Validation and Prognosis of Coronary Artery Calcium Scoring in Non-Triggered Thoracic Computed Tomography: Systematic Review and Meta-Analysis

Xie et al: Coronary Calcium Scoring in Non-Triggered CT

Xueqian Xie, MD*†; Yingru Zhao, MD*†; Geertruida H. de Bock, PhD,‡; Pim A. de Jong, MD, PhD§; Willem P. Mali, MD, PhD§; Matthijs Oudkerk, MD, PhD†; Rozemarijn Vliegenthart, MD, PhD*†

*Department of Radiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
†Center for Medical Imaging–North East Netherlands, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
‡Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
§Department of Radiology, University Medical Center Utrecht, The Netherlands

Correspondence to
Rozemarijn Vliegenthart, MD, PhD
Center for Medical Imaging–North East Netherlands
Department of Radiology, University of Groningen, University Medical Center Groningen
Hanzeplein 1, 9700 RB Groningen, The Netherlands
Telephone number: +31-50-3610251
Fax number: +31-50-3611798
E-mail: r.vliegenthart@umcg.nl

DOI: 10.1161/CIRCIMAGING.113.000092

Journal Subject Codes: Diagnostic testing:[30] CT and MRI
Abstract

Background—Coronary calcium score (CS), traditionally based on electrocardiography (ECG)-triggered computed tomography (CT), predicts cardiovascular risk. Nowadays, non-triggered thoracic CT is extensively utilized, such as in lung cancer screening. The study-purpose was to determine the correlation in CS between non-triggered and ECG-triggered CT, and to evaluate the prognostic performance of the CS derived from non-triggered CT.

Methods and Results—PubMed, Embase and Web of Knowledge were searched until November, 2012. Two reviewers independently screened 2,120 records to identify studies reporting the CS in non-triggered CT, and extracted information. Study quality was evaluated by standardized assessment tools. Cohen’s κ was extracted for agreement of CS categories between non-triggered and ECG-triggered CT (validation). Hazard ratio was extracted for prognostic performance. Five studies comprising 1,316 individuals were included regarding validation. Five studies comprising 34,028 cardiac asymptomatic individuals, mainly from lung cancer screening trials, were included regarding prognosis. All studies were of high quality. Meta-analysis could only be performed for validation studies, as studies on prognostic performance were highly heterogeneous. Pooled Cohen’s κ for agreement between the two techniques was 0.89 (95%CI: 0.83 to 0.95) for increasing CS categories. Increasing CS categories were associated with increasing risk of cardiovascular death or events. Non-triggered CT yielded false-negative CS in 8.8% of individuals, and underestimated high CS in 19.1%.

Conclusions—Our analysis shows the prognostic value and potential role of non-triggered assessment of coronary calcium, but it does not suggest that ECG-triggered CT should be replaced by non-triggered exams.

Key Words: systematic review, meta-analysis, computed tomography, calcium score; non-triggered cardiac imaging technique
Abbreviations and acronyms:

CC = correlation coefficient
CS = calcium score
CI = confidence interval
CT = computed tomography
ECG = electrocardiography
HR = hazard ratio
MDCT = multi-detector computed tomography
QUADAS = the Quality Assessment of Diagnostic Accuracy Studies
The amount of coronary artery calcium, based on computed tomography (CT) and traditionally expressed as calcium score (CS) according to Agatston,\(^1\) is a strong predictor of cardiovascular events.\(^2\)\(^-\)\(^5\) Calcium scoring has been found to improve cardiovascular risk stratification beyond cardiovascular risk factors.\(^4\)\(^,\)\(^6\) Due to the irregular and periodic movements of coronary arteries, electrocardiography (ECG)-triggered cardiac acquisition techniques are applied in CT to minimize motion artifacts and optimize calcium scoring.\(^3\)

Compared to ECG-triggered CT, non-triggered CT is extensively utilized. In 2007, 13.6 million non-triggered thoracic CT examinations were performed in the United States, in contrast to 0.7 million ECG-triggered CT examinations for calcium scoring.\(^7\) Recent trial results have increased the interest in lung cancer screening by thoracic CT.\(^8\) Thus, the number of non-triggered examinations will likely further increase. Age and smoking, the current selection criteria for lung cancer screening, are also correlated with coronary calcification and coronary heart disease.\(^9\) In lung cancer screening, coronary calcification is a frequent finding.\(^10\) If non-triggered CT can be used for calcium scoring, to stratify individuals in categories of cardiovascular risk and to identify those at high cardiovascular risk, there may be an enormous unused primary prevention potential.\(^11\) Also, deriving the CS from the same examination as used in lung cancer screening may positively impact the cost-effectiveness of screening.

Because motion of coronary arteries influences calcium scoring,\(^12\) the utilization of coronary calcium scoring in non-triggered CT is still being debated.\(^13\) With the increasing interest in lung cancer screening, this is an optimal moment to investigate the potential utilization of non-triggered CT for calcium scoring. However, compared with the extensive publications in ECG-triggered cardiac CT, the literature on calcium scoring in non-triggered thoracic CT is
relatively limited. Therefore, we conducted a systematic review and meta-analysis to investigate the validity and prognostic value of calcium scoring derived from non-triggered thoracic CT.

**Methods**

This study was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA).\textsuperscript{14}

**Information sources and search**

We searched PubMed, Embase, and Web of Knowledge until November, 2012, using terms related to computed tomography, non-triggered thoracic examination and coronary calcium without language restrictions (online-only Data Supplement Table A). Unpublished studies were not included.

