslipidaemia; family history; smoking; stress/rest perfusion measures; stress LVEF	No treatment covariate reported
Pieszko et al, 2023	Paired subgroup without prior coronary artery disease	2,502	PET myocardial perfusion imaging with low-dose PET/CT attenuation correction	ConvLSTM CAC segmentation and scoring model	0; 1–100; 101–400; >400 AU	Death, late revascularisation (>90 days), unstable-angina admission, or myocardial infarction	4.3 years	Not separately reported for the selected subgroup	Myocardial flow reserve; summed difference score	No treatment covariate reported
Dobrolinska et al, 2024	LDCT evaluation cohort	507	Consecutive patients with suspected coronary artery disease undergoing [15O]-water PET myocardial perfusion imaging, LDCT, and dedicated calcium-scoring CT	Deep-learning CAC-map/pseudo mass scoring from LDCT	0; 1–100; 101–400; >400 AU	Death, non-fatal myocardial infarction, late coronary revascularisation, or unstable angina	6.6 years	Not separately reported; 95 in the full n = 572 cohort	[15O]-water PET ischaemia/perfusion defect	No; baseline statin use was reported but not included in the selected model
Williams et al, 2024	External test cohort 2	5,969	SPECT/CT myocardial perfusion imaging with non-gated non-contrast CT attenuation correction	Vessel-specific ConvLSTM deep-learning CAC model	0; 1–100; 101–400; >400 AU	Death, revascularisation, myocardial infarction, or unstable-angina admission	2.3 years	841	Age; sex; BMI; hypertension; diabetes; dyslipidaemia; family history; smoking	No treatment covariate reported

Supplementary Table 14 continued

Report	Cohort	N	Setting	AI method	CAC categories	Outcome definition	Follow-up	Events	Adjustment covariates	Statin/treatment adjustment
Lee et al, 2025	Two-centre health-screening cohort	5,342	Asymptomatic adults aged ≥40 years undergoing non-gated non-contrast chest CT	ClariCardio version 1.0	0; 1–100; 101–300; 301–1000; >1000 AU	Unstable angina requiring admission or revascularisation, myocardial infarction, ischaemic stroke, or cardiovascular death	7.5 years	383	AI-derived thoracic aortic calcium; age; sex; BMI; smoking; hypertension; diabetes; dyslipidaemia; haemoglobin; HDL; LDL; creatinine; blood urea nitrogen	No treatment covariate reported
Yang et al, 2026	Chronic kidney disease chest-CT cohort	509	Adults with chronic kidney disease undergoing routine in-hospital non-gated chest CT; previous ASCVD excluded	DL-CACS	0; 1–100; 101–400; >400 AU	Death (including cardiac death), non-fatal myocardial infarction, revascularisation, or heart-failure/aggravated-angina rehospitalisation	4.6 years	48	Age; sex; smoking; CKD aetiology; BMI; blood pressure; antihypertensive and lipid-lowering drugs; lipid measures; eGFR; inflammatory, cardiac, mineral-metabolism, and albuminuria markers	Yes—antihypertensive and lipid-lowering drug use
Mets et al, 2013	NELSON derivation and external validation cohorts; descriptive prognostic evidence	3,648	Two-centre lung cancer screening cohort of current and former heavy smokers	Automated noncommercial CAC-volume quantification with manual correction	Continuous calcium volume; Agatston score evaluated as an alternative predictor	Three-year cardiovascular events from linked clinical outcome data	3 years	145 derivation; 120 validation cohort	Age; smoking status; pack-years; cardiovascular history; aortic calcium volume	No treatment covariate reported
Lessmann et al, 2019	NLST nested case-control cohort; descriptive cardiovascular-mortality evidence	5,718	Multicentre lung cancer screening cohort of heavy smokers	Fully automated in-house coronary and thoracic-aortic calcium scoring	0; 1–100; 101–1000; >1000 AU	Cardiovascular mortality, with cause-specific analyses	Median 6.5 years	Not reported by CAC category	Age; smoking; pack-years; BMI; race; education; marital status; cardiovascular history; diabetes; hypertension; CAC and thoracic-aortic calcium	Not assessed; preventive medication use was not recorded
Zhou et al, 2026	Multicentre mild-CAC/normal-perfusion cohort; descriptive prognostic evidence	4,552	SPECT/CT or PET/CT myocardial perfusion imaging; no prior coronary artery disease; AI CAC 1–99 AU	Automated vessel-specific CAC segmentation with proximal-lesion classification	CAC 1–99 AU; proximal CAC present versus absent	Study-defined MACE	Median 3.6 years	599	Total CAC; site; CAC density; age; sex; BMI; hypertension; diabetes; dyslipidaemia; family history; smoking; stress perfusion deficit	No treatment covariate reported

Report	Cohort	N	Setting	AI method	CAC categories	Outcome definition	Follow-up	Events	Adjustment covariates	Statin/treatment adjustment
Shanbhag et al, 2026	External Yale/Calgary CTAC cohort	5,969	Myocardial perfusion imaging; previous MI, PCI, or CABG excluded	Knowledge-adaptation deep-learning CAC scoring	0; 1–100; 101–400; >400 AU	Death or myocardial infarction	Median 813 days	450	Age; sex; hypertension; diabetes; dyslipidaemia; family history of CAD; smoking; stress total perfusion deficit; LVEF	No treatment covariate reported
Noureldine et al, 2026	Mayo Clinic; three sites	1,873	Immune checkpoint inhibitor treatment; prior coronary revascularisation excluded	AI-derived Agatston CAC	0; 1–99; 100–399; ≥400 AU	Myocardial infarction and all-cause mortality, assessed separately	1 year	MI 1.8%; all-cause mortality 50.7%	Not reported	Not reported
Yi et al, 2026	External test cohort	7,028	SPECT/CT or PET/CT myocardial perfusion imaging	CAC-only predictor from CT attenuation correction	Continuous CAC predictor	All-cause mortality; CAC-only AUC and C-index	Median 4.36 years	Not separately reported for CAC-only analyses	Not applicable to CAC-only discrimination metrics	Not applicable
Cheng et al, 2026	Development cohort; CAC-specific models	2,411	Chronic kidney disease; routine non-gated chest CT	Deep-learning CAC quantification	0; 1–100; 101–400; >400; doubling of CAC	All-cause and cardiovascular mortality, assessed separately	Median 4.1 years	387 all-cause deaths; 167 cardiovascular deaths	Age; sex; BMI; systolic blood pressure; smoking; diabetes; hypertension; coronary heart disease; stroke; eGFR; ln(UACR); HDL and LDL cholesterol; statins; ACEI/ARB/ARNI	Statin and ACEI/ARB/ARNI use included

ASCVD, atherosclerotic cardiovascular disease; AU=Agatston units; BMI, body-mass index; CAC, coronary artery calcium; CKD, chronic kidney disease; CT, computed tomography; CTAC, CT attenuation correction; CVD, cardiovascular disease; eGFR=estimated glomerular filtration rate; HDL, high-density lipoprotein; LDCT, low-dose CT; LDL, low-density lipoprotein; LVEF, left ventricular ejection fraction; MACE, major adverse cardiovascular events; MI, myocardial infarction; NR, not reported; PET, positron-emission tomography; SPECT, single-photon emission CT.

Supplementary Table 15. Prognostic pooled results

The table presents the primary study-defined MACE analyses.

CAC category comparison	k reports	k estimates	N	Pooled adjusted HR	95% CI	95% prediction interval	τ^2	I^2	Model
CAC 1–100 versus CAC 0	6	6	17,100	1.60	1.29–1.97	1.12–2.27	0.012	25.5%	Random effects on log HR scale; REML τ^2; Hartung–Knapp inference
CAC 101–400 versus CAC 0	5	5	11,758	2.40	1.49–3.89	0.94–6.15	0.084	71.0%	Random effects on log HR scale; REML τ^2; Hartung–Knapp inference

Supplementary Table 15 continued

CAC category comparison	k reports	k estimates	N	Pooled adjusted HR	95% CI	95% prediction interval	τ^2	I^2	Model
CAC >400 versus CAC 0	4	4	11,251	3.39	1.65–6.98	1.01–11.42	0.094	78.1%	Random effects on log HR scale; REML τ^2; Hartung–Knapp inference

CI, confidence interval; HR, hazard ratio; MACE, major adverse cardiovascular events; REML, restricted maximum likelihood.

Supplementary Table 16. GRADE certainty of evidence for prognostic associations between AI-derived CAC categories and major adverse cardiovascular events

Each row is a separate noncontextualised prognostic-factor evidence body. All contrasts use CAC 0 as the reference and adjusted hazard ratios for study-defined MACE.

Prognostic contrast	Evidence base	Pooled adjusted HR (95% CI)	95% prediction interval	Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Dose–response	Certainty
CAC 1–100 versus CAC 0	6 reports; 6 estimates; N = 17,100	1.60 (1.29–1.97)	1.12–2.27	Serious[a]	Not serious[b]	Not serious[c]	Not serious[e]	Not strongly suspected[f]	Supportive; no upgrade[g]	⊕⊕⊕◯ Moderate
CAC 101–400 versus CAC 0	5 reports; 5 estimates; N = 11,758	2.40 (1.49–3.89)	0.94–6.15	Serious[a]	Not serious[b]	Serious[d]	Not serious[e]	Not strongly suspected[f]	Supportive; no upgrade[g]	⊕⊕◯◯ Low
CAC >400 versus CAC 0	4 reports; 4 estimates; N = 11,251	3.39 (1.65–6.98)	1.01–11.42	Serious[a]	Not serious[b]	Serious[d]	Not serious[e]	Not strongly suspected[f]	Supportive; no upgrade[g]	⊕⊕◯◯ Low

a. Risk of bias was downgraded one level for each contrast. The estimate-level QUIPS profiles were CAC 1–100 versus CAC 0 (2 High and 4 Moderate); CAC 101–400 versus CAC 0 (2 High and 3 Moderate); CAC >400 versus CAC 0 (1 High and 3 Moderate). The principal concerns were residual confounding, participant selection or attrition, and possible influence of clinically visible CAC on clinician-driven MACE components. The evidence was not downgraded two levels because the concerns were not uniform, several adjusted cohorts had only moderate risk, and plausible bias directions were not consistently aligned.

b. Indirectness was not downgraded. The evidence directly addressed the association between AI-derived CAC categories on non-gated CT and subsequent study-defined MACE. Differences in MACE composition and clinical setting were considered when judging inconsistency rather than counted again as indirectness.

c. Inconsistency was not serious for CAC 1–100 versus CAC 0: I^2 was 25.5% and the 95% prediction interval (1.12–2.27) remained above 1.

d. Inconsistency was serious for the higher categories. For CAC 101–400 versus CAC 0, I^2 was 71.0% and the 95% prediction interval was 0.94–6.15; for CAC >400 versus CAC 0, I^2 was 78.1% and the prediction interval was 1.01–11.42. No additional downgrade was applied merely because the intervals were wide.

e. Imprecision was not downgraded. The pooled 95% confidence intervals for all three adjusted associations excluded HR=1 and supported at least a material increase in MACE risk. Prediction-interval dispersion was addressed under inconsistency and was not counted twice.

f. Publication bias was not strongly suspected. With only four to six reports per contrast, regression-based asymmetry tests would not be reliable; no specific missing report or selectively unavailable eligible MACE contrast was identified.

g. A graded association was present in the pooled estimates (HR 1.60, 2.40, and 3.39), and each of the four reports contributing all three harmonised categories showed a monotonic within-report gradient. This supported the prognostic interpretation but did not lead to formal upgrading because no prespecified dose–response model was fitted and residual confounding could contribute to the gradient.

GRADE approach. Certainty was assessed post hoc using GRADE guidance for prognostic factors. Each harmonised contrast—CAC 1–100, CAC 101–400, or CAC >400 versus CAC 0—was treated as a separate evidence body because the contributing reports and heterogeneity differed. Observational evidence addressing a prognostic-factor association began at high certainty.[131–132]

Risk-of-bias judgments for the harmonised prognostic evidence bodies were informed by the 15 category-based estimates contributing to the meta-analysis. Inconsistency considered adjusted hazard-ratio distributions, confidence-interval overlap, I^2, and prediction intervals. Indirectness was judged against AI-derived CAC category and study-defined MACE, imprecision against distinguishing no association from a material risk increase, and publication bias qualitatively because each contrast included four to six reports. The pooled and within-report gradients were supportive but did not justify upgrading because no prespecified dose–response model was fitted and residual confounding remained plausible.

AI, artificial intelligence; CAC, coronary artery calcium; CI, confidence interval; CT, computed tomography; GRADE, Grading of Recommendations Assessment, Development and Evaluation; HR, hazard ratio; MACE, major adverse cardiovascular events; QUIPS, Quality In Prognosis Studies.

Supplementary Section 8

Implementation evidence and certainty

Supplementary Table 17. Implementation evidence

The two randomised comparisons and their linked follow-up report are shown separately. Trial outcomes are reported at six months unless otherwise specified; the extension follows all notified participants without a concurrent control group.

