Use of 3×2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies

Georg M Schuetz research assistant1, Peter Schlattmann professor of medical statistics2, Marc Dewey chief consultant1

1The Charité-Universitätsmedizin Berlin, Humboldt-Universität zu Berlin, Freie Universität Berlin, Department of Radiology, 10117 Berlin, Germany; 2The Department of Medical Statistics, Informatics and Documentation, University Hospital of Friedrich Schiller University Jena, Jena, Germany

Abstract
Objective To determine whether a 3×2 table, using an intention to diagnose approach, is better than the “classic” 2×2 table at handling transparent reporting and non-evaluable results, when assessing the accuracy of a diagnostic test.

Design Based on a systematic search for diagnostic accuracy studies of coronary computed tomography (CT) angiography, full texts of relevant studies were evaluated to determine whether they could calculate an alternative 3×2 table. To quantify an overall effect, we pooled diagnostic accuracy values according to a meta-analytical approach.

Data sources Medline (via PubMed), Embase (via Ovid), and ISI Web of Science electronic databases.

Eligibility criteria Prospective English or German language studies comparing coronary CT with conventional coronary angiography in all patients and providing sufficient data for a patient level analysis.

Results 120 studies (10 287 patients) were eligible. Studies varied greatly in their approaches to handling non-evaluable findings. We found 26 studies (including 2298 patients) that allowed us to calculate both 2×2 tables and 3×2 tables. Using a bivariate random effects model, we compared the 2×2 table with the 3×2 table, and found significant differences for pooled sensitivity (98.2 (95% confidence interval 96.7 to 99.1) v 92.7 (88.5 to 95.3)), area under the curve (0.99 (0.98 to 1.00) v 0.93 (0.91 to 0.95)), positive likelihood ratio (9.1 (6.2 to 13.3) v 4.4 (3.3 to 6.0)), and negative likelihood ratio (0.02 (0.01 to 0.04) v 0.09 (0.06 to 0.15); (P<0.05)).

Conclusion Parameters for diagnostic performance significantly decrease if non-evaluable results are included by a 3×2 table for analysis (intention to diagnose approach). This approach provides a more realistic picture of the clinical potential of diagnostic tests.

Introduction
Clinical decisions in medicine are largely made on the basis of information gained from diagnostic testing. Against the background of more than 15 years of development and experience in evidence based medicine1 and in times of comparative effectiveness research,2 new diagnostic techniques have to be critically assessed and proven to be effective before they can be used on a wide scale. Diagnostic accuracy studies comparing an index test with a reference or gold standard and meta-analyses combining the results of many individual studies to explore a test’s diagnostic potential are an important and basic step in the overall evaluation process of the validity of a new diagnostic test.3 4 However, previous studies have shown that methodological deficits could affect the estimated diagnostic accuracy of a test.5 7

In non-invasive coronary imaging, technical innovations such as dual source8 and 320 row computed tomography (CT)9 have improved spatial and temporal resolution while reducing radiation. As a result of these developments, CT has evolved into the primary modality for non-invasively evaluating native coronary arteries over the past 10 years.10 11 Cardiac CT examinations performed on newer generation scanners (with at least 64 rows) have the potential to reliably rule out substantial stenoses in patients with a low to intermediate pretest likelihood, and thus can spare them an invasive catheterisation.10 12 Nonetheless, when exploring studies from this highly topical field of diagnostic imaging, we are confronted with a fundamental deficiency—non-evaluable results from the index test that are classifiable as neither positive nor negative. This problem has not yet been resolved adequately, although it has the most direct influence on diagnostic accuracy results. Coronary CT angiography studies commonly deal with
non-evaluable results of the index test in different ways, especially when transferring a segment based (or vessel based) to a patient based evaluation; non-evaluable segments are simply excluded, patients with non-evaluable segments are excluded, or patients with non-evaluable segments are generally considered either positive or negative. This exclusion leads to bias and overestimation of diagnostic accuracy at the study level, which is then introduced into meta-analyses pooling data from such studies.