**Study selection**

Two reviewers (XQ.X. and YR.Z.) with at least eight years of experience in thoracic and cardiovascular radiology participated in literature selection. Each record was evaluated independently. Disagreement in literature selection was resolved by consensus. Studies were included in the systematic review when they: 1) evaluated cardiac asymptomatic adult humans, or phantoms; 2) analyzed one of the following topics regarding calcium scoring in non-triggered CT: agreement between non-triggered and ECG-triggered CT, or prognostic performance to predict death or events; 3) used at least 16-row MDCT as ECG-triggered examination when ECG-triggered CT was used as reference examination. Sixteen-row MDCT was used as minimum CT generation because previous research showed higher accuracy and reproducibility in calcium scoring for 16-row MDCT compared to earlier generation CT systems.\textsuperscript{15}
Articles were excluded when they: 1) were reviews, abstracts, case reports or letters; 2) investigated participants with confounding factors, e.g., pacemaker or defibrillator implant, and cardiac surgery. When multiple similar publications based on the same trial were identified, only the study with the largest sample size was included to avoid possible duplicate reporting.

Subsequently, meta-analysis was performed in studies on agreement between non-triggered and ECG-triggered CT, when the studies used the same calcium scoring method, i.e., continuous CS and/or four CS categories (0, 1-99, 100-399, ≥400). No meta-analysis could be performed of the studies on prognostic value, because of large heterogeneity in calcium quantification methods, CS categorization and outcomes.

**Data collection process**

A standardized data extraction form was used to collect study and participant characteristics, methodology, and main results. Two reviewers (X.Q.X. and Y.R.Z.) collected data independently. Disagreement in data collection was resolved by consensus.

For results of studies on agreement of calcium scoring between non-triggered and ECG-triggered CT, a correlation coefficient (CC) was extracted for continuous data, and Cohen’s $\kappa$ and concordance percentage were extracted for categorical data. When available, the subject number with CS of >0, <400 and ≥400 was extracted for the two techniques. A CS of ≥400 is commonly considered as indicating high cardiovascular risk.3, 5 Thereafter, the diagnostic performance of non-triggered CT was calculated using ECG-triggered CT as reference. The percentage of false negative CS was calculated as the percentage of subjects with zero CS in non-triggered CT among subjects with CS>0 in ECG-triggered CT. The percentage of underestimated high-risk CS was considered as the percentage of subjects with CS<400 in non-triggered CT among subjects with CS≥400 in ECG-triggered CT. The percentage of
overestimated high-risk CS was calculated as the percentage of subjects with CS≥400 in non-triggered CT among subjects with CS<400 in ECG-triggered CT.

For prognostic performance of calcium scoring in non-triggered CT, hazard ratio (HR) for increasing CS categories derived from non-triggered CT to predict cardiovascular death or cardiovascular events (coronary heart disease, cerebrovascular disease, heart failure, peripheral arterial disease, aortic aneurysm, etc.) was extracted. When possible, unadjusted and adjusted HR with 95% confidence interval (CI) was extracted. Furthermore, the number of subjects with zero CS was extracted, as well as the number of subsequent cardiovascular deaths or events among these subjects.

**Study quality assessment**

Two reviewers (XQ.X. and YR.Z.) evaluated study quality independently on the studies included in the systematic review. Disagreement in quality assessment was resolved by consensus. Two quality assessment tools for different type of study were utilized to evaluate methodological quality and potential sources of bias, as described below.

For validation studies on agreement between non-triggered and ECG-triggered CT, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) with 14 standard items was utilized. For each study, a quality score was derived by assigning 1 point to each fulfilled item, 0.5 to an unclear item, and 0 to an unmet item, with a total possible score of 14 (online-only Data Supplement Table B).

For prognostic studies, the quality assessment criteria to evaluate reports on prognosis of CAC in American College of Cardiology Foundation/American Heart Association (ACCF/AHA) guideline with 8 standard criteria was utilized. For each study, a quality score was derived by
assigning 0 to 3 points to each criterion, with a total possible score of 16 (online-only Data Supplement Table C).

Synthesis of results and risk of bias

Pooling calculations of agreement between the two techniques were performed using the Hedges-Vevea random effects model and Z-test for overall effect. Pooling calculation was performed if there were at least two studies reporting the same measurement. Heterogeneity was tested using Q statistic and I² index. A two-sided \( P \) value for Q statistic<0.10 or \( I^2>50\% \) was considered to indicate heterogeneity. The random effects model was used regardless of the heterogeneity test, although results in Q statistics and I² index were still stated. Publication bias was evaluated with the Begg and Mazumdar rank correlation and Egger's regression test if the number of effect size in the included studies was at least three. For other statistical analysis, a two-sided \( P \) value<0.05 was considered as significant. Statistical analysis was performed using R 2.14.2 (R Foundation, Vienna, Austria) and Stata 11.0 (Statacorp LP, College Station, Texas).

Results

Study selection

The search of the three databases elicited 2,120 records after removal of duplicate records. Ten studies were included in systematic review, including five on agreement between non-triggered and ECG-triggered CT,\(^{17-21}\) and five on prognostic performance.\(^{10,22-25}\) Subsequently, meta-analysis was performed in three studies\(^{20,22,24}\) with consistent methodology on agreement between non-triggered and ECG-triggered CT (Figure 1).
Study characteristics

The systematic review included 35,344 participants (range of mean age, 51 to 65 years), comprising 21,558 (61%) men, 13,736 (39%) women and 50 (0.1%) individuals without indicated gender (Tables 1 and 2). Six (60%) studies were prospective, four (40%) retrospective. Four studies (40%) were from North America, three (30%) from Europe, and three (30%) from Asia. All studies were published in English.