Domain	Measure	Randomised notification trials		Uncontrolled extension
		Sandhu et al, 2023 (NOTIFY-1)	Dudum et al, 2026 (NOTIFY-PICTURE)	Bouladian et al, 2025 (NOTIFY-EXTEND)
Design and population	**Setting and design**	Single-centre randomised quality-improvement trial; retrospective screening of prior CTs; statin-naive adults without known ASCVD	Single-centre pragmatic prospective randomised trial; near-real-time CT screening; lipid-lowering-naive adults with or without ASCVD	Single-health-system follow-up of NOTIFY-1; pooled early- and delayed-notification recipients without a concurrent control group
	Participants, N	173 (86 vs 87)	202 (100 vs 102)	163 notified participants
Care pathway	**Intervention components**	DL-CAC screening; electronic-record exclusions; manual chart review; radiologist confirmation; actionable notification	FDA-cleared AI-CAC screening of routine non-contrast chest CT; eligibility review; near-real-time actionable notification	AI detection of incidental CAC; clinician/patient notification; preventive-treatment review; longitudinal electronic-record follow-up
	Communication	Clinician notified first; patient message 2 weeks later, with postal letter if unopened; CAC image and statin guidance provided	Clinician and patient notified; clinician received a CAC image and ACC/AHA statin guidance; patient workflow followed NOTIFY-1	Early and delayed NOTIFY-1 groups analysed from their respective notification dates
	Human confirmation	Radiologist confirmed CAC in 173/194 AI-positive scans (89.2%) before randomisation	No study-level confirmation step reported; CAC appeared in 97.0% of original radiology reports but in 4.0% of impressions	The parent NOTIFY-1 workflow included radiologist confirmation before randomisation
Preventive management	**Statin prescribing**	44/86 (51.2%) vs 6/87 (6.9%); RR 7.42 (95% CI 3.34–16.50)	44/100 (44.0%) vs 10/102 (9.8%); RR 4.49 (95% CI 2.39–8.42)	75/163 (46%) by 6 months; 86/163 (53%) by 12 months
	Absolute difference	+44.3 percentage points (reconstructed; CI not reported)	+34 percentage points (95% CI 23–46; trial reported)	Not estimable without a concurrent control group
	LDL cholesterol	Within-person change not reported; follow-up LDL among tested patients 97.2 vs 115.3 mg/dL (difference −18.1; P = 0.005)	Not reported; trial was not powered to detect LDL changes	Mean 115.6 to 90.6 mg/dL among 97 participants retested at 12 months (change −25.0 mg/dL)
	Adherence / persistence	Not assessed; persistence beyond 6 months not reported	Not assessed	Statin persistence 82/86 (95%) at 12 months and 75/86 (87%) at 18 months; any lipid-lowering therapy 78/86 (91%) at 18 months
Additional investigation	**Stress testing**	10/86 (11.6%) vs 2/87 (2.3%); P = 0.018	Included in cardiovascular-testing composite; not separately reported	3/163 (2%) before versus 19/163 (12%) during the 12 months after notification
	Gated CAC testing	3/86 (3.5%) vs 0/87 (0%); P = 0.121	Included in cardiovascular-testing composite; not separately reported	0/163 before versus 21/163 (13%) during the 12 months after notification
Follow-up and outcomes	**Health-service use**	Primary-care/endocrinology encounters 2.2 vs 1.4 per patient; new cardiology encounters 14/86 vs 4/87; CAD testing 13/86 vs 2/87	Primary-care visits 1.97 vs 1.42 per patient; cardiology visits 0.38 vs 0.22; cardiovascular testing 41/100 vs 28/102	Primary-care/endocrinology visits 2.2 to 2.7 per participant; cardiology visits 0.2 to 0.5
	Clinical-event assessment	No clinical-event endpoint; six-month process and risk-factor outcomes only	No clinical-event endpoint; six-month prescription and utilisation outcomes only	No clinical-outcome endpoint; follow-up extended to 18 months

Absolute difference is notification minus usual care. The NOTIFY-1 absolute statin-prescription difference was reconstructed from the arm proportions; the NOTIFY-PICTURE difference and confidence interval were trial reported. A fixed-effect inverse-variance synthesis of the two bundled pathways yielded RR 5.44 (95% CI 3.32–8.92); because only two heterogeneous workflows were available, this was treated as a secondary summary and not as the isolated effect of AI. NOTIFY-1 follow-up LDL values are not within-participant changes and were available for 50/86 and 29/87 participants. ACC/AHA, American College of Cardiology/American Heart Association; ASCVD, atherosclerotic cardiovascular disease; CAC, coronary artery calcium; CAD, coronary artery disease; CI, confidence interval; CT, computed tomography; DL-CAC, deep-learning coronary artery calcium; LDL, low-density lipoprotein cholesterol; RR, risk ratio.

Supplementary Table 18. GRADE certainty of evidence for AI-enabled notification and six-month statin prescribing

The evidence body addresses assignment to the complete clinician/patient notification pathway, not the isolated effect of the AI algorithm. Trial-level effects are shown; the secondary fixed-effect summary is reported with the implementation evidence in Supplementary Table 17.

Panel A. Trial-level effects

Study	N	Statin prescriptions	RR (95% CI)	Absolute difference
Sandhu et al, 2023 (NOTIFY-1)	173 (86 vs 87)	44/86 vs 6/87	7.42 (3.34–16.50)	+44.3 percentage points (CI not reported)
Dudum et al, 2026 (NOTIFY-PICTURE)	202 (100 vs 102)	44/100 vs 10/102	4.49 (2.39–8.42)	+34 percentage points (95% CI 23–46)

Panel B. Certainty profile

Outcome	Evidence base	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty
Statin prescription within 6 months	2 randomised trials; N = 375	Not serious[a]	Not serious[b]	Not serious[c]	Not serious[d]	Not strongly suspected[e]	⨁⨁⨁⨁ High

a. Risk of bias was not downgraded. RoB 2 was Low in 1 trial and Some concerns in 1 trial. The latter concerns involved incomplete public reporting of allocation concealment, outcome-extraction comparability, and the pre-unblinding analysis plan; complete intention-to-treat denominators, an objective prescription outcome, and no concrete material bias pathway made a body-level downgrade unwarranted.

b. Inconsistency was not downgraded. Both trials showed large effects in the same direction, with absolute increases of 34.0 and 44.3 percentage points and overlapping relative-effect confidence intervals.

c. Indirectness was not downgraded for the stated question of an AI-enabled bundled clinician/patient notification pathway. The rating does not isolate the AI component and does not apply to adherence, lipid change, or cardiovascular events.

d. Imprecision was not downgraded. Both trial confidence intervals excluded no effect and small effects; even the lower confidence limits supported a substantial increase in six-month statin prescribing.

e. Publication bias was not strongly suspected. Both trials were prospectively registered and reported the retained six-month prescription outcome; with two trials, statistical assessment of small-study effects was not appropriate.

GRADE approach. Certainty was assessed post hoc using intervention GRADE for six-month statin prescription after an AI-enabled bundled clinician/patient notification pathway versus usual care. Two individually randomised trials began at high certainty; the estimand was assignment to the complete pathway, not the isolated AI effect. Risk of bias used the prespecified result-level RoB 2 assessments.[133]

Because workflows and human confirmation differed, trial-specific relative and absolute effects were the principal presentation. Inconsistency considered direction, magnitude, and confidence-interval overlap; indirectness the bundled-notification question; imprecision each trial's compatible effects; and publication bias registration and outcome availability. Adherence, cardiovascular events, and an isolated AI effect were not assigned certainty ratings because they were not evaluated evidence bodies.

High certainty applies to six-month prescription orders in the two studied bundled-notification settings. Neither trial isolated the AI component or evaluated cardiovascular-event effects. AI, artificial intelligence; CI, confidence interval; GRADE, Grading of Recommendations Assessment, Development and Evaluation; RR, risk ratio; RoB 2, Risk of Bias 2.

References

1. Whiting PF, Tomlinson E, Rutjes AWS, et al. QUADAS-3: a revised tool for the quality assessment of diagnostic test accuracy studies. Ann Intern Med 2026; 179: 548–55.
2. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013; 158: 280–86.
3. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019; 366: l4898.
4. Chen Y, Wang Y, Wang Q, et al. Impact of scanning parameters on deep learning coronary artery calcium scoring in non-gated chest CT. *BMC Med Imaging* 2026; 26: 232.
5. Cho SG, Lee JE, Cho KH, et al. Coronary artery calcium measurement on attenuation correction computed tomography using artificial intelligence: correlation with coronary flow capacity and prognosis. *Eur J Nucl Med Mol Imaging* 2025; 52: 1050–59.
6. Dobrolinska MM, Lazarenko SV, van der Zant FM, et al. Performance of visual, manual, and automatic coronary calcium scoring of cardiac 13N-ammonia PET/low dose CT. *J Nucl Cardiol* 2023; 30: 239–50.

7. Dobrolinska MM, Jukema RA, van Velzen SGM, et al. The prognostic value of visual and automatic coronary calcium scoring from low-dose computed tomography-[15O]-water positron emission tomography. *Eur Heart J Cardiovasc Imaging* 2024; 25: 1186–96.
8. Eng D, Chute C, Khandwala N, et al. Automated coronary calcium scoring using deep learning with multicenter external validation. *NPJ Digit Med* 2021; 4: 88.
9. Gokmen MS, Haque MN, Leung SW, et al. A framework for cross-domain generalization in coronary artery calcium scoring across gated and non-gated computed tomography. *arXiv* 2026; published online Feb 25. https://doi.org/10.48550/arXiv.2602.21935.
10. Hagopian R, Strebel T, Bernatz S, et al. AI opportunistic coronary calcium screening at Veterans Affairs hospitals. *NEJM AI* 2025; 2. https://doi.org/10.1056/AIoa2400937.
11. Homayounieh F, Yan P, Digumarthy SR, et al. Prediction of coronary calcification and stenosis: role of radiomics from low-dose CT. *Acad Radiol* 2021; 28: 972–79.
12. Jung JH, Lee JE, Lee HS, Yang DH, Lee JG. Diffusion-based image translation model from low-dose chest CT to calcium scoring CT with random point sampling. *Comput Biol Med* 2025; 194: 110506.
13. Kang HW, Ahn WJ, Jeong JH, et al. Evaluation of fully automated commercial software for Agatston calcium scoring on non-ECG-gated low-dose chest CT with different slice thickness. *Eur Radiol* 2023; 33: 1973–81.
14. Kim C, Hong S, Choi H, et al. Impact of deep learning-based image conversion on fully automated coronary artery calcium scoring using thin-slice, sharp-kernel, non-gated, low-dose chest CT scans: a multi-center study. *Korean J Radiol* 2025; 26: 759–70.
15. Kim S, Park EA, Ahn C, et al. Performance of fully automated deep-learning-based coronary artery calcium scoring in ECG-gated calcium CT and non-gated low-dose chest CT. *Eur Radiol* 2025; 35: 7084–95.
16. Li J, Li A, Fang J, et al. Combating coronary calcium scoring bias for non-gated CT by semantic learning on gated CT. *Proc IEEE/CVF Int Conf Comput Vis Workshops* 2023: 2575–83.
17. Lin Y, Lin G, Peng MT, Kuo CT, Wan YL, Cherng WJ. The role of artificial intelligence in coronary calcium scoring in standard cardiac computed tomography and chest computed tomography with different reconstruction kernels. *J Thorac Imaging* 2024; 39: 111–18.
18. Ludwig KD, Hatt CR, Keith L, et al. AI-based coronary artery calcification on non-contrast CT: performance across calcium scoring, lung cancer screening, and liver transplant candidate cohorts. *medRxiv* 2026; published online May 15. https://doi.org/10.64898/2026.05.12.26352904.
19. Marjanski T, Chmielecki M, Klein-Awerjanow K, et al. Can artificial intelligence help us in the evaluation of coronary artery calcification scores by acting as a prognosticator in patients that are operated on due to non-small cell lung cancer? A pivotal study. *J Clin Med* 2024; 13: 6579.
20. McKinney SE, Mercaldo SF, Chin JK, et al. Evaluation of an artificial intelligence model for opportunistic Agatston scoring on non-gated chest computed tomography. *Sci Rep* 2025; 15: 39535.
21. Morf C, Sartoretti T, Gennari AG, et al. Diagnostic value of fully automated artificial intelligence powered coronary artery calcium scoring from 18F-FDG PET/CT. *Diagnostics (Basel)* 2022; 12: 1876.
22. Oh HS, Kim TH, Kim JW, et al. Feasibility and limitations of deep learning–based coronary calcium scoring in PET-CT: a comparison with coronary calcium score CT. *Eur Radiol* 2024; 34: 4077–88.
23. Pieszko K, Shanbhag A, Killekar A, et al. Calcium scoring in low-dose ungated chest CT scans using convolutional long-short term memory networks. Proc SPIE 2022; 12032: 120323A.
24. Pieszko K, Shanbhag A, Killekar A, et al. Deep learning of coronary calcium scores from PET/CT attenuation maps accurately predicts adverse cardiovascular events. *JACC Cardiovasc Imaging* 2023; 16: 675–87.
25. Sartoretti E, Gennari AG, Maurer A, et al. Opportunistic deep learning powered calcium scoring in oncologic patients with very high coronary artery calcium (≥ 1000) undergoing 18F-FDG PET/CT. *Sci Rep* 2022; 12: 19191.
26. van Assen M, Martin SS, Varga-Szemes A, et al. Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: a validation study. *Eur J Radiol* 2021; 134: 109428.
27. van Velzen SGM, Dobrolinska MM, Knaapen P, et al. Automated cardiovascular risk categorization through AI-driven coronary calcium quantification in cardiac PET acquired attenuation correction CT. *J Nucl Cardiol* 2023; 30: 955–69.
28. Xu J, Liu J, Guo N, et al. Performance of artificial intelligence-based coronary artery calcium scoring in non-gated chest CT. *Eur J Radiol* 2021; 145: 110034.
29. Xu C, Guo H, Xu M, et al. 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; 12: 2684–95.
30. Xu Y, Yu Y, Ding X, et al. Deep learning-based coronary calcium score derived from non-gated chest CT and major adverse cardiovascular events in patients with type 2 diabetes mellitus. *Cardiovasc Diabetol* 2025; 24: 389.
31. Yin K, Chen W, Qin G, et al. Performance assessment of an artificial intelligence-based coronary artery calcium scoring algorithm in non-gated chest CT scans of different slice thickness. *Quant Imaging Med Surg* 2024; 14: 5708–20.
32. Yu J, Qian L, Sun W, et al. Automated total and vessel-specific coronary artery calcium (CAC) quantification on chest CT: direct comparison with CAC scoring on non-contrast cardiac CT. *BMC Med Imaging* 2022; 22: 177.
33. Zhang J, Liu K, You C, Gong J. Vessel-specific reliability of artificial intelligence-based coronary artery calcium scoring on non-ECG-gated chest CT: a comparative study with ECG-gated cardiac CT. *Clin Imaging* 2025; 125: 110575.
34. Anifowose JO, Li Z, Agarwal G, et al. Automated opportunistic cardiovascular risk assessment in non-small cell lung cancer patients on routine chest CT using an optimised nnU-net framework. *BMC Med Imaging* 2026; 26: 179.
35. Balbi M, Sabia F, Ledda RE, et al. Automated coronary artery calcium and quantitative emphysema in lung cancer screening. *J Thorac Imaging* 2023; 38: W52–W63.
36. Barrett O, Shanbhag A, Zaid R, et al. Incremental diagnostic value of artificial intelligence-derived coronary artery calcium in 18F-flurpiridaz positron emission tomography myocardial perfusion imaging. *J Nucl Cardiol* 2025; 54: 102532.
37. Cano-Espinosa C, González G, Washko GR, Cazorla M, San José Estépar R. Automated Agatston score computation in non-ECG gated CT scans using deep learning. Proc SPIE 2018; 10574: 105742K.
38. Cano-Espinosa C, González G, Washko GR, et al. On the relevance of the loss function in the Agatston score regression from non-ECG gated CT scans. In: Stoyanov D, Taylor Z, Sarikaya D, et al, eds. *Image analysis for moving organ, breast, and thoracic images*. Cham: Springer, 2018: 326–34.