In this article, therefore, we aimed to investigate how different approaches of dealing with non-evaluable results lead to variations in overall diagnostic accuracy values, using a systematically compiled pool of studies of non-invasive coronary CT angiography. We proposed an approach for the transparent reporting of such results—by applying a 3×2 table—to avoid biased overestimation of diagnostic accuracy.

**Methods**

We performed a systematic search for CT studies of coronary angiography on a patient level using recently reported methods. Briefly, we searched Medline (via PubMed), Embase (via Ovid), and ISI Web of Science databases. The main inclusion criteria were prospective study design, conventional coronary angiography as the reference standard in evaluating native coronary arteries, both tests performed in all patients, and CT scanners with at least 12 detector rows. The studies also had to provide results allowing calculation of per patient 2×2 tables for obstructive coronary artery disease (defined as at least one coronary stenosis of at least 50%) and had to be published in English or German. We excluded studies explicitly stated to be retrospective or if they potentially overlapped with other studies. The original meta-analysis has further methodological details.

The update search was performed on 2 February 2011. We then checked the full texts of the pool of relevant studies for the possibility to calculate an alternative 3×2 table on the per patient level—that is, giving adequate background information on single patients with coronary CT images of non-evaluable quality and the patient’s real health status defined by the invasive catheter examination (gold standard).

**Statistical analysis**

For the meta-analytical data evaluation, we used an exact binomial rendition of the bivariate, mixed effects regression model developed by van Houwelingen and colleagues and modified for synthesis of diagnostic test data. We calculated summary diagnostic performance values including 95% confidence intervals from standard data of a 2×2 table after excluding non-evaluable results or the 3×2 table, including non-evaluable results either in the “false negative” or the “false positive” cell of a 2×2 table (worst case scenario) according to the results of the reference standard (intention to diagnose principle). For visually illustrating and directly comparing these two approaches, we combined two summary receiver operating characteristics curves into one graph. We also evaluated two further common approaches, categorically declaring non-evaluable results as either positive or negative.

Figures 1 to 3 summarise the different approaches and their influence on sensitivity and specificity. The “classic” 2×2 table (fig 1) does not take into account non-evaluable results. Figure 2 shows the effects of excluding non-evaluable results or declaring them as either positive or negative. Figure 3 presents the 3×2 table, suggested for transparent reporting and for avoiding overestimation of sensitivity and specificity (with an intention to diagnose approach). We used the MIDAS module for Stata, version 11 (StataCorp), and Proc GLIMMIX in SAS, version 9.2 (SAS Institute), to perform the analysis.

**Results**

We found 120 eligible studies of coronary CT angiography (including 10 287 patients) that compared each patient’s results with conventional coronary angiography. We checked their full texts for the possibility to calculate an alternative 3×2 table. Eleven (11/120=0.092=9%) studies did not have non-evaluable results. Twenty six (22%) studies simply excluded non-evaluable segments, and 23 (19%) excluded patients with non-evaluable segments from analysis. Twenty six (22%) studies declared all patients with non-evaluable segments as positive, and seven (6%) declared them as negative. Three (3%) studies reported findings in a 3×2 table while retaining all non-evaluable results. For 24 (20%) studies, it remained unclear how non-evaluable segments were transferred to the patient level.

For 26 studies (including 2298 patients), it was possible to calculate alternative 3×2 tables (table 1). Sensitivity, area under the curve, and positive and negative likelihood ratios indicated significantly decreased diagnostic performance (P<0.05), compared with the 2×2 table (table 2).

Other approaches, which are sometimes also referred to as “intention to diagnose” but only declare patients with non-evaluable segments as either positive or negative, overestimated either sensitivity or specificity (table 2). Figure 4 compares the findings for the 2×2 table and the 3×2 table calculations combining summary receiver operating characteristics curves in one graph.