Different CT modalities were utilized, ranging from single-slice to 64-row MDCT. Also CT systems as part of single-photon emission computed tomography (SPECT)/CT and positron emission tomography (PET)/CT were utilized.18 Low-dose acquisition was applied in eight studies (80%), normal dose in one study (10%). In one (10%) study the radiation dose was not reported. Scan data derived from non-triggered CT were reconstructed with different slice thicknesses, ranging from 1.25 to 10 mm (Tables 1 and 2). The most commonly used slice thicknesses were 2.5/3 mm and 5 mm. Four studies used a (medium-)smooth kernel,10, 20, 21, 25 the other studies did not indicate the applied kernel. Six studies utilized Agatston scoring,10, 17-19, 21, 25 while four others utilized visual grading of the presence and extent of coronary calcification. No study utilized contrast media. No phantom study was included.

Study quality

All five studies on agreement between the non-triggered and ECG-triggered CT were of high quality (score ≥ 10 according to the QUADAS tool). Suboptimal scores were present in two QUADAS items: no study mentioned uninterpretable results (item 13); three studies did not mention whether there were withdrawals (item 14) (online-only Data Supplement Table B).
All five studies on prognostic performance were of high quality (a score ≥12 according to the quality assessment criteria to evaluate prognosis of coronary calcification). Suboptimal scores were present for criterion 4: no study reported results by ethnicity (online-only Data Supplement Table C).

Validation of calcium scoring in non-triggered CT

Five studies were included in the systematic review, comprising 1,316 cardiac asymptomatic participants (Table 1).17–21 Diagnostic performance of non-triggered CT was calculated in four studies with 1,153 subjects (Table 3),17–19, 21 in which, 137 (11.9%) had CS of 100-400 in non-triggered CT. Fifty-five subjects (8.8%) showed no coronary calcification in non-triggered CT examination among 625 subjects with CS>0 in ECG-triggered CT. In those fifty-five subjects, fifty-two (8.3%) had CS 1-100 in ECG-triggered CT, and three (0.5%) had CS 100-400. Among 162 subjects with CS≥400 in ECG-triggered CT, non-triggered CT underestimated the CS in 31 subjects (19.1%). In these thirty-one subjects, two (1.2%) had CS 1-100 in non-triggered CT, and twenty-nine (17.9%) had CS 100-400. On the other hand, among 991 subjects with CS<400 in ECG-triggered CT, non-triggered CT showed a CS≥400 in twenty-six subjects (2.6%) and thus overestimated the CS. In those twenty-six subjects, one (0.1%) had CS 1-100 in ECG-triggered CT, and twenty-five (2.5%) had CS 100-400.

Meta-analysis was performed in three studies comprising 661 participants (Figure 2).17, 19, 21 The study by Kirsch could not be included because it evaluated the amount of coronary calcification using visual grading score.20 The pooled CC for CS was 0.94 (95%CI: 0.89 to 0.97). The pooled Cohen’s κ was 0.89 (95%CI: 0.83 to 0.95) for four categories of the calcium score. Heterogeneity was found in the pooling calculation of the calcium score (P for Q statistic<0.001 and I² >50%). No publication bias was found in the pooling calculation of the calcium score.
($P$>0.05). Publication bias testing was not performed in the pooling calculation of four CS categories, due to insufficient number of studies.

**Prognosis of calcium scoring in non-triggered CT**

Five studies were included, comprising 34,028 cardiac asymptomatic participants (Table 2).\(^{10,22-25}\) In the five studies, mean follow-up duration was 45 months (range, 10 to 72 months). None of the participants in the included studies had a history or symptoms of cardiovascular diseases before CT examination. During follow-up, 207 cardiovascular deaths and 675 cardiovascular events were observed. Overall, with increasing CS categories, increasing unadjusted and adjusted hazard ratio for cardiovascular death or events was observed. Risks in CS categories were not consistently reported, however in one study, unadjusted and adjusted hazard ratio increased up to 7.5 and 5.3 for CS>1000, respectively (Table 4).\(^{10}\)

A small percentage of subjects with zero CS in non-triggered CT suffered cardiovascular death or events. During a mean follow-up of 45 months, 47 cardiovascular deaths (0.55%) were found in 8,487 subjects with zero CS,\(^{22,24}\) whereas 72 cardiovascular events (1.3%) occurred in 5,249 subjects with zero CS.\(^{10,23}\) However, the event rate for subjects with positive CS was higher. During follow-up, 160 cardiovascular deaths (2.5%) were found in 6,415 subjects with positive CS,\(^{22,24}\) whereas 570 cardiovascular events (4.5%) occurred in 12,718 subjects with positive CS.\(^{10,23}\)

**Discussion**

In this systematic review and meta-analysis, we aimed to investigate whether coronary calcium scoring can be performed in non-triggered thoracic CT, for instance utilized in lung cancer screening. A strong correlation in calcium score categories between non-triggered and ECG-
triggered CT was found. In cardiac asymptomatic elderly and smokers, mainly from lung cancer screening trials, increasing coronary calcium burden translated into a higher risk of cardiovascular death or events.