39. Cano-Espinosa C, Gonzalez G, Washko GR, Cazorla M, Estepar RSJ. Biomarker localization from deep learning regression networks. *IEEE Trans Med Imaging* 2020; 39: 2121–32.
40. Chamberlin J, Kocher MR, Waltz J, et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. *BMC Med* 2021; 19: 55.
41. Ching K, Carlson B, Robinson E, et al. A manually annotated, open-source dataset for coronary, aortic, and valvular calcium segmentation in non-gated chest CT. *Int J Imaging Syst Technol* 2026; 36: e70311.
42. Choi JH, Cha MJ, Cho I, et al. Validation of deep learning-based fully automated coronary artery calcium scoring using non-ECG-gated chest CT in patients with cancer. *Front Oncol* 2022; 12: 989250.
43. 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: 2127–38.
44. Gernaat SAM, Išgum I, de Vos BD, et al. Automatic coronary artery calcium scoring on radiotherapy planning CT scans of breast cancer patients: reproducibility and association with traditional cardiovascular risk factors. *PLoS One* 2016; 11: e0167925.
45. Gernaat SA, van Velzen SG, Koh V, et al. Automatic quantification of calcifications in the coronary arteries and thoracic aorta on radiotherapy planning CT scans of Western and Asian breast cancer patients. *Radiother Oncol* 2018; 127: 487–92.
46. Gonzalez G, Washko GR, Estepar RSJ. Automated Agatston score computation in a large dataset of non ECG-gated chest computed tomography. *IEEE Int Symp Biomed Imaging* 2016: 53–57.
47. Hamelink I, Nie ZZ, Severijn TT, et al. Validation of artificial intelligence software for automatic calcium scoring in cardiac and chest computed tomography. *Eur J Radiol* 2025; 191: 112323.
48. Huo Y, Terry JG, Wang J, et al. Coronary calcium detection using 3D attention identical dual deep network based on weakly supervised learning. Proc SPIE 2019; 10949: 1094917.
49. Imuro SES, Khan SS, Ahmed EN, et al. Classifying coronary artery calcification risk using radiomics from non-gated CT scans. *IEEE Int Symp Biomed Imaging* 2025: 1–4.
50. Išgum I, Prokop M, Jacobs PC, et al. Automatic coronary calcium scoring in low-dose non-ECG-synchronized thoracic CT scans. *Proc SPIE* 2010; 7624: 76240M.
51. Išgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging* 2012; 31: 2322–34.
52. Išgum I, de Vos BD, Wolterink JM, et al. Automatic detection of cardiovascular risk in CT attenuation correction maps in Rb-82 PET/CTs. *Proc SPIE* 2016; 9784: 978405.
53. Lessmann N, Išgum I, Setio AAA, et al. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest CT. *Proc SPIE* 2016; 9785: 978511.
54. Lessmann N, van Ginneken B, Zreik M, et al. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans Med Imaging* 2018; 37: 615–25.
55. Miller RJ, Pieszko K, Shanbhag A, et al. Deep learning coronary artery calcium scores from SPECT/CT attenuation maps improve prediction of major adverse cardiac events. *J Nucl Med* 2023; 64: 652–58.
56. Mineo E, Assuncao-Jr AN, Grego da Silva CF, et al. Automated coronary artery calcium scoring using deep learning: validation across diverse chest CT protocols. *Acad Radiol* 2025; 32: 7177–86.
57. Nguyen A, Liu J, Mathai TS, Grayson PC, Summers RM. Automated plaque detection and Agatston score estimation on non-contrast CT scans: a multicenter study. Proc SPIE 2024; 12927: 1292710.
58. Rathore S, Gautam A, Raghav P, et al. Fully automated coronary artery calcium score and risk categorization from chest CT using deep learning and multiorgan segmentation: a validation study from the National Lung Screening Trial (NLST). *Int J Cardiol Heart Vasc* 2025; 56: 101593.
59. Shadmi R, Mazo V, Bregman-Amitai O, Elnekave E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. *IEEE Int Symp Biomed Imaging* 2018: 24–28.
60. Shen H, Lian Y, Yin J, et al. Cardiovascular risk stratification by automatic coronary artery calcium scoring on pretreatment chest computed tomography in diffuse large B-cell lymphoma receiving anthracycline-based chemotherapy: a multicenter study. *Circ Cardiovasc Imaging* 2023; 16: e014829.
61. Suh YJ, Kim C, Lee JG, et al. Fully automatic coronary calcium scoring in non-ECG-gated low-dose chest CT: comparison with ECG-gated cardiac CT. Eur Radiol 2023; 33: 1254–65.
62. Takx RAP, de Jong PA, Leiner T, et al. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. *PLoS One* 2014; 9: e91239.
63. Tang H, Moradi M, Prasanna P, Wang H, Syeda-Mahmood T. An algorithm for fully automatic detection of calcium in chest CT imaging. *IEEE Int Symp Biomed Imaging* 2017: 265–69.
64. van Velzen SGM, Lessmann N, Velthuis BK, et al. Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols. *Radiology* 2020; 295: 66–79.
65. van Velzen SGM, de Vos BD, Noothout JMH, Verkooijen HM, Viergever MA, Išgum I. Generative models for reproducible coronary calcium scoring. *J Med Imaging (Bellingham)* 2022; 9: 052406.
66. Wang M, Zhang J, Zhu R, Gu J, Fan L, Chen Q. Quantifying the impact of slice thickness on cardiovascular risk stratification in lung cancer screening: a multi-center "RESCUE" study. *Quant Imaging Med Surg* 2026; 16: 568.
67. Zeleznik R, Foldyna B, Eslami P, et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nat Commun* 2021; 12: 715.
68. Alnasser TN, Hokmabadi A, Sharkey MJ, et al. A novel fully automated deep learning model for coronary artery calcification detection on computed tomography. *Diagnostics (Basel)* 2026; 16: 646.
69. Bonatti M, Proner B, Vingiani V, Valletta R, Saba L. The role of AI assistance in the evaluation of unenhanced chest CT scans in an emergency setting. *Pol J Radiol* 2026; 91: 132–39.
70. Chamberlin J, Smith C, Schoepf U, et al. A deep convolutional neural network ensemble for composite identification of pulmonary nodules and incidental findings on routine PET/CT. *Clin Radiol* 2023; 78: e368–e76.
71. Kerndt CC, Chopra R, Weber P, et al. Using artificial intelligence to semi-quantitate coronary calcium as an 'incidentaloma' on non-gated, non-contrast CT scans, a single-center descriptive study in West Michigan. *Spartan Med Res J* 2023; 8: 89132.
72. Park C, Lee BC, Jeong WG, Park WJ, Jin GY, Kim YH. Coronary artery calcification on low-dose lung cancer screening CT in South Korea: visual and artificial intelligence–based assessment and association with cardiovascular events. *AJR Am J Roentgenol* 2024; 222: e2430852.
73. Shear B, Graby J, Murphy D, et al. Assessing the diagnostic accuracy and prognostic utility of artificial intelligence detection and grading of coronary artery calcification on nongated computed tomography (CT) thorax. *Clin Radiol* 2025; 85: 106909.
74. Smith C, Nance S, Chamberlin JH, et al. Application of an artificial intelligence ensemble for detection of important secondary findings on lung ventilation and perfusion SPECT-CT. *Clin Imaging* 2023; 100: 24–29.
75. Wu KC, Hsieh TC, Hsu ZK, et al. Application of deep learning in automated localization and interpretation of coronary artery calcification in oncological PET/CT scans. *Int J Cardiovasc Imaging* 2025; 41: 453–65.
76. Xie Y, Liu S, Miller A, et al. Coronary artery calcification identification and labeling in low-dose chest CT images. *Proc SPIE* 2017; 10134: 101340L.

77. Atkins KM, Weiss J, Zeleznik R, et al. Elevated coronary artery calcium quantified by a validated deep learning model from lung cancer radiotherapy planning scans predicts mortality. *JCO Clin Cancer Inform* 2022; 6: e2100095.
78. Barda N, Dagan N, Stemmer A, et al. Improving cardiovascular disease prediction using automated coronary artery calcium scoring from existing chest CTs. *J Digit Imaging* 2022; 35: 962–69.
79. de Vos BD, de Jong PA, Wolterink JM, et al. Automatic machine learning based prediction of cardiovascular events in lung cancer screening data. *Proc SPIE* 2015; 9414: 94140D.
80. de Vos BD, Lessmann N, de Jong PA, Išgum I. Deep learning–quantified calcium scores for automatic cardiovascular mortality prediction at lung screening low-dose CT. *Radiol Cardiothorac Imaging* 2021; 3: e190219.
81. Dekker M, Waissi F, Bank IE, et al. The prognostic value of automated coronary calcium derived by a deep learning approach on non-ECG gated CT images from 82Rb-PET/CT myocardial perfusion imaging. *Int J Cardiol* 2021; 329: 9–15.
82. Kim EJ, Cho SW, Yang JH, Jeong WG. Artificial intelligence-assisted longitudinal assessment of coronary artery calcification in the Korean lung cancer screening CT program. *Clin Imaging* 2025; 125: 110558.
83. Kim K, Chung SY, Oh C, et al. Automated coronary artery calcium scoring in patients with breast cancer to assess the risk of heart disease following adjuvant radiation therapy. *Breast* 2022; 65: 77–83.
84. Kim NY, Kim YH, Lee JE, Suh YJ. Opportunistic assessment of coronary artery calcium volume and density from non-electrocardiogram-gated chest CT using artificial intelligence: prognostic implications in a screening cohort. *Korean J Radiol* 2026; 27: 532–42.
85. Kim S, Kim S, Cha MJ, et al. Association between automated coronary artery calcium from routine chest computed tomography scans and cardiovascular risk in patients with colorectal or gastric cancer. *Circ Cardiovasc Qual Outcomes* 2025; 18: e011656.
86. Lee JE, Kim NY, Kim YH, et al. Long-term prognostic implications of thoracic aortic calcification on CT using artificial intelligence–based quantification in a screening population: a two-center study. *AJR Am J Roentgenol* 2025; 224: e2532697.
87. Lessmann N, de Jong PA, Celeng C, et al. Sex differences in coronary artery and thoracic aorta calcification and their association with cardiovascular mortality in heavy smokers. *JACC Cardiovasc Imaging* 2019; 12: 1808–17.
88. Mets OM, Vliegenthart R, Gondrie MJ, et al. Lung cancer screening CT-based prediction of cardiovascular events. *JACC Cardiovasc Imaging* 2013; 6: 899–907.
89. Sato H, Sakakura K, Jinnouchi H, et al. The impact of the coronary artery calcium score on the clinical outcomes in patients with acute myocardial infarction. *J Clin Med* 2024; 13: 7136.
90. Stemmer A, Shadmi R, Bregman-Amitai O, et al. Using machine learning algorithms to review computed tomography scans and assess risk for cardiovascular disease: retrospective analysis from the National Lung Screening Trial (NLST). *PLoS One* 2020; 15: e0236021.
91. Takx RA, 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: 50–57.
92. Takx RAP, Vliegenthart R, Hoesein FAAM, et al. Pulmonary function and CT biomarkers as risk factors for cardiovascular events in male lung cancer screening participants: the NELSON study. *Eur Radiol* 2015; 25: 65–71.
93. Weber BN, Biery DW, Petranovic M, et al. Prevalence and prognostic value of incidentally detected coronary artery calcium using artificial intelligence among individuals with immune-mediated inflammatory diseases. *JACC Cardiovasc Imaging* 2026; 19: 64–75.
94. Williams MC, Shanbhag AD, Zhou J, et al. Automated vessel-specific coronary artery calcification quantification with deep learning in a large multi-centre registry. *Eur Heart J Cardiovasc Imaging* 2024; 25: 976–85.
95. Xing Z, Pan H, Ding H, et al. Incremental value of non-gated chest CT coronary artery calcium score in predicting major adverse cardiovascular events by GRACE score after percutaneous coronary intervention in patients with acute coronary syndrome. *Clin Cardiol* 2026; 49: e70242.
96. Yang K, Li M, Wang J, et al. Association between deep learning-based coronary artery calcium score on non-gated chest CT and progression of chronic kidney disease: a retrospective observational cohort study. *BMC Med Imaging* 2026; 26: 291.
97. Zhou J, Miller RJH, Shanbhag A, et al. Hidden risk in normal myocardial perfusion scans: AI-detected proximal coronary calcium on CT attenuation maps improves prognosis. *Eur J Nucl Med Mol Imaging* 2026; published online July 28. https://doi.org/10.1007/s00259-026-08084-x.
98. Bouladian SG, Ngo S, Mastrodicasa D, et al. Statin therapy persistence following opportunistic screening for coronary artery calcium on nongated chest CTs (NOTIFY-EXTEND). *JACC Cardiovasc Imaging* 2025; 18: 841–43.
99. Chernina VY, Belyaev MG, Silin AY, et al. Analysis of the diagnostic and economic impact of the combined artificial intelligence algorithm for analysis of 10 pathological findings on chest computed tomography. *Digit Diagn* 2023; 4: 105–32.
100. Dudum R, Jain SS, Mastrodicasa D, et al. Effects of real-time notification of AI-detected incidental coronary artery calcium on statin prescription: the NOTIFY-PICTURE trial. *Circulation* 2026; 153: 210–12.
101. Sandhu AT, Rodriguez F, Ngo S, et al. Incidental coronary artery calcium: opportunistic screening of previous nongated chest computed tomography scans to improve statin rates (NOTIFY-1 project). *Circulation* 2023; 147: 703–14.
102. Bunk S, Bruins Slot T, Bennink E, et al. Repeatability of AI-quantified incidental findings on lung cancer screening CT scans in the NELSON trial. *Eur Radiol* 2026; published online June 17. https://doi.org/10.1007/s00330-026-12669-3.
103. Hamelink I, van Tuinen M, Kwee TC, van Ooijen PMA, Vliegenthart R. Repeatability of AI-based, automatic measurement of vertebral and cardiovascular imaging biomarkers in low-dose chest CT: the ImaLife cohort. *Eur Radiol* 2025; 35: 3833–41.
104. Ledda RE, Milanese G, Balbi M, et al. Coronary calcium score and emphysema extent on different CT radiation dose protocols in lung cancer screening. *Eur Radiol* 2025; 35: 3781–87.
105. Yacoub B, Kabakus IM, Schoepf UJ, et al. Performance of an artificial intelligence-based platform against clinical radiology reports for the evaluation of noncontrast chest CT. *Acad Radiol* 2022; 29 (suppl 2): S108–17.
106. Liu Y, Xin W, Wang L, et al. Validation of Artificial Intelligence-based Non-gated Chest CT for Coronary Artery Calcium Scoring Across Multiple CT Scanners. J Imaging Inform Med 2026. https://doi.org/10.1007/s10278-026-02288-1.