**Discussion**

Our analysis indicated a lack of consensus on how studies of non-invasive CT coronary angiography handle non-evaluable outcomes, and our meta-analytical examination shows how different yet common strategies can distort diagnostic accuracy results. The “classic” 2×2 table (fig 1) does not hold enough information to show the true range of possible results, and forces investigators to use one of the approaches in figure 2: by simply excluding non-evaluable results, sensitivity and specificity are artificially increased; and by declaring non-evaluable results as either positive or negative, either sensitivity or specificity is overestimated, and the absolute numbers of non-evaluable results are not accessible.

From a clinical perspective, patients with non-evaluable results will have to be further evaluated to rule out or confirm significant disease. Therefore, classifying these patients as positive and taking into account that they will need further investigation seems to be clinically appropriate in several scenarios. However, in relation to the true diagnostic capabilities of the test itself, such an approach might be misleading. Only by transforming the 2×2 table into a 3×2 table and reporting all results accordingly will researchers make study outcomes fully transparent. Furthermore, using an intention to diagnose principle (fig 3) for calculation ensures that both sensitivity and specificity are not overestimated. The range of possible outcomes for sensitivity and specificity between the two scenarios of declaring non-evaluable results as positive or negative represents the overall effect of non-evaluable results. But summarising these two scenarios with a conservative approach—that is, including non-evaluable results as false positive
positives and false negatives—seems to summarise the true clinical potential of the diagnostic test most adequately. Acknowledging the fact that the reference standard can also yield non-evaluable results, the 3×2 table could even be extended further to a 3×3 table, also transparently reporting these non-evaluable results of the reference standard. However, non-evaluable results of conventional coronary angiography were rare in our analysis (0.1%).

As early as 1987, Simel and colleagues proposed using the 3×2 table as the standard method for reporting absolute numbers of diagnostic test results. To our knowledge, they were the first to propose a solution to overcome the problem. To appraise diagnostic accuracy, they used new operational definitions of sensitivity and specificity, calculating them traditionally but including further test values (especially the overall test yield) necessary to account for non-evaluable results and to characterise the diagnostic test. Based on the 3×2 table, we systematically analysed different approaches of handling non-evaluable results on coronary CT angiography and found that this method significantly altered the results regarding diagnostic accuracy in a meta-analytical evaluation. By applying an intention to diagnose approach, we avoided the overestimation of sensitivity and specificity by including non-evaluable results in our calculation. We believe that this approach is most convenient, because the customary method of characterising diagnostic accuracy as a pair of sensitivity and specificity, without the need for further test yield values, is preserved.

In addition to the fact that there is no consensus on how to handle non-evaluable results, our full text evaluation showed a general lack of comprehensive reporting of results by coronary CT angiography studies. Thus, in 109 (109/120=0.908=91%) studies, the authors encountered non-evaluable results on the segment or vessel level, but for 24 (22%) of these studies, it remained unclear how, or if at all, the results were transferred to the patient level. Furthermore, only 26 (26/85=0.306=31%) of the remaining 85 studies provided enough background information to enable us to calculate alternative 3×2 tables. Only three (3/120=0.025=3%) studies from our pool originally reported non-evaluable results in a 3×2 table.

Our findings indicate a lack of awareness of these issues of poor reporting and inconsistent handling of non-evaluable results, which continues to persist; of the investigated 120 studies from the young field of non-invasive coronary CT angiography, 106 (88%) were published after 2005. This problem is probably not restricted to coronary CT angiography but could greatly affect diagnostic accuracy studies in general. Firstly, the methodological differences in handling non-evaluable results compromise the comparability of diagnostic accuracy data from different studies. Furthermore, the common approaches of handling non-evaluable results (fig 2) distort findings regarding diagnostic accuracy. This distortion affects not only sensitivity and specificity but also predictive values and likelihood ratios, which has important implications for clinical decision making. If we assume that the pool of studies investigated here is representative of all studies available on the topic, the overall potential of coronary CT angiography as a test to rule out significant stenoses is weakened, because the confidence interval of the negative likelihood ratio exceeded the value of 0.1 (0.09 to 0.15).13 Therefore, these biased results will probably also have an effect at higher levels of evidence, and will therefore affect the evaluation of new diagnostic technologies even more. This scenario could happen when such biased data are combined in systematic reviews and meta-analyses that constitute an important basis for health technology assessment reports, which influence decision and policy makers.