Calcium score for individual atherosclerotic lesions is greatly influenced by motion. Regardless, we found that the correlation in CS between non-triggered and ECG-triggered CT was excellent ($r=0.94$) on a group level. In broad CS categories, we found a high agreement between non-triggered and ECG-triggered CT (Cohen’s $\kappa=0.89$). Thus, for an individual patient, although variability in CS between non-triggered and ECG-triggered CT is likely considerable, broad CS categories can potentially be utilized for cardiovascular risk stratification.

Absence of coronary calcification in ECG-triggered CT is associated with a very low cardiovascular risk, and thus is commonly utilized to rule out coronary artery disease. We found that non-triggered CT can yield a false-negative CS in about 9% of individuals compared to ECG-triggered CT. Furthermore, we found that a zero CS in non-triggered CT indicates a low cardiovascular risk, although non-triggered CT cannot reliably exclude coronary calcification. When a high CS ($\geq 400$) is found in asymptomatic individuals, the risk of cardiovascular events is elevated. The ACCF/AHA consensus document suggests to regard these individuals as candidates for intensive preventive therapies. The probability of overestimating the CS is low, and thus, it is reasonable to assume an elevated cardiovascular risk in case of a CS$\geq 400$ in non-triggered CT. On the other hand, non-triggered CT underestimated the CS in a non-negligible percentage of individuals with CS$\geq 400$ in ECG-triggered CT, thus underestimating cardiovascular risk. In the validation study, 11.9% had a CS of 100-400 in the non-triggered CT examination. In this relatively small proportion of the included study populations, dedicated ECG-triggered CT could be considered, to assess whether the CS is actually $\geq 400$. This
proportion is much lower than the population percentage in which calcium scoring could be considered according to current consensus documents (40%).

In this study, hazard ratios of CS categories for cardiovascular events were generally lower than in a previous systematic review on calcium scoring derived from ECG-triggered CT. For example, adjusted HR for cardiovascular events was up to 5.3 for CS>1000 in our study, lower than 10.8 in a previous report in ECG-triggered CT in an elderly population. The relative risk is usually based on the risk of subjects without coronary calcium at baseline as reference category. During a mean follow-up of 45 months, we found that 1.3% subjects without coronary calcium had a cardiovascular event. In contrast, a meta-analysis by Sarwar et al. on ECG-triggered CT reported only 0.47% of subjects without coronary calcium suffered a cardiovascular event during a mean follow-up of 50 months. In that meta-analysis, studies mainly consisted of middle-aged individuals at low-to-intermediate cardiovascular risk, referred for cardiovascular risk evaluation. In contrast, the majority of the populations in the prognostic studies on calcium scoring using non-triggered CT comprised participants of lung cancer screening trials. The generally higher age and heavier smoking history in the prognostic studies included in our study likely at least partly explain the higher event rate in individuals with zero CS. Besides, this higher event rate in case of zero CS for non-triggered CT may also be explained by the fact that a proportion of the individuals without coronary calcification on the non-triggered CT, actually have a positive CS in ECG-triggered CT. As this reference risk is higher, the relative risk for increasing CS categories also yield lower values. Our finding does suggest that presence of coronary calcification in non-triggered CT is an independent predictor of cardiovascular events. Also, we found that higher calcium burden translated into a higher cardiovascular risk in a large aggregated sample.
Reproducibility of calcium scoring in repeated non-triggered CT has been investigated. Jacobs investigated 584 subjects who underwent two non-triggered examinations of the thorax, and calculated the CS for both exams. The calcium scores were divided into the commonly used categories of 0, 1-100, 101-400, >400. In 440 cases (75%), the calcium scores of the two CT examinations fell in the same category. In 138 subjects (24%), calcium scores differed by one category, and in 6 subjects (1%) by more than one category. The intra-class correlation coefficient was 0.94. On the other hand, reproducibility of calcium scoring in ECG-triggered CT is also not perfect. Using ECG-triggered CT, Rutten reported that 76% to 85% of individuals ended up in the same CS category, and in 15% to 24%, the results differed by one category.

The agreement of repeated calcium scoring in non-triggered CT within and between observers is high, although slightly lower than in ECG-triggered CT. Nearly all studies in this systematic review investigated either intra- or inter-observer variability of calcium scoring in non-triggered CT. For example, in 483 subjects, Wu reported an inter-observer variability of 3.6% for ECG-triggered CT, and of 9.6% for non-triggered CT. However, all studies found a very strong concordance in score categorization within and between observers (kappa values 0.77 to 0.91, intraclass correlation coefficient 0.93 to 0.99).

The majority of included studies (80%) were based on low-dose thoracic CT, which has a lower radiation dose than a dedicated cardiac CT for calcium scoring. A typical effective radiation dose for low-dose CT utilized in lung cancer screening is 0.8-0.9mSv for normal sized body. However, the mean dose for a cardiac CT for calcium scoring is approximately 1.0-2.9mSv, depending on scanner type and scanning protocol.
Clinical implication

A large number of non-triggered CT examinations are annually performed world-wide. In the aging and smoking population, coronary calcification is a common finding. A lung cancer screening trial reported that over 70% of the participants had coronary calcification.10 The group at risk for lung cancer overlaps with the group at highest risk of cardiovascular diseases, because at least aging and smoking are two major risk factors for both diseases. There may be an enormous primary prevention potential if the calcium score can be derived from the same examination, at least in participants of lung cancer screening trials. While results from the one study in a clinical population suggest that the extent of coronary calcification is also predictive outside lung cancer screening setting, more studies are needed to confirm the value of calcium scoring in routine clinical thoracic CT.