107. Shanbhag A, Miller RJH, Killekar A, et al. Cross-Domain Knowledge Transfer from Expert-Annotated Gated CT via Synthetic Ungated CT Improves Coronary Artery Calcium Scoring on CT Attenuation Correction Scans. medRxiv 2026; published online July 6 (preprint). https://doi.org/10.64898/2026.07.02.26356002.
108. Ochs K, Ensle F, Happe J, Jungblut L, Frauenfelder T, Kroschke J. Agreement and workflow efficiency of AI-based coronary artery calcification quantification in lung cancer screening: Comparison with semi-automated and visual assessment. Eur J Radiol Open 2026; 17: 100799. https://doi.org/10.1016/j.ejro.2026.100799.
109. Noureldine HA, Scialo V, Farina JM, et al. Assessing the prognostic value of artificial intelligence for opportunistic prediction of coronary artery calcium score on routine non-gated computed tomography in patients treated with immune checkpoint inhibitors. Digit Health 2026; 12: 20552076261480207. https://doi.org/10.1177/20552076261480207.
110. Liu YK, Wu FY, Jiang CP, et al. Artificial intelligence-based chest computed tomography assessment of bone, vascular, and epicardial adipose changes in type 2 diabetes. World J Diabetes 2026; 17: 121422. https://doi.org/10.4239/wjd.121422.
111. Yi J, Patel KK, Miller RJH, et al. Beyond BMI: an interpretable integrated body composition index from low-dose chest CT for all-cause mortality risk stratification: a multicentre study. medRxiv 2026; published online Aug 10 (preprint). https://doi.org/10.64898/2026.08.05.26359437.
112. Cheng D, Liu C, Fei Y, et al. Deep learning-derived cardiac phenotyping from routine non-gated chest CT for mortality prediction and risk stratification in chronic kidney disease: a development and external validation cohort study. SSRN 2026 (preprint). https://doi.org/10.2139/ssrn.7040888.
113. Hussain M, Saeed HY, Farid MI, Ismoilovich IO, Kumar S, Talha HU. Deep Learning for Automated Coronary Artery Calcium Scoring in Chest CTs. Int J Drug Deliv Technol 2026; 16 (suppl 76): 1535–41. https://doi.org/10.25258/ijddt.16.76s.140.
114. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PMM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58: 982–90.
115. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001; 20: 2865–84.
116. Doebler P. mada: meta-analysis of diagnostic accuracy. R package version 0.5.12. 2025.
117. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019; 22: 153–60.
118. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010; 36: 1–48.
119. R Core Team. R: a language and environment for statistical computing, version 4.5.3. Vienna: R Foundation for Statistical Computing, 2026.
120. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. J Stat Softw 2017; 76: 1–32.
121. Gabry J, Češnovar R, Johnson A. cmdstanr: R interface to CmdStan. R package version 0.9.0. 2025.
122. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005; 58: 882–93.
123. Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. J Clin Epidemiol 2020; 122: 129–41.
124. Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. J Clin Epidemiol 2020; 122: 142–52.
125. Mustafa RA, El Mikati IK, Murad MH, et al. GRADE guidance 37: rating imprecision in a body of evidence on test accuracy. J Clin Epidemiol 2024; 165: 111189.
126. Bipat S, Zwinderman AH, Bossuyt PMM, Stoker J. Multivariate random-effects approach: for meta-analysis of cancer staging studies. Acad Radiol 2007; 14: 974–84.
127. Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. BMC Med Res Methodol 2009; 9: 73.
128. Met R, Bipat S, Legemate DA, Reekers JA, Koelemay MJW. Diagnostic performance of computed tomography angiography in peripheral arterial disease: a systematic review and meta-analysis. JAMA 2009; 301: 415–24.
129. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995; 57: 289–300.
130. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003; 22: 2693–710.
131. Foroutan F, Guyatt G, Zuk V, et al. GRADE guidelines 28: use of GRADE for the assessment of evidence about prognostic factors. J Clin Epidemiol 2020; 121: 62–70.
132. Murad MH, Verbeek J, Schwingshackl L, et al. GRADE guidance 38: updated guidance for rating up certainty of evidence due to a dose–response gradient. J Clin Epidemiol 2023; 164: 45–53.
133. Schünemann HJ, Higgins JPT, Vist GE, et al. Chapter 14: Completing summary-of-findings tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane handbook for systematic reviews of interventions, version 6.5.1. Cochrane, 2024.

Supplementary Section 9. Reporting checklists

PRISMA 2020 checklist

Manuscript: AI-based opportunistic coronary artery calcium assessment on non-gated chest CT: a systematic review and meta-analysis

This file also contains PRISMA 2020 for Abstracts, PRISMA-DTA, PRISMA-DTA for Abstracts, and PRISMA-S.

Section and topic	Item	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title.
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract; PRISMA 2020 for Abstracts checklist in this file.
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction.
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction.
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods—Eligibility criteria; Methods—Reference standards and outcomes.
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods—Search strategy; Supplementary Note 1.
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Supplementary Note 1.
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods—Study selection and data extraction.
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods—Study selection and data extraction.
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods—Reference standards and outcomes; Methods—Statistics and reproducibility.
Data items	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Methods—Study selection and data extraction; Supplementary Table 5A; Supplementary Data 1, sheets S1 and S12.
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods—Risk of bias and applicability; Fig. 2; Supplementary Fig. 1; Supplementary Tables 3A–4.
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods—Reference standards and outcomes; Methods—Statistics and reproducibility.
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods—Reference standards and outcomes; Methods—Statistics and reproducibility.
Synthesis methods	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods—Study selection and data extraction; Methods—Statistics and reproducibility (Diagnostic accuracy; Continuous agreement and sources of discrepancy).
Synthesis methods	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Figs. 1–8; Tables 1–3; Supplementary Figs. 1–8; Supplementary Tables 2–18.
Synthesis methods	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods—Statistics and reproducibility.
Synthesis methods	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods—Statistics and reproducibility (Subgroup analyses and meta-regression); Supplementary Table 13; Supplementary Figs. 7 and 8; Supplementary Data 1, sheet S6.
Synthesis methods	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Methods—Statistics and reproducibility; Supplementary Table 6; Supplementary Tables 8 and 9; Supplementary Figs. 4 and 5.
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Methods—Statistics and reproducibility (Diagnostic accuracy); Supplementary Table 6; Supplementary Fig. 3; publication-bias judgments in Supplementary Tables 7, 16, and 18.
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods—Certainty of evidence; Supplementary Tables 7, 16, and 18.
RESULTS			

PRISMA 2020 checklist continued

Section and topic	Item	Checklist item	Location where item is reported
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Results—Study selection and characteristics; Fig. 1.
Study selection	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Supplementary Section 10, Supplementary Table 19 (reports excluded after full-text assessment, with reasons).
Study characteristics	17	Cite each included study and present its characteristics.	Results—Study selection and characteristics; Table 1; Supplementary Tables 5A and 5B; Supplementary Data 1, sheets S1 and S12.
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Results—Study selection and characteristics; Fig. 2; Supplementary Fig. 1; Supplementary Tables 3A–4.
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Figs. 3–6 and 8; Supplementary Figs. 2–8; Supplementary Tables 6–18; Supplementary Data 1, sheets S1–S6.
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results; Supplementary Tables 3A–5B; Supplementary Tables 7, 16, and 18.
Results of syntheses	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results; Table 2; Figs. 3–8; Supplementary Figs. 2–8; Supplementary Tables 6–18.
Results of syntheses	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results—Sensitivity analyses, publication bias, and meta-regression; Supplementary Table 13; Supplementary Figs. 7 and 8; Supplementary Data 1, sheet S6.
Results of syntheses	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Results—Sensitivity analyses, publication bias, and meta-regression; Supplementary Table 6; Supplementary Tables 8 and 9; Supplementary Figs. 4 and 5.
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Results—Sensitivity analyses, publication bias, and meta-regression; Supplementary Table 6 (threshold-specific Deeks tests); Supplementary Fig. 3; Supplementary Tables 7, 16, and 18 (publication-bias judgments).
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Results; Table 2; Supplementary Tables 7, 16, and 18.
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion.
Discussion	23b	Discuss any limitations of the evidence included in the review.	Discussion.
Discussion	23c	Discuss any limitations of the review processes used.	Discussion.
Discussion	23d	Discuss implications of the results for practice, policy, and future research.	Discussion.
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Methods—Protocol and registration (INPLASY202660006).
Registration and protocol	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Methods—Protocol and registration; Data availability (https://doi.org/10.37766/inplasy2026.6.0006).
Registration and protocol	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Supplementary Table 1.
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Acknowledgements.
Competing interests	26	Declare any competing interests of review authors.	Competing interests.
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data availability; Code availability.

Checklist source: Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71. CC BY 4.0.

PRISMA 2020 for Abstracts checklist

Section and topic	Item	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title.
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Abstract (first two sentences).
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Abstract (AI-based CAC assessment on routine non-gated chest CT); full criteria in Methods—Eligibility criteria.
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Not reported in the abstract owing to the 150-word limit; Methods—Search strategy (PubMed, Embase, Web of Science Core Collection, and CENTRAL; final search 2 October 2026).
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Not reported in the abstract owing to the 150-word limit; Methods—Risk of bias and applicability.
Synthesis of results	6	Specify the methods used to present and synthesise results.	Abstract (meta-analyses by reference standard and threshold); Methods—Statistics and reproducibility.
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Abstract (110 included reports; 59 contributed to meta-analyses).
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Abstract (pooled sensitivity and specificity, classification probabilities, hazard ratios, and trial effects).
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Not reported in the abstract owing to the 150-word limit; Discussion.
Interpretation	10	Provide a general interpretation of the results and important implications.	Abstract (final sentence).
OTHER			
Funding	11	Specify the primary source of funding for the review.	Not reported in the abstract owing to the 150-word limit; Acknowledgements.
Registration	12	Provide the register name and registration number.	Not reported in the abstract owing to the 150-word limit; Methods—Protocol and registration (INPLASY202660006).

Checklist source: Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71. CC BY 4.0.

PRISMA-DTA checklist

Section and topic	Item	Checklist item	Location where item is reported
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	Title (systematic review and meta-analysis); the diagnostic test accuracy component is identified in the Abstract and Introduction.
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	Abstract; PRISMA-DTA for Abstracts checklist in this file.
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Introduction.
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	Introduction.
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	Introduction (review objectives); Methods—Eligibility criteria; Methods—Reference standards and outcomes.
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Methods—Protocol and registration.
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Methods—Eligibility criteria; Methods—Reference standards and outcomes.
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Methods—Search strategy; Supplementary Note 1.
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Supplementary Note 1.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Methods—Study selection and data extraction.
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Methods—Study selection and data extraction.
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	Methods—Reference standards and outcomes; Methods—Risk of bias and applicability; Supplementary Table 2.
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	Methods—Risk of bias and applicability; Fig. 2; Supplementary Fig. 1; Supplementary Tables 3A–4.
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	Methods—Reference standards and outcomes.
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	Methods—Statistics and reproducibility (Diagnostic accuracy).
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	Methods—Statistics and reproducibility (Diagnostic accuracy).
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Methods—Statistics and reproducibility (Diagnostic accuracy; Category agreement; Subgroup analyses and meta-regression).
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	Results—Study selection and characteristics; Fig. 1; Supplementary Tables 19 and 20.
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	Results—Study selection and characteristics; Table 1; Supplementary Tables 5A and 5B; Supplementary Data 1, sheets S1 and S12.
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	Results—Study selection and characteristics; Fig. 2; Supplementary Fig. 1; Supplementary Tables 3A–4.
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2×2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	Figs. 3–6 (TP, FN, TN, and FP with individual estimates and 95% CIs); Supplementary Data 1, sheet S3; Supplementary Fig. 2 (SROC curves).
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	Results—Pooled sensitivity and specificity by reference standard and threshold; Table 2; Figs. 3–6; Supplementary Table 6; Supplementary Fig. 2.

PRISMA-DTA checklist continued

Section and topic	Item	Checklist item	Location where item is reported
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	Results; Supplementary Table 6; Supplementary Tables 8–10 and Supplementary Figs. 4 and 5; Supplementary Table 13 and Supplementary Figs. 7 and 8; Supplementary Data 1.
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	Discussion; Table 2 (GRADE certainty); Supplementary Table 7.
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	Discussion.
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	Discussion.
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	Acknowledgements.

Checklist source: McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. JAMA. 2018;319:388–396. doi:10.1001/jama.2017.19163.