The STARD checklist (Standards for the Reporting of Diagnostic Accuracy Studies)14 and the QUADAS tool (Quality Assessment of Diagnostic Accuracy Studies Included in Systematic Reviews),15 the most common reporting guideline and assessment tool for diagnostic accuracy studies, do not pay enough attention to the problem of non-evaluable results. STARD item 19 postulates the reporting of results as “including indeterminate and missing” results, and STARD item 22 requests investigators to report on how these results were handled. Item 13 of the original QUADAS checklist asks “Were uninterpretable/intermediate results reported?”—despite recognising the topic, the question is restricted to the level of reporting. But the mere inclusion of “non-evaluable results as positives” does not necessarily mean that absolute numbers of non-evaluable results will be accessible to readers. In this regard, it is especially disappointing that QUADAS-2,142 the revised version of the original QUADAS tool,141 does not consider non-evaluable results explicitly. This situation underlines the need to find a consensus on how to report and integrate non-evaluable diagnostic test results in the future.

Although we believe that the issues discussed here are inherent to all diagnostic tests, our analysis is limited to a specific field of medical imaging. Therefore, the effect on diagnostic accuracy estimates in other medical fields should be evaluated in further research.

Responsible reporting is an essential component of research conduct.143 To improve the reliability and value of medical research by promoting transparent and accurate reporting, the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) network has been established.144 This international initiative aims to bring together all stakeholders with an interest in the improvement of publications and medical research, including authors, journal editors, and peer reviewers. Among other things, the network’s website (www.equator-network.org) offers a comprehensive online library with the available reporting guidelines.

Conclusions

For diagnostic accuracy studies in particular, complete reporting of all results on all levels (at the segment, vessel, and patient levels), if applicable, is the basis for fully characterising a test’s diagnostic potential. As a minimum, authors of diagnostic accuracy studies should adopt the STARD checklist14 to meet general reporting standards, and medical journals should consistently encourage (or even demand) the use of this checklist for submitting manuscripts on studies of diagnostic accuracy. Beyond that, a standardised approach for authors of diagnostic accuracy studies on coronary CT angiography (and on any other topic) could be to report their findings in a 3×2 table. If this approach is performed following intention to diagnose principles, all results become transparent to readers, and authors will be more cautious in interpreting an overly optimistic presentation of diagnostic test accuracy.

Funding: This study was supported by a grant of the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF; FKZ: 01KG1110) for meta-analyses as part of the joint programme “clinical trials” of the BMBF and the German Science Foundation (DFG). The supporting organisation had no involvement in the design, conduct, analysis, or manuscript preparation of this study.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request

No commercial reuse: See rights and reprints http://www.bmj.com/permissions
Subscribe: http://www.bmj.com/subscribe
What is already known on this topic

Diagnostic accuracy studies and meta-analyses of pooled results constitute an important step in the evaluation of diagnostic tests.

There is no consensus on how diagnostic accuracy studies should handle non-evaluable results, and common approaches to do so overestimate diagnostic accuracy.

What this study adds

In a pool of studies of non-invasive coronary CT angiography, we saw no consensus on how to handle non-evaluable results.

Common approaches of dealing with non-evaluable results led to significant differences in overall diagnostic accuracy estimates.