We observed that CS categorization between non-triggered and ECG-triggered CT correlated very well, and increasing CS categories based on non-triggered CT are predictive of increasing cardiovascular risk. Thus, for subjects who were examined by non-triggered thoracic CT, the cardiovascular risk could potentially be stratified by performing calcium scoring. Subjects identified in non-triggered CT as having high CS could be considered as candidates of intensive risk factor modification, especially in an aging and smoking population such as the participants in lung cancer screening. However, a zero calcium score in non-triggered CT does not exclude coronary calcification.

Furthermore, cardiovascular event rate of subjects without CS in non-triggered CT is higher than in ECG-triggered CT. Absent coronary calcification in non-triggered CT may not reliably exclude the risk of cardiovascular events. Future studies on this topic are needed to provide stronger support for coronary calcium scoring in non-triggered CT.
Limitations

Firstly, despite our favorable results it remains to be clarified whether differences in the accuracy between non-triggered and ECG-triggered CS measures translate into differences in prognostic value. For example, a zero CS in non-triggered CT may render a positive CS in ECG-triggered exams. Secondly, for agreement in calcium scoring between non-triggered and ECG-triggered CT, the number of studies and participants in the meta-analysis was relatively low. To compare calcium scoring between non-triggered and ECG-triggered CT, the patients have to be scanned twice in a short time-interval, and at doubled radiation dose. This could contribute to the relatively small number of studies in this field. Our conclusions are not only based on the pooling calculations but also on the systematic review. Besides, for the second part of our study, regarding the prognostic value, the aggregated sample size was over 30,000 individuals. Thirdly, different calcium scoring methods were utilized in studies on prognostic performance. Meta-analysis could therefore not be performed to assess predictive value of the calcium score derived from non-triggered CT. However, heavier calcium burden was in the systematic review associated with increasing cardiovascular risk. Finally, included studies were fairly heterogeneous in terms of participant population, imaging equipment and acquisition protocol. Those factors weakened the strength of meta-analysis. We used a random effects model to compensate for at least some of the heterogeneity in the pooling calculation. On the other hand, the differences in imaging procedures also reflect the heterogeneity of procedures in clinical practice. Despite different CT equipment and calcium scoring methods, at least the presence or absence of coronary calcium is clear. Results on the presence and absence of coronary calcification should not differ significantly based on important CS categories. Thus, our conclusions based on findings related to those points are solid.
Conclusions

In this systematic review and meta-analysis, strong agreement in CS categorization was found between non-triggered CT and ECG-triggered CT. Compared to ECG-triggered CT, a high calcium score category in non-triggered CT is a fairly reliable finding. However, non-triggered CT yielded false-negative CS in 8.8% of individuals, and underestimated high CS in 19.1%. In cardiac asymptomatic participants mainly from lung cancer screening trials, increasing CS categories in non-triggered CT were associated with increasing risk of cardiovascular events. Our analysis presents preliminary evidence for the prognostic value and potential role of calcium scoring in non-triggered CT. However, it does not suggest that non-triggered examinations can replace dedicated, ECG-triggered CT.

Acknowledgements

The authors wish to thank Dr. Estelle J.K. Noach for revision of the manuscript.

Disclosures

None.

References


Table 1. Characteristics of studies on agreement of calcium scoring between non-triggered and triggered CT

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Patients n</th>
<th>Men, %</th>
<th>Age, year±SD</th>
<th>Setting of study</th>
<th>Type of participants</th>
<th>CT type</th>
<th>Radiation dose setting</th>
<th>Slice thickness, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budoff 2011</td>
<td>50</td>
<td>n/a</td>
<td>n/a</td>
<td>COPD cohort</td>
<td>Invited smokers of &gt;10 pack-years</td>
<td>64-MDCT</td>
<td>Low</td>
<td>2.5</td>
</tr>
<tr>
<td>Einstein 2010</td>
<td>492</td>
<td>44</td>
<td>n/a</td>
<td>Routine clinical population referred adults</td>
<td>Consecutively referred adults</td>
<td>16-SPECT/CT 16-PET/CT 64-PET/CT</td>
<td>Low</td>
<td>n/a</td>
</tr>
<tr>
<td>Kim 2008</td>
<td>128</td>
<td>100</td>
<td>52±7</td>
<td>Lung cancer screening</td>
<td>Consecutively referred elder smokers</td>
<td>40-MDCT</td>
<td>Low</td>
<td>2.5</td>
</tr>
<tr>
<td>Kirsch 2011</td>
<td>163</td>
<td>78</td>
<td>51±9</td>
<td>Asymptomatic</td>
<td>Consecutively referred elder adults</td>
<td>16- and 64MDCT</td>
<td>Normal</td>
<td>5.0</td>
</tr>
<tr>
<td>Wu 2008</td>
<td>483</td>
<td>66</td>
<td>62±13</td>
<td>Lung cancer screening</td>
<td>Self-referred elder adults</td>
<td>16MDCT</td>
<td>Low</td>
<td>3.0</td>
</tr>
</tbody>
</table>