PRISMA-DTA for Abstracts checklist

Section and topic	Item	Checklist item	Location where item is reported
TITLE and PURPOSE			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	Title (systematic review and meta-analysis); the diagnostic test accuracy component is identified in the Abstract and Introduction.
Objectives	2	Indicate the research question, including components such as participants, index test, and target conditions.	Abstract (first sentence: participants, index test, and target condition).
METHODS			
Eligibility criteria	3	Include study characteristics used as criteria for eligibility.	Abstract (AI-based CAC assessment on routine non-gated chest CT); full criteria in Methods—Eligibility criteria.
Information sources	4	List the key databases searched and the search dates.	Not reported in the abstract owing to the 150-word limit; Methods—Search strategy.
Risk of bias & applicability	5	Indicate the methods of assessing risk of bias and applicability.	Not reported in the abstract owing to the 150-word limit; Methods—Risk of bias and applicability.
Synthesis of results	A1	Indicate the methods for the data synthesis.	Abstract (meta-analyses); Methods—Statistics and reproducibility.
RESULTS			
Included studies	6	Indicate the number and type of included studies and the participants and relevant characteristics of the studies (including the reference standard).	Abstract (110 reports; 59 in meta-analyses; paired ECG-gated CT reference standard); further characteristics in Results and Table 1.
Synthesis of results	7	Include the results for the analysis of diagnostic accuracy, preferably indicating the number of studies and participants. Describe test accuracy including variability; if meta-analysis was done, include summary results and confidence intervals.	Abstract (sensitivity and specificity at CAC >0 and ≥100; specificity above 97% at ≥300 and ≥400); complete estimates, 95% CIs, and per-analysis counts in Table 2.
DISCUSSION			
Strengths and limitations	9	Provide a brief summary of the strengths and limitations of the evidence	Not reported in the abstract owing to the 150-word limit; Discussion.
Interpretation	10	Provide a general interpretation of the results and the important implications.	Abstract (final sentence).
OTHER			
Funding	11	Indicate the primary source of funding for the review.	Not reported in the abstract owing to the 150-word limit; Acknowledgements.
Registration	12	Provide the registration number and the registry name	Not reported in the abstract owing to the 150-word limit; Methods—Protocol and registration (INPLASY202660006).

Checklist source: Cohen JF, Deeks JJ, Hooft L, et al. PRISMA-DTA for Abstracts. BMJ. 2021;372:n265. doi:10.1136/bmj.n265.

PRISMA-S checklist

Section and topic	Item	Checklist item	Location where item is reported
INFORMATION SOURCES AND METHODS			
Database name	1	Name each individual database searched, stating the platform for each.	Methods—Search strategy; Supplementary Note 1.
Multi-database searching	2	If databases were searched simultaneously on a single platform, state the name of the platform, listing all of the databases searched.	Not applicable—databases were searched separately rather than simultaneously through one multi-database platform.
Study registries	3	List any study registries searched.	Not used as a separate source.
Online resources and browsing	4	Describe any online or print source purposefully searched or browsed (e.g., tables of contents, print conference proceedings, web sites), and how this was done.	Not used as a separate source.
Citation searching	5	Indicate whether cited references or citing references were examined, and describe any methods used for locating cited/citing references (e.g., browsing reference lists, using a citation index, setting up email alerts for references citing included studies).	Not used as a separate source.
Contacts	6	Indicate whether additional studies or data were sought by contacting authors, experts, manufacturers, or others.	Methods—Search strategy (study authors were not contacted).
Other methods	7	Describe any additional information sources or search methods used.	Methods—Search strategy (supplementary implementation and preventive-care search; Search B in Supplementary Note 1).
SEARCH STRATEGIES			
Full search strategies	8	Include the search strategies for each database and information source, copied and pasted exactly as run.	Supplementary Note 1 (PubMed, Web of Science, CENTRAL, and Embase).
Limits and restrictions	9	Specify that no limits were used, or describe any limits or restrictions applied to a search (e.g., date or time period, language, study design) and provide justification for their use.	Methods—Search strategy (no date or language restrictions); Supplementary Note 1.
Search filters	10	Indicate whether published search filters were used (as originally designed or modified), and if so, cite the filter(s) used.	No published search filter was used. The complete syntax, including the PubMed animal-only exclusion, is reported in Supplementary Note 1.
Prior work	11	Indicate when search strategies from other literature reviews were adapted or reused for a substantive part or all of the search, citing the previous review(s).	Not applicable—no previous review search strategy was reported as adapted or reused.
Updates	12	Report the methods used to update the search(es) (e.g., rerunning searches, email alerts).	Methods—Search strategy (all databases searched from inception to 2 October 2026).
Dates of searches	13	For each search strategy, provide the date when the last search occurred.	Methods—Search strategy (2 October 2026); Supplementary Note 1.
PEER REVIEW			
Peer review	14	Describe any search peer review process.	Not applicable—no formal PRESS peer review was undertaken.
MANAGING RECORDS			
Total Records	15	Document the total number of records identified from each database and other information sources.	Fig. 1: Embase 2,090; PubMed 2,084; Web of Science 1,803; CENTRAL 565; total 6,542 records.
Deduplication	16	Describe the processes and any software used to deduplicate records from multiple database searches and other information sources.	Methods—Search strategy; Fig. 1 (2,467 duplicates removed).

PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews

Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB, PRISMA-S Group.

Checklist source: Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10:39. doi:10.1186/s13643-020-01542-z.

Supplementary Section 10. Reports excluded after full-text assessment and reports not retrieved

Supplementary Table 19. Reports excluded after full-text assessment (n = 224)