Transparent reporting of findings in a 3+2 table including non-evaluable results and applying an intention to diagnose approach can provide a more realistic picture of the clinical potential of diagnostic tests.

from the corresponding author) and declare: PS and MD are supported by a grant of the German Federal Ministry of Education and Research (BMBF) for meta-analyses as part of the joint programme “clinical trials” of the BMBF and the German Science Foundation (DFG); PS is also supported by another grant of the DFG (Schl 3-1) and has received lecture fees from Bayer-Scherling; MD has received grant support from Heisenberg Program of the German Research Foundation (DFG) for a professorship (DE 1361/14-1), European Regional Development Fund (2007/2013 2/05, 2007/2013 2/48), German Heart Foundation/German Foundation of Heart Research (F/2308, F/27/10), a joint programme from the German Research Foundation (DFG) and the German Federal Ministry of Education and Research (BMBF) for meta-analyses (01KG1013, 01KG1110), GE Healthcare, Bracco, Guerbet, and Toshiba Medical Systems; MD has received lecture fees from Toshiba Medical Systems, Guerbet, Cardiac MR Academy Berlin, and Bayer (Schering-Berlex); MD is a consultant to Guerbet and one of the principal investigators of multicentre studies (CORE-64 and 320) on coronary CT angiography sponsored by Toshiba Medical Systems; MD is the editor of Coronary CT Angiography and Cardiac CT, both published by Springer; MD gives lectures and offers hands-on workshops on cardiovascular imaging; institutional master research agreements for MD exist with Siemens Medical Solutions, Philips Medical Systems, and Toshiba Medical Systems; GMS is a physician working as a research assistant in MD’s working group, whose salary is financed by a BMBF grant for meta-analyses granted to MD; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval: Not required.

Contributors: MD had the initial idea for the manuscript; MD and GMS conceived and designed the study, MD, GMS, and PS were responsible for analysis and interpretation of data; GMS drafted and MD critically revised the article for important intellectual content; All authors approved the final manuscript; MD is the guarantor. The article was initiated by the authors, who had full control of the data analysis and interpretation, and there was no industry sponsorship.

Data sharing: No additional data available.
35 Dewey M, Tofig F, Schnappauf D, Laube M, Borges AC, Werneczki KD, et al. Noninvasive detection of coronary artery stenoses with multislice computed tomography or magnetic resonance imaging. *Ann Intern Med* 2006;445:107-15.

36 Ehrab M, Wittershagen M, Katon M, Malladi A, Terashima T, et al. Diagnostic accuracy of 64-slice computed tomography for detecting angiographically significant coronary artery stenoses in an unselected consecutive patient population: comparison to angiography. *Circulation* 2005;112:2177-81.

37 Erdogan N, Akar N, Vural M, Cambay A, Kayhan T, Sahin D, et al. Diagnostic value of 16-slice multidetector computed tomography in symptomatic patients with suspected obstructive coronary artery disease. *Heart* 2005;91:2179-84.

38 Gaudio C, Marfelli F, Pollicino C, Francoise M, Tantigli G, Mielleche S, et al. Early detection of coronary artery disease by 64-slice multidetector computed tomography in asymptomatic hypertensive high-risk patients. *J Cardiovasc Comput Tomogr* 2009;3:250-6.

39 Ghersin E, Limanov D, Dragu R, Rispler S, Lessjo J, Ofer A, et al. 16-MDCT coronary angiography versus invasive coronary angiography in acute chest pain syndromes. In: *J Am Coll Cardiol* 2007;52:2090-4.

40 Gilard M, Comly JC, Pennie PY, Joret C, Le Gal G, Mansorati J, et al. Accuracy of multislice computed tomography in the preoperative assessment of coronary artery disease in patients referred for cardiac valve surgery. *J Am Coll Cardiol* 2006;47:2001-8.

41 Gross S, Globits S, Hergan K. Forty-slice spiral computed tomography of the coronary arteries: assessment of image quality and diagnostic accuracy in a non-selected patient population. *Radiology* 2006;239:714-22.