SD=standard deviation; CT=computed tomography; n/a=not available; COPD = chronic obstructive pulmonary disease; ECG=electrocardiographic; MDCT=multi-detector computed tomography; SPECT=single-photon emission computed tomography; PET=positron emission tomography
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Patients, n</th>
<th>Men, %</th>
<th>Age, year±SD</th>
<th>Setting of study</th>
<th>Type of participants</th>
<th>CT type</th>
<th>Radiation dose setting</th>
<th>Slice thickness, mm</th>
<th>Cohort name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itani 2004</td>
<td>6120</td>
<td>55</td>
<td>61±11</td>
<td>Lung cancer screening</td>
<td>Invited elder adults</td>
<td>Single-slice CT</td>
<td>Low</td>
<td>10</td>
<td>Nagano</td>
</tr>
<tr>
<td>Jacobs 2011</td>
<td>10410</td>
<td>58</td>
<td>62±12</td>
<td>Routine clinical population</td>
<td>Retrospectively included elder adults</td>
<td>4-, 8-, 16-, 40- and 64-MDCT</td>
<td>n/a</td>
<td>3.0-10</td>
<td>PROVIDI</td>
</tr>
<tr>
<td>Jacobs 2012</td>
<td>7557</td>
<td>83</td>
<td>60±6</td>
<td>Lung cancer screening</td>
<td>Invited elder heavy smokers</td>
<td>16-MDCT</td>
<td>Low</td>
<td>3.1</td>
<td>NELSON</td>
</tr>
<tr>
<td>Shemesh 2010</td>
<td>8782</td>
<td>49</td>
<td>65±7</td>
<td>Lung cancer screening</td>
<td>Invited elder heavy smokers</td>
<td>Single- and multi-slice CT</td>
<td>Low</td>
<td>1.25-5</td>
<td>I-ELCAP</td>
</tr>
<tr>
<td>Sverzellati</td>
<td>1159</td>
<td>68</td>
<td>58±6</td>
<td>Lung cancer screening</td>
<td>Invited elder heavy smokers</td>
<td>16-MDCT</td>
<td>Low</td>
<td>5</td>
<td>MILD</td>
</tr>
</tbody>
</table>

SD=standard deviation; CT=computed tomography; n/a=not available; MDCT=multi-detector computed tomography; PROVIDI=the prognostic value of unrequested information in diagnostic imaging; NELSON=the Dutch-Belgian randomized lung cancer screening trial; I-ELCAP = international early lung cancer action program; MILD=multi-centric Italian lung detection.
Table 3. Agreement of coronary calcium scoring between non-triggered thoracic and ECG-triggered cardiac CT, and diagnostic performance of calcium scoring in non-triggered CT

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Agreement</th>
<th>Diagnostic performance*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scoring in non-triggered CT</td>
<td>Reference scoring in triggered CT</td>
</tr>
<tr>
<td>Budoff 201117</td>
<td>CS</td>
<td>CS</td>
</tr>
<tr>
<td>Einstein 201018</td>
<td>4 categories of CS†</td>
<td>4 categories of CS†</td>
</tr>
<tr>
<td>Kim 200819</td>
<td>CS</td>
<td>CS</td>
</tr>
<tr>
<td>Kirsch 201120</td>
<td>Visual grading score§</td>
<td>CS</td>
</tr>
<tr>
<td>Wu 200821</td>
<td>CS</td>
<td>CS</td>
</tr>
</tbody>
</table>

CT=computed tomography; CS=calcium score; n/c=not calculated because different scoring systems were used in non-triggered and ECG-triggered CT.

*False negative calcium score is indicated as percentage of CS=0 in non-triggered CT among those with CS>0 in triggered CT. Underestimated high-risk calcium score is indicated as percentage of CS<400 in non-triggered CT among those with CS≥400 in triggered CT. Overestimated high-risk calcium score is indicated as percentage of CS≥400 in non-triggered CT among those with CS<400 in triggered CT.

†Four categories of CS defined as 0, 1-99, 100-399, and ≥400.
‡Six categories of CS, defined as 0, 1-9, 10-99, 100-399, 400-999 and ≥1000.

§Score assigned to each major coronary artery. 0: No calcification; 1: single pixel calcification; 3: dense calcification with blooming artifact; 2: calcification between 1 and 3. The visual grading score (range 0-12) was calculated by summing the score for each artery.
Table 4. Prognostic performance of coronary calcium scoring for cardiovascular death or events in non-triggered CT

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Follow-up, months (range)</th>
<th>Endpoint event, n</th>
<th>Calcium scoring method in non-triggered CT</th>
<th>Calcium scoring cut-off</th>
<th>Event number/category number, percentage</th>
<th>Unadjusted hazard ratio of calcium score (95%CI)</th>
<th>Adjusted hazard ratio of calcium score (95%CI)</th>
<th>Adjusted for factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itani 200422</td>
<td>48</td>
<td>Cardiac death, 14</td>
<td>Presence of coronary calcification</td>
<td>Zero calcium score</td>
<td>4/4914,0.08%</td>
<td>1.0(reference)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Jacobs 201123</td>
<td>18</td>
<td>Cardiovascular event, 515</td>
<td>4 categories of visual grading score*</td>
<td>Positive calcium score</td>
<td>10/1206,0.83%</td>
<td>2.7(0.8,9.4)</td>
<td>1.0(reference)</td>
<td>Age, sex, indication for CT, image quality and type of medical center</td>
</tr>
<tr>
<td>Jacobs 201224</td>
<td>10(1-21)</td>
<td>Cardiovascular event, 127</td>
<td>4 categories of calcium score</td>
<td>Calcium score:0</td>
<td>62/3435,1.8%</td>
<td>1.0(reference)</td>
<td>1.0(reference)</td>
<td>Age, sex, smoking status, hypertension, hypercholesterolemia and diabetes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Calcium score:1-100</td>
<td>113/2498,4.5%</td>
<td>2.8(2.0,3.8)</td>
<td>2.2(1.6,3.0)</td>
<td>Age, sex and smoking pack-years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Calcium score:101-1000</td>
<td>149/2603,5.7%</td>
<td>3.8(2.9,5.2)</td>
<td>2.5(1.8,3.4)</td>
<td>Age, sex, diabetes, hypertension, smoking status and smoking duration</td>
</tr>
</tbody>
</table>
|             |                           |                    |                                            | Calcium score:>1000     | 191/1874,10.2%                        | 6.9(5.2,9.2)                                 | 3.7(2.7,5.2) 
| Shemesh 201025 | 72(1-92)                  | Cardiovascular death, 193 | 3 categories of visual grading score†      | Visual score:0          | 43/3573,1.2%                            | 1.0(reference)                               | 1.0(reference)                               | Age, sex and smoking pack-years |
|             |                           |                    |                                            | Visual score:1-3        | 66/3569,1.8%                            | 1.6(1.1,2.4)                                 | 1.0(0.7,1.5)                                 | Age, sex, diabetes, hypertension, smoking status and smoking duration |
|             |                           |                    |                                            | Visual score:4-12       | 84/1640,5.1%                            | 4.7(3.6,6.8)                                 | 2.1(1.4,3.1)                                 | Age, sex, diabetes, hypertension, smoking status and smoking duration |
| Sverzellati 201226 | 36(1-54)                  | Cardiovascular event, 33 | 2 categories of calcium score              | Calcium score:<400      | 26/1079,2.4%                            | n/a                                          | 1.0(reference)                               | Age, sex, diabetes, hypertension, smoking status and smoking duration |
|             |                           |                    |                                            | Calcium score:>400      | 7/80,8.8%                              | n/a                                          | 2.9(1.1,7.3) 