Report	Author and year	Title	Reason for exclusion
FTX001	Nozaki, Y et al (2025)	A novel artificial intelligence-powered framework for predicting coronary artery calcium score from non-gated computed tomography	Conference or meeting abstract without an eligible full report.
FTX002	Mansoor, M et al (2026)	A VISION TRANSFORMER OUTPERFORMS A CONVOLUTIONAL NEURAL NETWORK FOR AUTOMATED CORONARY ARTERY CALCIUM SCORING ON NON-GATED CHEST CT: AN ANALYSIS OF THE NATIONAL LUNG SCREENING TRIAL	Conference or meeting abstract without an eligible full report.
FTX003	Aviv, Y et al (2025)	AI-driven opportunistic screening for coronary artery calcification on non-gated CT: 1-year follow-up results from Nanox AI	Conference or meeting abstract without an eligible full report.
FTX004	Kumar, K et al (2025)	AI-POWERED APPLICATIONS IMPROVE THE UNDER-RECOGNITION OF CORONARY ARTERY DISEASE BY IDENTIFYING AND SCORING CORONARY ARTERY CALCIUM ON NON-CARDIAC CHEST CT	Conference or meeting abstract without an eligible full report.
FTX005	Graby, J et al (2023)	An artifical intelligence tool improves coronary artery calcification reporting on non-gated, non-cardiac chest ct	Conference or meeting abstract without an eligible full report.
FTX006	Shears, B et al (2022)	ASSESSING THE USE OF AN AI TOOL FOR GRADING CORONARY ARTERY CALCIFICATION ON NON-GATED, NON-CARDIAC CHEST CT	Conference or meeting abstract without an eligible full report.
FTX007	Lu, MT et al (2019)	Association of Automated Deep Learning Coronary Artery Calcium Score on ECG-Gated and Non-ECG Gated CT With Future Cardiac Events: Insights From ROMICAT-II and NLST	Conference or meeting abstract without an eligible full report.
FTX008	Mansoor, M (2025)	AUTOMATED CORONARY ARTERY CALCIUM SCORING USING CONVOLUTIONAL NEURAL NETWORKS: ENHANCING CARDIOVASCULAR RISK ASSESSMENT IN CHEST CT SCANS	Conference or meeting abstract without an eligible full report.
FTX009	Follmer, B et al (2023)	Automated segment-level coronary artery calcium scoring on noncontrast CT using deep learning	Conference or meeting abstract without an eligible full report.
FTX010	Satoh, H et al (2010)	Computer-aided diagnosis workstation and teleradiology network system for chest diagnosis using the web medical image conference system with a new information security solution	Conference or meeting abstract without an eligible full report.
FTX011	Han, D et al (2024)	Dedicated cardiovascular screening in lung cancer screening: preliminary results from the European 4-IN-THE-LUNG-RUN trial	Conference or meeting abstract without an eligible full report.
FTX012	Pieszko, K et al (2022)	DEEP LEARNING FROM UNGATED LOW-DOSE CT ATTENUATION CORRECTION MAPS PREDICTS MAJOR ADVERSE CARDIAC EVENTS SIMILAR TO STANDARD CORONARY CALCIUM SCORES	Conference or meeting abstract without an eligible full report.
FTX013	Fuhrman, JD et al (2019)	Deep learning in the task of detecting coronary artery calcifications on low-dose thoracic CTs	Conference or meeting abstract without an eligible full report.
FTX014	Alnasser, TN et al (2025)	Diagnostic accuracy of a deep learning model for coronary calcification detection on unenhanced CT in pulmonary hypertension patients	Conference or meeting abstract without an eligible full report.
FTX015	Qu, JT et al (2025)	DOES GRADING OF CORONARY ARTERY CALCIFICATION IN LDCT LUNG CANCER SCREENING AFFECT MANAGEMENT OF PATIENTS? THE REAL-WORLD EXPERIENCE OF A LARGE INTEGRATED HEALTH CARE SYSTEM	Conference or meeting abstract without an eligible full report.
FTX016	Kiewsky, J et al (2025)	Evaluation of a Deep Learning-driven Incidental Detection of Coronary Artery Calcium on Non-gated CT Scans	Conference or meeting abstract without an eligible full report.
FTX017	Millar, K et al (2019)	INCIDENTAL CORONARY ARTERY CALCIFICATION ON NON-CARDIAC COMPUTED TOMOGRAPHY: A THREE YEAR FOLLOW-UP STUDY	Conference or meeting abstract without an eligible full report.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX018	Kolli, KK et al (2017)	Machine Learning Algorithms and Coronary Calcium Score for Risk Stratification in Asymptomatic Healthy Population	Conference or meeting abstract without an eligible full report.
FTX019	Aviv, Y et al (2024)	Opportunistic screening of coronary artery calcification on non-gated conventional CT scans using artificial intelligence	Conference or meeting abstract without an eligible full report.
FTX020	Peng, A et al (2023)	OPPORTUNISTIC SCREENING OF INCIDENTAL CORONARY ARTERY CALCIUM WITH DEEP-LEARNING ALGORITHM ON NON-ECG GATED CHEST CT IMAGING AND ASSOCIATION WITH CARDIOVASCULAR EVENTS AND MORTALITY	Conference or meeting abstract without an eligible full report.
FTX021	Rahiminejad, M et al (2020)	REPORTING OF CORONARY ARTERY CALCIFICATION ON NON-GATED PRE-SURGICAL CT THORAX	Conference or meeting abstract without an eligible full report.
FTX022	Gernaat, SAM et al (2016)	Reproducibility of automated coronary artery calcification scoring on radiotherapy treatment planning computed tomography scans of breast cancer patients	Conference or meeting abstract without an eligible full report.
FTX023	Chopra, R et al (2022)	USING ARTIFICIAL INTELLIGENCE TO ACCURATELY QUANTIFY CORONARY ARTERY CALCIFICATIONS ON NON-GATED NON-CONTRAST CT SCANS TO DIAGNOSE INDIVIDUALS WITH HIGH RISK CORONARY ARTERY DISEASE	Conference or meeting abstract without an eligible full report.
FTX024	Bednarski, B et al (2022)	USING ARTIFICIAL INTELLIGENCE TO EVALUATE ADDED VALUE OF CORONARY ARTERY CALCIUM SCORING FROM STANDARD AND LOW-DOSE UNGATED CT IN MYOCARDIAL PERFUSION PET/CT IMAGING MORTALITY RISK ASSESSMENT	Conference or meeting abstract without an eligible full report.
FTX025	Ukai, Y et al (1998)	A coronary calcification diagnosis system based on helical CT images	Duplicate or superseded report.
FTX026	Dudum R et al (2025)	Effects of Real-Time Notification of AI-Derived Incidental Coronary Artery Calcium on Statin Initiation: the NOTIFY-PICTURE Trial	Meeting abstract superseded by the retrieved and included 2026 Circulation full report (DOI 10.1161/CIRCULATIONAHA.125.078155).
FTX027	Wang Z et al (2024)	Enhanced CT imaging artificial neural network coronary artery calcification score assisted diagnosis.	Duplicate or superseded report.
FTX028	Michalowska AM et al (2024)	Holistic AI analysis of hybrid cardiac perfusion images for mortality prediction.	Duplicate or superseded report.
FTX029	Sandhu, AT et al (2022)	Incidental Coronary Artery Calcium: Opportunistic Screening of Prior Non-Gated Chest Cts to Improve Statin Rates	Duplicate or superseded report.
FTX030	Barrett O et al (2025)	Incremental diagnostic value of AI-derived coronary artery calcium in (18)F-flurpiridaz PET Myocardial Perfusion Imaging.	Duplicate or superseded report.
FTX031	Manava P et al (2020)	[Artificial intelligence-based algorithms : Decision-making support for computed tomography of the chest].	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX032	Ukai, Y et al (1998)	A coronary calcification diagnosis system based on helical CT images	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX033	Ukai, Y et al (1998)	A coronary calcification diagnosis system based on helical CT images	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX034	Zhang, S et al (2023)	A ROBUST AGATSTON SCORE FOR CORONARY ARTERY CALCIUM SCORING FROM NON-ECG-GATED CT WITH DIFFERENT RECONSTRUCTION KERNELS	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX035	Marcinkiewicz AM et al (2024)	AI for Multistructure Incidental Findings and Mortality Prediction at Chest CT in Lung Cancer Screening.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX036	Palm, V et al (2022)	AI-Supported Comprehensive Detection and Quantification of Biomarkers of Subclinical Widespread Diseases at Chest CT for Preventive Medicine	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX037	Haseltine JM et al (2023)	Association of cardiac calcium burden with overall survival after radiotherapy for non-small cell lung cancer.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX038	Peng, AW et al (2023)	Association of Coronary Artery Calcium Detected by Routine Ungated CT Imaging With Cardiovascular Outcomes	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX039	Sabia, F et al (2025)	Automated Measurement of Coronary Artery Calcifications and Routine Perioperative Blood Tests Predict Survival in Resected Stage I Lung Cancer	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX040	Ukai, Y et al (1999)	CAD system for coronary calcifications based on helical CT images	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX041	Fuhrman, JD et al (2020)	Cascade of U-Nets in the detection and classification of coronary artery calcium in thoracic low-dose CT	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX042	Palmisano A et al (2023)	Chest CT opportunistic biomarkers for phenotyping high-risk COVID-19 patients: a retrospective multicentre study.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX043	Jirapatnakul A. et al (2024)	Comparison of the extent and progression of respiratory and cardiovascular disease in World Trade Center responders to lung screening participants	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX044	Ukai, Y et al (1999)	Computer assisted diagnosis system for coronary calcification based on helical CT image	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX045	Jaubert, O et al (2025)	CONFORMAL CORONARY CALCIFICATION VOLUME ESTIMATION WITH CONDITIONAL COVERAGE VIA HISTOGRAM CLUSTERING	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX046	Wenning C et al (2021)	Coronary artery calcium burden, carotid atherosclerotic plaque burden, and myocardial blood flow in patients with end-stage renal disease: A non-invasive imaging study combining PET/CT and 3D ultrasound.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX047	Takx, RAP et al (2017)	Coronary artery calcium in breast cancer survivors after radiation therapy	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX048	Takeshita, Y et al (2021)	Coronary artery calcium score may be a novel predictor of COVID-19 prognosis: a retrospective study	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX049	Upadhyayula S.K. (2024)	Deep Learning based CT-scan Coronary Artery Segmentation and Calcium Scoring	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX050	Fuhrman, JD et al (2019)	Detection and Classification of Coronary Artery Calcifications in Low Dose Thoracic CT Using Deep Learning	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX051	Chen V. et al (2026)	End-to-end opportunistic cardiovascular risk reporting from non-contrast computed tomography using foundation-model segmentation and survival modeling: a proof-of-concept pipeline	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX052	Mura, R et al (2026)	Exploring the prevalence and clinical impact of carotid plaque burden by Doppler ultrasound in lung cancer screening participants with limited coronary artery calcification	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX053	Sabia, F et al (2023)	[Prognostic report outside the eligible outcome scope]	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX054	Gal, R et al (2021)	Identification of Risk of Cardiovascular Disease by Automatic Quantification of Coronary Artery Calcifications on Radiotherapy Planning CT Scans in Patients With Breast Cancer	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX055	Ruggirello M et al (2023)	Long-term outcomes of lung cancer screening in males and females.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX056	Ngamdu KS et al (2023)	Long-term statin therapy is associated with severe coronary artery calcification.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX057	Sidorenkov, G et al (2023)	Multi-source data approach for personalized outcome prediction in lung cancer screening: update from the NELSON trial	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX058	van der Aalst, C et al (2023)	P1.14-04 European Lung Cancer Screening Implementation: 4-IN-THE-LUNG-RUN trial	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX059	Vasilev Y.A. et al (2023)	Population-Based Study of Coronary Artery Calcification Using the Automated Analysis of Radiology Reports in Moscow	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX060	Miller, RJH et al (2024)	Predicting mortality from AI cardiac volumes mass and coronary calcium on chest computed tomography	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX061	Sabia F. et al (2022)	Predictive value of automated coronary calcium scoring in lung cancer screening with low dose computed tomography	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX062	Nardocci, C et al (2026)	Prognostic value of deep learning-based coronary artery calcium score and quantitative pneumonia burden in patients hospitalized with COVID-19	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX063	Yang T et al (2024)	The association of cardiovascular disease risk with coronary artery calcification and thoracic aortic dilation: a study in idiopathic inflammatory myopathies and systemic lupus erythematosus.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX064	Lassau N et al (2021)	Three artificial intelligence data challenges based on CT and ultrasound.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX065	Wang WC et al (2025)	TQGDNet: Coronary artery calcium deposit detection on computed tomography.	No eligible diagnostic, agreement, prognostic, or implementation outcome was reported.
FTX066	Feher, A et al (2024)	Artificial Intelligence Predicts Hospitalization for Acute Heart Failure Exacerbation in Patients Undergoing Myocardial Perfusion Imaging	Review, guideline, commentary, or other report without eligible original data.
FTX067	Ning, Y et al (2021)	CAC-EMVT: Efficient Coronary Artery Calcium Segmentation with Multi-scale Vision Transformers	Review, guideline, commentary, or other report without eligible original data.
FTX068	Xiao H et al (2024)	Coronary artery calcium scoring assessment in ultra-low-dose chest computed tomography.	Review, guideline, commentary, or other report without eligible original data.
FTX069	Budoff MJ et al (2007)	Coronary calcium: does it still play a role in the age of CT angiography.	Review, guideline, commentary, or other report without eligible original data.
FTX070	Otero-Pla E et al (2025)	Mapping clinical and imaging factors that might predict cardiac events in breast cancer patients.	Review, guideline, commentary, or other report without eligible original data.
FTX071	Miller, RJH et al (2024)	The Updated Registry of Fast Myocardial Perfusion Imaging with Next-Generation SPECT (REFINE SPECT 2.0)	Review, guideline, commentary, or other report without eligible original data.
FTX072	Opitz A et al (2005)	[New CT procedures in cardiovascular diagnosis].	No eligible AI-based or fully automated CAC assessment was reported.
FTX073	Naghavi, M et al (2025)	AI-enabled opportunistic measurement of liver steatosis in coronary artery calcium scans predicts cardiovascular events and mortality: an AI-CVD study	No eligible AI-based or fully automated CAC assessment was reported.
FTX074	Shih YT et al (2026)	Application of the Novel Two-Compartmental Model to Quantify Coronary Artery Calcium: A Pilot Study.	No eligible AI-based or fully automated CAC assessment was reported.
FTX075	Lei, MH et al (2024)	Assessing mortality risk in Type 2 Diabetes patients with prolonged ASCVD risk factors: the inclusive Poh-Ai predictive scoring system with CAC Score integration	No eligible AI-based or fully automated CAC assessment was reported.
FTX076	Zheng X et al (2022)	Association between coronary artery calcium score and in-stent restenosis after drug-eluting stent implantation.	No eligible AI-based or fully automated CAC assessment was reported.
FTX077	Kazibwe R et al (2023)	Association of silent myocardial infarction on electrocardiogram and coronary artery calcium: The Multi-Ethnic Study of Atherosclerosis.	No eligible AI-based or fully automated CAC assessment was reported.
FTX078	Xie, YT et al (2014)	Automated coronary artery calcification detection on low-dose chest CT images	No eligible AI-based or fully automated CAC assessment was reported.
FTX079	Douthit NT et al (2021)	Clinical Impact of Reporting Coronary Artery Calcium Scores of Non-Gated Chest Computed Tomography on Statin Management.	No eligible AI-based or fully automated CAC assessment was reported.
FTX080	Brodov Y et al (2015)	Combined Quantitative Assessment of Myocardial Perfusion and Coronary Artery Calcium Score by Hybrid 82Rb PET/CT Improves Detection of Coronary Artery Disease.	No eligible AI-based or fully automated CAC assessment was reported.
FTX081	Shu YL et al (2025)	Comparison of chest computed tomography in Turbo FLASH mode with conventional mode for coronary artery disease screening: radiation dose, image quality, and calcium scoring performance.	No eligible AI-based or fully automated CAC assessment was reported.
FTX082	Feher, A et al (2024)	Comparison of the prognostic value between quantification and visual estimation of coronary calcification from attenuation CT in patients undergoing SPECT myocardial perfusion imaging	No eligible AI-based or fully automated CAC assessment was reported.
FTX083	de Mattos, JN et al (2022)	Computed tomography on lung cancer screening is useful for adjuvant comorbidity diagnosis in developing countries	No eligible AI-based or fully automated CAC assessment was reported.
FTX084	Satoh, H et al (2008)	Computer-aided diagnosis workstation and network system for chest diagnosis based on multislice CT images	No eligible AI-based or fully automated CAC assessment was reported.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX085	Rasmussen, T et al (2015)	Coronary artery calcification detected in lung cancer screening predicts cardiovascular death	No eligible AI-based or fully automated CAC assessment was reported.
FTX086	Watase, M et al (2025)	Coronary Artery Calcification on Chest Computed Tomography as a Predictor of Cardiovascular Adverse Events in Patients With COVID-19 ― A Multicenter Retrospective Study in Japan ―	No eligible AI-based or fully automated CAC assessment was reported.
FTX087	Wang K et al (2022)	Coronary Artery Calcifications and Cardiac Risk After Radiation Therapy for Stage III Lung Cancer.	No eligible AI-based or fully automated CAC assessment was reported.
FTX088	Gal R et al (2021)	Coronary artery calcifications on breast cancer radiotherapy planning CT scans and cardiovascular risk: What do patients want to know?	No eligible AI-based or fully automated CAC assessment was reported.
FTX089	Jacobs PC et al (2012)	[Prognostic report outside the eligible outcome scope]	No eligible AI-based or fully automated CAC assessment was reported.
FTX090	Zorzi, A et al (2021)	Coronary artery calcium on standard chest computed tomography predicts cardiovascular events after liver transplantation	No eligible AI-based or fully automated CAC assessment was reported.
FTX091	Cook, DN et al (2026)	Coronary Artery Calcium Score Predicts Cardiovascular and Transplant-Related Outcomes in Lung Transplant Recipients	No eligible AI-based or fully automated CAC assessment was reported.
FTX092	Shin, JM et al (2020)	Coronary artery calcium scoring on non-gated, non-contrast chest computed tomography (CT) using wide-detector, high-pitch and fast gantry rotation: comparison with dedicated calcium scoring CT	No eligible AI-based or fully automated CAC assessment was reported.
FTX093	Groen RA et al (2023)	Coronary calcium score in COVID-19 survivors: Association with cardiac injury and cardiac function after 6 weeks.	No eligible AI-based or fully automated CAC assessment was reported.
FTX094	Šprem J et al (2018)	Coronary calcium scoring with partial volume correction in anthropomorphic thorax phantom and screening chest CT images.	No eligible AI-based or fully automated CAC assessment was reported.
FTX095	Puliti D et al (2019)	Decreased cardiovascular mortality in the ITALUNG lung cancer screening trial: Analysis of underlying factors.	No eligible AI-based or fully automated CAC assessment was reported.
FTX096	Pelandré GL et al (2018)	Detection of coronary artery calcification with nontriggered computed tomography of the chest.	No eligible AI-based or fully automated CAC assessment was reported.
FTX097	Hou KY et al (2026)	Developing a calibration method to utilize low-dose chest CT for assessment of coronary artery calcification score.	No eligible AI-based or fully automated CAC assessment was reported.
FTX098	Wang, XM et al (2025)	Feasibility analysis of non-electrocardiogram-triggered chest low-dose computed tomography using a kV-independent reconstruction algorithm for predicting cardiovascular disease risk in patients receiving maintenance hemodialysis	No eligible AI-based or fully automated CAC assessment was reported.
FTX099	Regan, EA et al (2019)	Identifying Smoking-Related Disease on Lung Cancer Screening CT Scans: Increasing the Value	No eligible AI-based or fully automated CAC assessment was reported.
FTX100	Gong B et al (2024)	Immune checkpoint inhibitors in cancer: the increased risk of atherosclerotic cardiovascular disease events and progression of coronary artery calcium.	No eligible AI-based or fully automated CAC assessment was reported.
FTX101	Sprem, J et al (2018)	Impact of automatically detected motion artifacts on coronary calcium scoring in chest computed tomography	No eligible AI-based or fully automated CAC assessment was reported.
FTX102	Rodriguez-Granillo, GA et al (2017)	Impact on mortality of coronary and non-coronary cardiovascular findings in non- gated thoracic CT by malignancy status	No eligible AI-based or fully automated CAC assessment was reported.
FTX103	Lee, JE et al (2026)	Incidental cardiovascular calcifications detected on screening CT: prognostic impact over long-term follow-up	No eligible AI-based or fully automated CAC assessment was reported.
FTX104	Shaikh PA et al (2024)	Incidental Coronary Arterial Calcification for Cardiovascular Risk Assessment in Men With Prostate Cancer Undergoing PET/CT Imaging.	No eligible AI-based or fully automated CAC assessment was reported.
FTX105	Millar, K et al (2022)	INCIDENTAL CORONARY ARTERY CALCIFICATION ON NON-CARDIAC COMPUTED TOMOGRAPHY: A 5 YEAR FOLLOW UP STUDY	No eligible AI-based or fully automated CAC assessment was reported.
FTX106	Han, D et al (2015)	[Prognostic report outside the eligible outcome scope]	No eligible AI-based or fully automated CAC assessment was reported.
FTX107	Ladak, AM et al (2026)	Large Language Model Agentic Workflows for Automated Opportunistic Screening of Incidental Coronary Artery Calcium on Chest CT Reports	No eligible AI-based or fully automated CAC assessment was reported.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX108	Chattopadhyay, R et al (2023)	Lipid lowering therapy in patients with incidental coronary artery calcification: a single-centre observational study	No eligible AI-based or fully automated CAC assessment was reported.
FTX109	Stephan JT et al (2023)	Low-Dose Computed Tomography Scan Features Are Associated With Annual Risk of Hospitalization.	No eligible AI-based or fully automated CAC assessment was reported.
FTX110	Lu MT et al (2016)	Lung Cancer Screening Eligibility in the Community: Cardiovascular Risk Factors, Coronary Artery Calcification, and Cardiovascular Events.	No eligible AI-based or fully automated CAC assessment was reported.
FTX111	Kermanian, Riley et al (2021)	MANAGEMENT OF CORONARY ARTERY CALCIFICATION IN PATIENTS ENROLLED IN A LOW-DOSE COMPUTERIZED TOMOGRAPHY LUNG CANCER SCREENING PROGRAM	No eligible AI-based or fully automated CAC assessment was reported.
FTX112	Qu, JT et al (2024)	MEDICAL INTERVENTIONS AND CARDIOLOGY REFERRAL PATTERNS AFTER REPORTING OF INCIDENTAL CORONARY ARTERY CALCIFICATION DURING LUNG CANCER SCREENING: REAL-WORLD EXPERIENCE FROM A LARGE INTEGRATED HEALTHCARE SYSTEM	No eligible AI-based or fully automated CAC assessment was reported.
FTX113	Zhang, Q et al (2026)	Opportunistic coronary calcium progression on routine chest CT improves cardiovascular risk stratification in patients with inflammatory bowel disease: a multicenter study	No eligible AI-based or fully automated CAC assessment was reported.
FTX114	Shim, J et al (2026)	Optimization of Coronary Artery Calcium Scoring Accuracy in Low-Dose Chest Computed Tomography Using Fast Non-Local Means Denoising with High-Pitch Acquisition and Tin Filtration	No eligible AI-based or fully automated CAC assessment was reported.