42 Hacker M, Jakobs T, Hack N, Nicolaou K, Becker C, von Ziegler F, et al. Combined use of 64-slice computed tomography angiography and gated myocardial perfusion SPECT for the detection of angiographically relevant coronary artery stenoses. First results in a clinical setting concerning patients with stable angina. *Nuklearmedizin* 2007;46:23-93.

43 Halon DA, Gaspar T, Atalay S, Rubinstein R, Schlammier J, Peled N, et al. Use and limitations of 16-multidetector computed tomography for diagnosing coronary lesions in unselected patients referred for routine invasive coronary angiography. *Circulation* 2007;208:209-21.

44 Hausleiter J, Freiberger W, Dombrowsky T, Manki Z, Geier P, Donner K, et al. Non-invasive computed tomographic angiography for patients with suspected coronary artery disease: The Coronary Angiography By Computed Tomography With the Use of a Submillimetric Radiation Dose (CACTUS) trial. *Eur Heart J* 2008;29:3034-41.

45 Hennemann MM, Schuff JD, Jukwiek JW, Lam-Huj, S, D(per) A, Dibbs PE, et al. Comprehensive cardiac assessment with multislice computed tomography evaluation of left ventricular function in addition to coronary anatomy in patients with previous myocardial infarction. *Heart* 2008;92:1779-83.

46 Hennemann MM, Schuff JD, Pundtze G, van Werkhoven JM, van der Wal EE, Jukwiek JW, et al. Noninvasive evaluation of multislice computed tomography in suspected acute coronary syndromes. A single-center study with 64-slice computed tomography versus coronary calcium score. *J Am Coll Cardiol* 2008;52:216-22.

47 Herzog EA, Huxsoll L, Burkhard N, Guapier O, Valerita T, Tatsugami F, et al. Accuracy of low-dose multislice computed tomography angiography for detecting small coronary artery disease in symptomatic patients. *Eur Heart J* 2008;30:3307-12.

48 Herzog C, Nguyen SA, Savino G, Zwerner PL, Doll J, Nielsen CD, et al. Does two-segment image reconstruction improve image quality and diagnostic accuracy? *Radiology* 2007;244:121-9.

49 Herzog C, Zwemer PL, Doll JP, Nielsen CD, Aygun SA, Savino G, et al. Significant coronary artery stenoses: a direct comparison of patient-based and per-patient or per-segment basis at 64-slice CT coronary angiography. *Radiology* 2007;244:121-20.

50 Hoffmann MH, Sih H, Schmitz BL, Schmitz FT, Liebichm K, Schulze R, et al. Noninvasive coronary angiography with multislice computed tomography. *JAMA* 2005;293:2471-8.

51 Hoffmann U, Mosswelk F, Curry RC, Fieremak M, Jiang H, Diaz LJ, et al. Predictive value of 64-slice computed tomography for detection of significant coronary artery disease in patients at high risk for the disease: patient-based analysis. *Circulation* 2004;110:3689-97.

52 Johnson TR, Heuser B, Siebel AW, Becker C, Leber AW, Rosenberg B, et al. Diagnostic accuracy of dual-source computed tomography of the diagnosis of coronary artery disease. *Invest Radiol* 2007;42:884-91.

53 Kaiser C, Bredenbeck GH, El-Brunner-LaRoca HP, Bongartz G, Pfeiffer M, et al. Limited diagnostic yield of non-invasive coronary angiography by 16-slice detector spiral computed tomography in routine patients referred for evaluation of coronary artery disease. *Radiology* 2007;245:969-72.

54 Keller J, Coche A, Legros O, Pasquet A, Grandin C, Van Der Bees BE, et al. Head-to-head comparison of three-dimensional navigator-gated magnetic resonance imaging and 16-slice computed tomography to detect coronary artery stenosis in patients. *Am J Cardiol* 2005;96:90-100.