CI=confidence interval; n/a=not available.

*Grading score assigned to each major coronary artery. 0:No calcification. 1:1-2 calcifications. 2:≥2 calcifications or one calcification extending ≥2 slices. 3:Calcification covering a large coronary segment. Four visual grades were stratified by the sum of the score (0, 1-2, 3-5 and 6-12).
† Grading score assigned to each major coronary artery. 0: No calcification. 1: $\leq \frac{1}{6}$ of artery length showing calcification. 2: $\frac{1}{6} < \frac{\sqrt{3}}{2}$. 3: $\geq \frac{\sqrt{3}}{2}$. Three visual grades were stratified by the sum of the score (0, 1-3 and 4-12).
Figure Legends

Figure 1. Flowchart of literature review and selection

Figure 2. Forest plots for agreement of coronary calcium scoring between non-triggered and electrocardiography-triggered computed tomography examinations

\( P(Q) = P \) value for Q statistic; \( P(Z) = P \) value for Z test; \( P(B) = P \) value for Begg and Mazumdar rank correlation test; \( P(E) = P \) value for Egger’s regression test.

*The four categories of the calcium score were defined as 0, 1-99, 100-399, and \( \geq 400 \).
Records identified through database searching (n=3059)

Records screened by title and abstract after duplicates removed (n=2120)

Full-text articles assessed for eligibility (n=77)

Studies included in systematic review:
- Studies on agreement (n=5)
- Studies on prognosis (n=5)

Studies included in meta-analysis:
- Studies on agreement (n=3)

Records excluded (n=2043)

Full-text articles excluded:
- Not in humans, or phantom (n=1)
- BCG-triggered examination (n=43)
- Not preferred topics (n=19)
- <16-MDCT as reference (n=2)
- Possible duplicate (n=2)
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Effect size (95% CI)</th>
<th>Size, n</th>
<th>Forest plots</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Calcium score</strong></td>
<td>Correlation coefficient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budoff 2011\textsuperscript{17}</td>
<td>0.96 (0.93, 0.98)</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Kim 2008\textsuperscript{19}</td>
<td>0.89 (0.85, 0.92)</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>Wu 2008\textsuperscript{21}</td>
<td>0.95 (0.94, 0.96)</td>
<td>483</td>
<td></td>
</tr>
<tr>
<td>Pooled</td>
<td>0.94 (0.89, 0.97)</td>
<td>661</td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td>$P(Q)&lt;0.001$, $I^2= 89.1%$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall effect</strong></td>
<td>$P(Z)&lt;0.001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Publication bias</strong></td>
<td>$P(B)=0.60$, $P(E)= 0.80$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**4 categories of calcium score***

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Concordance</th>
<th>Cohen’s $\kappa$</th>
<th>Size, n</th>
<th>Forest plots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budoff 2011\textsuperscript{17}</td>
<td>94%</td>
<td>0.90 (0.79, 1.00)</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Wu 2008\textsuperscript{21}</td>
<td>93%</td>
<td>0.89 (0.82, 0.96)</td>
<td>483</td>
<td></td>
</tr>
<tr>
<td>Pooled</td>
<td>0.89 (0.83, 0.95)</td>
<td>533</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td>$P(Q)= 0.88$, $I^2= 0%$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall effect</strong></td>
<td>$P(Z)&lt; 0.001$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Online-only Data Supplement Table A. Literature search strategy

**PubMed**

("Tomography, X-Ray Computed"[MeSH] OR "computed tomography"[tiab] OR CT[tiab] OR “MDCT”) AND ("untriggered" OR “ungated” OR “non-gated” OR “non-triggered” OR “non-electrocardiogram” OR “thorax”[MeSH] OR “chest” OR “thoracic” OR “lung” OR “pulmonary” OR “torso”) AND (“coronary vessels”[MeSH] OR "Coronary") AND (“Calcium” OR “calcification” OR “calcific” OR “calcified”) AND 1900/01:2012/11[dp]