FTX115	Choi H et al (2023)	Performance of 1-mm non-gated low-dose chest computed tomography using deep learning-based noise reduction for coronary artery calcium scoring.	No eligible AI-based or fully automated CAC assessment was reported.
FTX116	Qu, JT et al (2025)	PHARMACOLOGIC INTERVENTIONS AFTER REPORTING OF INCIDENTAL CORONARY ARTERY CALCIFICATION DURING LUNG CANCER SCREENING: REAL-WORLD EXPERIENCE FROM A LARGE HEALTH CARE SYSTEM	No eligible AI-based or fully automated CAC assessment was reported.
FTX117	Kroschke J et al (2026)	Photon-Counting Chest CT at Radiography-Comparable Dose Levels: Impact on Opportunistic Visual and Semiautomated Coronary Calcium Quantification.	No eligible AI-based or fully automated CAC assessment was reported.
FTX118	Broadbent R et al (2024)	Pilot study of lung cancer screening for survivors of Hodgkin lymphoma.	No eligible AI-based or fully automated CAC assessment was reported.
FTX119	Schwartz AW et al (2025)	Predictive value of tibial and coronary artery calcification scores for cardiac and lower extremity events.	No eligible AI-based or fully automated CAC assessment was reported.
FTX120	Tisi S et al (2023)	Prevalence and clinical characteristics of non-malignant CT detected incidental findings in the SUMMIT lung cancer screening cohort.	No eligible AI-based or fully automated CAC assessment was reported.
FTX121	Zahergivar A et al (2024)	Prognostic value of coronary artery calcium scoring in patients with non-small cell lung cancer using initial staging computed tomography.	No eligible AI-based or fully automated CAC assessment was reported.
FTX122	Wakaki, T et al (2025)	Prognostic value of coronary calcification detected via non-electrocardiogram-gated computed tomography in patients with cardiovascular disease: A retrospective cohort study	No eligible AI-based or fully automated CAC assessment was reported.
FTX123	Willemink, MJ et al (2015)	Prognostic value of heart valve calcifications for cardiovascular events in a lung cancer screening population	No eligible AI-based or fully automated CAC assessment was reported.
FTX124	Nappi, C et al (2022)	Quantification of Coronary Artery Atherosclerotic Burden and Muscle Mass: Exploratory Comparison of Two Freely Available Software Programs	No eligible AI-based or fully automated CAC assessment was reported.
FTX125	Pieszko K et al (2022)	Reproducibility of quantitative coronary calcium scoring from PET/CT attenuation maps: comparison to ECG-gated CT scans.	No eligible AI-based or fully automated CAC assessment was reported.
FTX126	Field, J et al (2019)	S01.03 Session I: 2019 Status of the International Maturity of CT Trial Outcomes and Their Implications	No eligible AI-based or fully automated CAC assessment was reported.
FTX127	Jungblut, L et al (2023)	Swiss Pilot Low-Dose CT Lung Cancer Screening Study: First Baseline Screening Results	No eligible AI-based or fully automated CAC assessment was reported.
FTX128	Aydeniz E et al (2024)	The association between coronary artery calcification and vectorcardiography in mechanically ventilated COVID-19 patients: the Maastricht Intensive Care COVID cohort.	No eligible AI-based or fully automated CAC assessment was reported.
FTX129	Celeng C et al (2020)	The Association Between Marital Status, Coronary Computed Tomography Imaging Biomarkers, and Mortality in a Lung Cancer Screening Population.	No eligible AI-based or fully automated CAC assessment was reported.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX130	Qu, JT et al (2025)	THE CLINICAL MANAGEMENT OF PATIENTS WITH CORONARY ARTERY CALCIFICATIONS DETECTED BY LDCT LUNG CANCER SCREENING: THE SIGNIFICANCE OF FINDINGS IN THE REAL WORLD	No eligible AI-based or fully automated CAC assessment was reported.
FTX131	Kvist, TV et al (2017)	The DanCavas Pilot Study of Multifaceted Screening for Subclinical Cardiovascular Disease in Men and Women Aged 65-74 Years	No eligible AI-based or fully automated CAC assessment was reported.
FTX132	Zhang, L et al (2025)	The role of coronary artery calcium scoring in the prediction of coronary artery disease based on non-contrast non-cardiac chest CT scans in airline pilots	No eligible AI-based or fully automated CAC assessment was reported.
FTX133	Antonicelli A et al (2024)	The Smokers Health Multiple ACtions (SMAC-1) Trial: Study Design and Results of the Baseline Round.	No eligible AI-based or fully automated CAC assessment was reported.
FTX134	Diederichsen, ACP et al (2024)	User-defined outcomes of the Danish cardiovascular screening (DANCAVAS) trial: a post hoc analyses of a population-based, randomised controlled trial	No eligible AI-based or fully automated CAC assessment was reported.
FTX135	Pletcher, MJ et al (2014)	Using the Coronary Artery Calcium Score to Guide Statin Therapy A Cost-Effectiveness Analysis	No eligible AI-based or fully automated CAC assessment was reported.
FTX136	Araúo, ADC et al (2024)	A Cascade Approach for Automatic Segmentation of Coronary Arteries Calcification in Computed Tomography Images Using Deep Learning	Population or CT imaging context was outside the eligibility criteria.
FTX137	Wu, J et al (2012)	A Fully Automated Multi-modal Computer Aided Diagnosis Approach to Coronary Calcium Scoring of MSCT Images	Population or CT imaging context was outside the eligibility criteria.
FTX138	Išgum, I et al (2004)	A pattern recognition approach to automated coronary calcium scoring	Population or CT imaging context was outside the eligibility criteria.
FTX139	Hoori, A et al (2021)	A Sequential Deep Learning Approach for Automated Detection and Attribution of Coronary Artery Calcification in Non-Contrast CT	Population or CT imaging context was outside the eligibility criteria.
FTX140	Kurkure U et al (2010)	A supervised classification-based method for coronary calcium detection in non-contrast CT.	Population or CT imaging context was outside the eligibility criteria.
FTX141	Naghavi, M et al (2025)	Abstract 4371006: Identification of High Risk Cases in Coronary Artery Calcium (CAC) Scans based on CAC Score and AI-driven Cardiometabolic Biomarkers: An AI-CVD Study within the Multi-Ethnic Study of Atherosclerosis (MESA)	Population or CT imaging context was outside the eligibility criteria.
FTX142	Bengtsen, K et al (2024)	AI-based algorithm for assessing coronary artery calcium score on contrast-enhanced cardiac computed tomography scans	Population or CT imaging context was outside the eligibility criteria.
FTX143	Naghavi, M et al (2024)	[Prognostic report outside the eligible outcome scope]	Population or CT imaging context was outside the eligibility criteria.
FTX144	Dehmeshki J et al (2004)	An accurate and reproducible scheme for quantification of coronary artery calcification in CT scans.	Population or CT imaging context was outside the eligibility criteria.
FTX145	Liang CJ et al (2012)	An alternative method for quantifying coronary artery calcification: the multi-ethnic study of atherosclerosis (MESA).	Population or CT imaging context was outside the eligibility criteria.
FTX146	Zair AM et al (2023)	An automated segmentation of coronary artery calcification using deep learning in specific region limitation.	Population or CT imaging context was outside the eligibility criteria.
FTX147	Zhang, WW et al (2019)	An end-to-end joint learning framework of artery-specific coronary calcium scoring in non-contrast cardiac CT	Population or CT imaging context was outside the eligibility criteria.
FTX148	Delgado, Y et al (2025)	Analysis of Discrepancies Between Radiologist and AI CCTA Readings Across Clinical Thresholds	Population or CT imaging context was outside the eligibility criteria.
FTX149	Nitta, K et al (2004)	Assessment of coronary artery calcification in hemodialysis patients using multi-detector spiral CT scan	Population or CT imaging context was outside the eligibility criteria.
FTX150	Jurgens PT et al (2021)	Association of Abdominal Aorta Calcium and Coronary Artery Calcium with Incident Cardiovascular and Coronary Heart Disease Events in Black and White Middle-Aged People: The Coronary Artery Risk Development in Young Adults Study.	Population or CT imaging context was outside the eligibility criteria.
FTX151	Craiem, D et al (2016)	Association of thoracic aorta calcium and non cardiac vascular events in cardiac disease-free individuals	Population or CT imaging context was outside the eligibility criteria.
FTX152	Nguyen, AM et al (2025)	Automated coronary calcium detection and scoring on multicenter, multiprotocol noncontrast CT	Population or CT imaging context was outside the eligibility criteria.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX153	Wu, J et al (2009)	Automated Coronary Calcium Scoring Using Predictive Active Contour Segmentation	Population or CT imaging context was outside the eligibility criteria.
FTX154	Kianoush, S et al (2024)	AUTOMATED PLAQUE ANALYSIS IN PATIENTS WITHOUT CORONARY ARTERY CALCIUM ON CORONARY CT ANGIOGRAPHY	Population or CT imaging context was outside the eligibility criteria.
FTX155	Föllmer B et al (2024)	Automated segment-level coronary artery calcium scoring on non-contrast CT: a multi-task deep-learning approach.	Population or CT imaging context was outside the eligibility criteria.
FTX156	Saur SC et al (2008)	Automatic detection of calcified coronary plaques in computed tomography data sets.	Population or CT imaging context was outside the eligibility criteria.
FTX157	Išgum I et al (2018)	Automatic determination of cardiovascular risk by CT attenuation correction maps in Rb-82 PET/CT.	Population or CT imaging context was outside the eligibility criteria.
FTX158	Asakawa, T et al (2023)	Cardiac Detection in Non-Contrast CT and Application to Calcium Scoring	Population or CT imaging context was outside the eligibility criteria.
FTX159	Shi Y et al (2026)	Clinical Metadata Guided Limited-Angle CT Image Reconstruction.	Population or CT imaging context was outside the eligibility criteria.
FTX160	Abdelmageed, A et al (2024)	CLOUD-BASED AI-ENHANCED DUAL-MODE SYSTEM FOR AUTOMATIC CORONARY ARTERY CALCIFICATION DETECTION AND QUANTIFICATION	Population or CT imaging context was outside the eligibility criteria.
FTX161	Waheed, S et al (2016)	Collective impact of conventional cardiovascular risk factors and coronary calcium score on clinical outcomes with or without statin therapy: the St Francis Heart Study	Population or CT imaging context was outside the eligibility criteria.
FTX162	Shanbhag, AD et al (2023)	Comparative analysis between convolutional long short-term memory networks and vision transformers for Coronary Calcium scoring in non-contrast CT	Population or CT imaging context was outside the eligibility criteria.
FTX163	Sánchez, CI et al (2012)	Contextual computer-aided detection: Improving bright lesion detection in retinal images and coronary calcification identification in CT scans	Population or CT imaging context was outside the eligibility criteria.
FTX164	Castro R et al (2025)	Contrastive Coronary Artery Calcification Image Retrieval in Computed Tomography.	Population or CT imaging context was outside the eligibility criteria.
FTX165	Serafin Z et al (2009)	Coronary artery calcium distribution and interscan measurement variability in end-stage renal and coronary heart disease patients.	Population or CT imaging context was outside the eligibility criteria.
FTX166	Datong C et al (2019)	Coronary Calcium Detection Based on Improved Deep Residual Network in Mimics.	Population or CT imaging context was outside the eligibility criteria.
FTX167	Esmaeili, Z et al (2026)	DEEP LEARNING-DERIVED PERICARDIAL ADIPOSE TISSUE BY ECG-GATED COMPUTED TOMOGRAPHY PREDICTS CARDIOVASCULAR EVENTS BEYOND CORONARY CALCIUM	Population or CT imaging context was outside the eligibility criteria.
FTX168	Baskaran L et al (2025)	Design and Rationale of GPS-CAD: Global Pretest Probability Study of Coronary Artery Disease.	Population or CT imaging context was outside the eligibility criteria.
FTX169	Gokmen MS et al (2026)	DINO-LG: Enhancing vision transformers with label guidance for coronary artery calcium detection.	Population or CT imaging context was outside the eligibility criteria.
FTX170	Singh, G et al (2021)	End-to-End, Pixel-Wise Vessel-Specific Coronary and Aortic Calcium Detection and Scoring Using Deep Learning	Population or CT imaging context was outside the eligibility criteria.
FTX171	Graves, CV et al (2024)	Enhancing Calcium Score Quantification in Cardiac Images through Robust Mediastinum Segmentation	Population or CT imaging context was outside the eligibility criteria.
FTX172	Becker CR (2005)	Estimation of cardiac event risk by MDCT.	Population or CT imaging context was outside the eligibility criteria.
FTX173	Park D et al (2025)	Evaluating the generalizability of an automated coronary artery calcium segmentation and scoring algorithm using multi-vendor dataset.	Population or CT imaging context was outside the eligibility criteria.
FTX174	Martin SS et al (2020)	Evaluation of a Deep Learning-Based Automated CT Coronary Artery Calcium Scoring Algorithm.	Population or CT imaging context was outside the eligibility criteria.
FTX175	Kitamura A et al (2005)	Evaluation of coronary artery calcification by multi-detector row computed tomography for the detection of coronary artery stenosis in Japanese patients.	Population or CT imaging context was outside the eligibility criteria.
FTX176	Xing K et al (2026)	Evaluation of the impact of cardiopulmonary rehabilitation exercise training on cardiopulmonary function in patients with chronic obstructive pulmonary disease complicated by unstable angina pectoris using a hierarchical deep learning CT image model.	Population or CT imaging context was outside the eligibility criteria.
FTX177	Dey, D et al (2009)	Fast Computer-Aided CT-Based Quantitation of Pericardial and Thoracic Fat and Their Association With Coronary Calcium	Population or CT imaging context was outside the eligibility criteria.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX178	Winkelmann MT et al (2022)	Fully Automated Artery-Specific Calcium Scoring Based on Machine Learning in Low-Dose Computed Tomography Screening.	Population or CT imaging context was outside the eligibility criteria.
FTX179	Sartoretti T et al (2023)	Fully automated deep learning powered calcium scoring in patients undergoing myocardial perfusion imaging.	Population or CT imaging context was outside the eligibility criteria.
FTX180	Kim S et al (2021)	Histogram-derived modified thresholds for coronary artery calcium scoring with lower tube voltage.	Population or CT imaging context was outside the eligibility criteria.
FTX181	Marcinkiewicz AM et al (2025)	Holistic AI analysis of hybrid cardiac perfusion images for mortality prediction.	Population or CT imaging context was outside the eligibility criteria.
FTX182	Kruk, M et al (2022)	Identification of non-calcified coronary plaque characteristics using machine learning radiomic analysis of non-contrast high-resolution CT	Population or CT imaging context was outside the eligibility criteria.
FTX183	Miller RJH et al (2025)	Integrating perfusion with AI-derived coronary calcium on CT attenuation scans to improve selection of low-risk studies for stress-only SPECT myocardial perfusion imaging.	Population or CT imaging context was outside the eligibility criteria.
FTX184	Zhai Z et al (2022)	Learning coronary artery calcium scoring in coronary CTA from non-contrast CT using unsupervised domain adaptation.	Population or CT imaging context was outside the eligibility criteria.
FTX185	Hou, JJ et al (2024)	LEARNING DISENTANGLED REPRESENTATION FOR VESSEL-SPECIFIC CORONARY ARTERY CALCIUM SCORING	Population or CT imaging context was outside the eligibility criteria.
FTX186	Liu Q et al (2010)	Lesion-specific coronary artery calcium quantification for predicting cardiac event with multiple instance support vector machines.	Population or CT imaging context was outside the eligibility criteria.
FTX187	Santos, R et al (2024)	Lightweight 3D CNN for the Segmentation of Coronary Calcifications and Calcium Scoring	Population or CT imaging context was outside the eligibility criteria.
FTX188	Wang, SW et al (2023)	LPE-Unet: An Improved UNet Network Based on Perceptual Enhancement	Population or CT imaging context was outside the eligibility criteria.
FTX189	Nakanishi, R et al (2017)	Machine Learning Adds to Standard Clinical & CAC Assessments in Predicting 10-Year Coronary Heart Disease & Cardiovascular Disease Deaths: Insight From the Coronary Artery Calcium Consortium of 66,636 Patients	Population or CT imaging context was outside the eligibility criteria.
FTX190	Lee, J et al (2025)	Machine learning prediction of myocardial ischemia using quantitative assessment of coronary calcifications in non-contrast CT calcium scoring scans	Population or CT imaging context was outside the eligibility criteria.
FTX191	Zimmermann GS et al (2019)	Predictive value of coronary calcifications for future cardiac events in asymptomatic patients: underestimation of risk in asymptomatic smokers.	Population or CT imaging context was outside the eligibility criteria.
FTX192	Javaid, A et al (2021)	Predictors of Coronary Artery Calcium and Long-Term Risks of Death, Myocardial Infarction, and Stroke in Young Adults	Population or CT imaging context was outside the eligibility criteria.
FTX193	Schulman-Marcus J et al (2014)	Prognostic utility of coronary artery calcium scoring in active smokers: a 15-year follow-up study.	Population or CT imaging context was outside the eligibility criteria.
FTX194	Wu, X et al (2019)	Pulse pressure correlates with coronary artery calcification and risk for coronary heart disease: a study of elderly individuals in the rural region of Southwest China	Population or CT imaging context was outside the eligibility criteria.
FTX195	Jayawardena, E et al (2018)	RELIABILITY OF CARDIAC FAT VOLUMES MEASURED BY NON-CONTRAST CORONARY CALCIUM COMPUTED TOMOGRAPHY USING A SEMI-AUTOMATED CARDIAC FAT ANALYSIS SOFTWARE	Population or CT imaging context was outside the eligibility criteria.
FTX196	Li, BY et al (2026)	Robustness of Deep Learning-Based Opportunistic Quantitative Assessment in Low-Dose CT	Population or CT imaging context was outside the eligibility criteria.
FTX197	van der Aalst, CM et al (2020)	Screening for cardiovascular disease risk using traditional risk factor assessment or coronary artery calcium scoring: the ROBINSCA trial	Population or CT imaging context was outside the eligibility criteria.
FTX198	Santos R et al (2025)	Segmentation of coronary calcifications with a domain knowledge-based lightweight 3D convolutional neural network.	Population or CT imaging context was outside the eligibility criteria.
FTX199	TANG, WY et al (1994)	THE AFFECTS OF PARTICLE-SIZE, SLICE THICKNESS AND RECONSTRUCTION ALGORITHM ON CORONARY CALCIUM QUANTITATION USING ULTRAFAST COMPUTED-TOMOGRAPHY	Population or CT imaging context was outside the eligibility criteria.
FTX200	Brunner, G et al (2010)	Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data	Population or CT imaging context was outside the eligibility criteria.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX201	Soudagar A et al (2025)	Transfer Learning Strategies for Cardiovascular Disease Detection in ECG Imagery.	Population or CT imaging context was outside the eligibility criteria.
FTX202	Fentanes E et al (2018)	Treatment in a preventive cardiology clinic utilizing advanced practice providers effectively closes atherosclerotic cardiovascular disease risk-management gaps among a primary-prevention population compared with a propensity-matched primary-care cohort: A team-based care model and its impact on lipid and blood pressure management.	Population or CT imaging context was outside the eligibility criteria.
FTX203	Paiva, M et al (2022)	Use of coronary calcium score to refine the cardiovascular risk classification of the new SCORE2 and SCORE2-OP algorithms in patients undergoing coronary CT angiography	Population or CT imaging context was outside the eligibility criteria.
FTX204	Sekar B et al (2015)	Using coronary artery calcification combined with pretest clinical risk assessment as a means of determining investigation and treatment in patients presenting with chest pain in a rural setting.	Population or CT imaging context was outside the eligibility criteria.
FTX205	Muscogiuri E et al (2023)	Validation of a convolutional neural network algorithm for calcium score quantification using a multivendor dataset.	Population or CT imaging context was outside the eligibility criteria.
FTX206	Lehmann, N et al (2018)	Value of Progression of Coronary Artery Calcification for Risk Prediction of Coronary and Cardiovascular Events Result of the HNR Study (Heinz Nixdorf Recall)	Population or CT imaging context was outside the eligibility criteria.
FTX207	Arnold, BA et al (2012)	Very small calcifications are detected and scored in the coronary arteries from small voxel MDCT images using a new automated/calibrated scoring method with statistical and patient specific plaque definitions	Population or CT imaging context was outside the eligibility criteria.
FTX208	Dehmeshki J et al (2007)	Volumetric quantification of atherosclerotic plaque in CT considering partial volume effect.	Population or CT imaging context was outside the eligibility criteria.
FTX209	Kim, H et al (2023)	Clinical outcomes of a CT protocol for simultaneous examination of the aorta and coronary artery in patients with aortic aneurysm	Target condition was not coronary artery calcium.
FTX210	Zhou, Jianhang et al (2026)	Hidden risk in normal myocardial perfusion scans: AI-detected proximal coronary calcium on CT attenuation maps improves prognosis.	Same report already included as R094 / STUDY120. The journal publication updates the previously retrieved preprint; no additional study or independent cohort.
FTX211	Fujito, Hidesato et al (2026)	Quantitative coronary artery and aortic valve calcification by virtual non-contrast Images from dual-energy delayed CT scans in patients with severe aortic stenosis	ECG-gated cardiac CT only; no separately extractable non-gated CAC assessment.
FTX212	Zhou, Sang et al (2026)	Age-Dependent Efficacy of Incidental Coronary Artery Calcium on Pre-CAG Chest CT: Potentially Missed Prevention Opportunities in Chinese Population.	Visual CAC assessment by radiologists; no AI-based CAC assessment.
FTX213	Belardo A. et al (2026)	Evidence of different predictors for different types of cardiotoxicities after hypo-fractionated whole breast radiotherapy	Total cardiac calcification combines coronary, valvular, and myocardial calcification; no separate coronary CAC result.
FTX214	He, Zhong-Yun et al (2026)	Development and validation of a clinical factor-based nomogram for predicting imaging-defined cardiopulmonary abnormality risk in an asymptomatic screening population.	AI-CAC is one component of a contemporaneous composite imaging phenotype; no eligible AI-CAC measurement, clinical prognosis, or care outcome.
FTX215	Burzić, Amna et al (2026)	Structured triage of incidental findings in lung cancer screening: Outcomes from a UK programme.	Clinical pathway uses radiologists' qualitative CAC grading; no AI-CAC assessment or AI-triggered pathway.
FTX216	Gaisinsky, Adam et al (2026)	Actionable Incidental Findings at Baseline Lung Cancer Screening: Identifying Opportunities in a Universal Healthcare System.	Automated extraction of existing visual radiology reports; no AI image-based CAC assessment.
FTX217	Pan, Abigail S et al (2026)	Association of Hyperlipidemia and Cardiovascular-Kidney-Metabolic Risk Factor Severity With Incidental Coronary Artery Calcium.	Visual CAC grading by two thoracic radiologists; no AI-based CAC assessment.
FTX218	Cheng, Yung-Yin (2026)	Opportunistic Coronary Artery Calcium Reporting on Non-Gated Chest CT: From Detection to Clinical Action.	Commentary letter without original study data.
FTX219	Zhang, He et al (2026)	High-pitch Low-dose Chest Computed Tomography With Calcium-aware Reconstruction for Coronary Artery Calcium Scoring: Accuracy and Dose Reduction.	Deterministic acquisition/reconstruction correction with physician CAC scoring; no AI CAC index assessment.
FTX220	Fletcher, Abbey Louise et al (2026)	Patient acceptability and preferences for communication of coronary artery calcification findings in the NHS targeted lung health checks: a qualitative study.	Interviews following routine visual CAC reporting; no AI-CAC assessment or AI-triggered pathway.
FTX221	Dharmavaram, Gourisree et al (2026)	Incidental steatotic liver in young patients with ischemic stroke: A marker of high vascular burden and underrecognized prevention opportunity.	Manual review of existing clinical imaging reports; no AI-CAC assessment.

Supplementary Table 19 continued

Report	Author and year	Title	Reason for exclusion
FTX222	Groner, Lauren K et al (2026)	Opportunistic Screening for Cardiovascular and Chronic Obstructive Pulmonary Disease Biomarkers and Implications for Downstream Care: A Retrospective Lung Cancer Screening Cohort Study.	Manual review of completed radiology reports; no AI-CAC assessment.
FTX223	Nalepa J. et al (2026)	Enhancing coronary artery calcium scoring via multi-organ segmentation on non-contrast cardiac computed tomography	The formal journal report evaluates ECG-gated cardiac CT only; no eligible non-gated CAC results.
FTX224	Fleming M.R. et al (2026)	Initial Staging (18)F-FDG PET/CT for Coronary Artery Calcium Scoring to Assess Cardiovascular Risk in Women with Breast Cancer	Traditional semiautomated Agatston scoring; no AI or fully automatic CAC index assessment.

Supplementary Table 20. Reports sought but not retrieved (n = 7)

Report	Author and year	Title	Retrieval status and attempts
NR001	McConachie P et al (2022)	Accurate measurement of coronary artery calcium in cancer patients using the CT component of PET/CT scans.	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.
NR002	Priyatharshini R. et al (2016)	An active contour model based multi objective optimization technique for coronary calcium scoring	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.
NR003	Satoh, H et al (2007)	Computer-aided diagnosis workstation and network system for chest diagnosis based on miltislice CT images	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.
NR004	Sato, H et al (2005)	Computer-aided diagnosis workstation for chest diagnosis based on multihelical CT images	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.
NR005	Satoh, H et al (2006)	Computer-aided diagnosis workstation and database system for chest diagnosis based on multihelical CT images	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.
NR006	Ukai, Y et al (1996)	An algorithm for coronary calcification diagnosis based on helical CT images	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.
NR007	Sato, H et al (2004)	Computer-aided diagnosis workstation and database system for chest diagnosis based on multihelical CT images	Full report not retrieved The exact report could not be obtained after searches by title, author, DOI or proceedings record, and checks of bibliographic databases, publisher pages, institutional repositories, and related-report reference lists.

Supplementary Data 1. Aggregate datasets accompanying the Article

Microsoft Excel workbook with 39 worksheets. Read me: contents, analysis unit, and missing-data conventions. S1 Report characteristics: characteristics of all 110 included reports. S2 RoB applicability: risk-of-bias and applicability judgments with reasons. S3 Selected estimates: selected diagnostic-accuracy estimates and 2×2 counts. S4 Signalling responses: QUADAS-3 signalling responses and rationale. S5 Continuous agreement, S5 Measurement calibration, and S5 Data dictionary: continuous-agreement measures, cutpoint derivation and validation, and field definitions. S6 Global tests, S6 Component contrasts, and S6 Subgroup estimates: subgroup analyses and meta-regression. S7 Full-text exclusions and S8 Reports not retrieved. S9A–S9D: diagnostic sensitivity analyses, likelihood ratios, predictive values, and failed-scan mapping. S10 Evidence context. S11A–S11E: cross-classification evidence base, table inventory, count tables, and category agreement. S12A–S12E: cohort map, cohort records, overlap pairs, algorithm lineage, and estimate selection. S13 Prognostic estimates. S14 Implementation outcomes. Source-data worksheets: Table 1 source, Table 2 source, Table 3 source, Figure 1 source, Figures 3–6 source, Figure 7 matrix, Figure 7 probabilities, Figure 8 source, and Figure 8 studies.