**EmBase**

#1: ((Computed tomography) OR CT:ab,ti OR MDCT) AND (untriggered OR ungated OR non-gated OR non-triggered OR non-electrocardiogram OR thorax OR chest OR thoracic OR lung:ab,ti OR pulmonary:ab,ti OR torso) AND Coronary AND (Calcium OR calcification OR calcific OR calcified)

Grammar in advanced search: #1 AND [1-1-1900]/sd NOT (#1 AND [30-11-2012]/sd)

**Web of Knowledge**

#1 topic: ((Computed tomography) OR CT OR MDCT)

#2 topic: (untriggered OR ungated OR non-gated OR non-triggered OR non-electrocardiogram OR thorax OR chest OR thoracic OR lung OR pulmonary OR torso)

#3 topic: Coronary

#4 topic: (Calcium OR calcification OR calcific OR calcified)

Grammar: #1 topic and #2 topic and #3 topic and #4 topic
Online-only Data Supplement Table B. Quality assessment for validation studies on agreement and diagnostic performance, by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool

<table>
<thead>
<tr>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1: Representative patient sample</td>
</tr>
<tr>
<td>Item 2: Selection criteria clearly described</td>
</tr>
<tr>
<td>Item 3: Acceptable reference standard</td>
</tr>
<tr>
<td>Item 4: Acceptable delay between tests*</td>
</tr>
<tr>
<td>Item 5: Partial verification avoided</td>
</tr>
<tr>
<td>Item 6: Differential verification avoided</td>
</tr>
<tr>
<td>Item 7: Incorporation avoided</td>
</tr>
<tr>
<td>Item 8: Adequate index test description</td>
</tr>
<tr>
<td>Item 9: Adequate reference standard description</td>
</tr>
<tr>
<td>Item 10: Index test blinded to reference standard</td>
</tr>
<tr>
<td>Item 11: Reference standard blinded to index test</td>
</tr>
<tr>
<td>Item 12: Clinical data available as in practice</td>
</tr>
<tr>
<td>Item 13: Uninterpretable test results reported</td>
</tr>
<tr>
<td>Item 14: Withdrawals explained</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budoff 2011&lt;sup&gt;1&lt;/sup&gt;</td>
<td>12.0</td>
</tr>
<tr>
<td>Einstein 2010&lt;sup&gt;2&lt;/sup&gt;</td>
<td>12.0</td>
</tr>
<tr>
<td>Kim 2008&lt;sup&gt;3&lt;/sup&gt;</td>
<td>12.5</td>
</tr>
<tr>
<td>Kirsch 2011&lt;sup&gt;4&lt;/sup&gt;</td>
<td>13.0</td>
</tr>
<tr>
<td>Wu 2008&lt;sup&gt;5&lt;/sup&gt;</td>
<td>13.5</td>
</tr>
</tbody>
</table>

*Maximum delay of 2 months between non-triggered and reference examination was considered as acceptable.

†For each study, a quality score was accumulated by assigning 1 point to “yes” item, 0.5 point to “unclear” item, and 0 to “no” item. The total possible score was 14. A score of ≥10 points was considered as high quality, and a score between 6 and 9 points as moderate quality, a score of ≤5 as low quality.
Online-only Data Supplement Table C. Quality assessment for studies on prognosis, by the quality assessment criteria of prognostic studies on coronary artery calcium in American College of Cardiology Foundation / American Heart Association (ACCF/AHA) guideline

<table>
<thead>
<tr>
<th>Study</th>
<th>Criterion 1: Retrospective vs. prospective study</th>
<th>Criterion 2: Potential for referral bias</th>
<th>Criterion 3: Reporting coronary calcification by CHD death or myocardial infarction</th>
<th>Criterion 4: Reporting of results by gender or ethnicity</th>
<th>Criterion 5: Sample size greater than 1000</th>
<th>Criterion 6: Potential for limited challenge</th>
<th>Criterion 7: Risk factor reporting</th>
<th>Criterion 8: Covariate or risk-adjusted outcomes</th>
<th>Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itani 2004&lt;sup&gt;6&lt;/sup&gt;</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Jacobs 2011&lt;sup&gt;7&lt;/sup&gt;</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Jacobs 2012&lt;sup&gt;8&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Shemesh 2010&lt;sup&gt;9&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Sverzellati 2012&lt;sup&gt;10&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>14</td>
</tr>
</tbody>
</table>

CHD = coronary heart disease.

*For each study, a quality score was accumulated by assigning a score for each criterion as the following:

- Criterion 1: Retrospective vs. prospective study (1=retrospective, 2=prospective).
- Criterion 2: Potential for referral bias (0=clinically referred patients, 1=unselected cohort, 2=population sample).
- Criterion 3: Reporting coronary calcification by CHD death or myocardial infarction (1=no, 2=yes).
• Criterion 4: Reporting of results by gender or ethnicity (0=no, 1=gender only, 2=ethnicity only, 3=both).

• Criterion 5: Sample size ≥ 1000 (0=no, 1=yes).

• Criterion 6: Potential for limited challenge (1=no reporting of calcium outcomes in low- to high-risk global risk scores, 2=reporting of calcium outcomes in low- to high-risk global risk scores).

• Criterion 7: Risk factor reporting (1=historical only, 2=measured in subset, 3=measured in all subjects).

• Criterion 8: Covariate or risk-adjusted outcomes (0=no, 1=yes).

Total possible score was 16. A score of ≥11 points was considered as high quality, and a score between 7 and 10 points as moderate quality, a score of ≤6 as low quality.
References