55 Kilias K, Velle CH, Haireis S, Heigom K, Wissh M. Multislice computed tomography coronary angiography for diagnostic coronary artery disease and 4D analysis in patients with unstable angina pectoris. *Heart* 2007;93:2143-9.

56 Kilias K, Velle CH, Haireis S, Heigom K, Wissh M. Multislice computed tomography coronary angiography for diagnostic coronary artery disease and 4D analysis in patients with unstable angina pectoris. *Heart* 2007;93:2143-9.

57 Kilias K, Velle CH, Haireis S, Heigom K, Wissh M. Multislice computed tomography coronary angiography for diagnostic coronary artery disease and 4D analysis in patients with unstable angina pectoris. *Heart* 2007;93:2143-9.

58 Knott C, Bredenbeck GH, El-Brunner-LaRoca HP, Bongartz G, Pfeiffer M, et al. Limited diagnostic yield of non-invasive coronary angiography by 16-slice detector spiral computed tomography in routine patients referred for evaluation of coronary artery disease. *Radiology* 2007;245:969-72.

59 Keller J, Coche A, Legros O, Pasquet A, Grandin C, Van Der Bees BE, et al. Head-to-head comparison of three-dimensional navigator-gated magnetic resonance imaging and 16-slice computed tomography to detect coronary artery stenosis in patients. *Am J Cardiol* 2005;96:90-100.

60 Kilias K, Velle CH, Haireis S, Heigom K, Wissh M. Multislice computed tomography coronary angiography for diagnostic coronary artery disease and 4D analysis in patients with unstable angina pectoris. *Heart* 2007;93:2143-9.

61 Kilias K, Velle CH, Haireis S, Heigom K, Wissh M. Multislice computed tomography coronary angiography for diagnostic coronary artery disease and 4D analysis in patients with unstable angina pectoris. *Heart* 2007;93:2143-9.

62 Knott C, Bredenbeck GH, El-Brunner-LaRoca HP, Bongartz G, Pfeiffer M, et al. Limited diagnostic yield of non-invasive coronary angiography by 16-slice detector spiral computed tomography in routine patients referred for evaluation of coronary artery disease. *Radiology* 2007;245:969-72.

63 Keller J, Coche A, Legros O, Pasquet A, Grandin C, Van Der Bees BE, et al. Head-to-head comparison of three-dimensional navigator-gated magnetic resonance imaging and 16-slice computed tomography to detect coronary artery stenosis in patients. *Am J Cardiol* 2005;96:90-100.
Ropers D, Baum U, Pohle K, Andersen K, Lethiser S, Ohnohara B, et al. Detection of coronary artery stenoses with thin-slice multi-detector row spiral computed tomography and multiparametric reconstruction. Circulation 2003;107:694-6.

Ropers D, Ropers A, Andersen K, Katus A, Lauterburg A, Baurwinkel W, et al. Usefulness of multidetector row spiral computed tomography with 64 × 0.6-mm collimation and 330-ms rotation for the noninvasive detection of significant coronary artery stenoses. Am J Cardio 2007;100:701-6.

Ropers U, Ropers D, Pfitzner T, Andersen K, Kuetten A, Stilmanek N, et al. Influence of heart rate on the diagnostic accuracy of dual-source coronary computed tomography angiography. J Am Coll Cardiol 2007;50:2395-8.

Scheffel H, Ackla R, Leschka S, Plass A, Desloches L, Guber I, et al. Low-dose CT coronary angiography in the step-and-stop mode: Diagnostic performance. Heart 2008;94:1132-7.

Scheffel H, Ackla R, Plass A, Vachenauer R, Desloches L, Gaemperli O, et al. Accuracy of dual-source CT coronary angiography: First experience in a high-pre-test probability population without heart rate control. Eur Radiol 2006;16:2739-47.

Scheffel H, Leschka S, Plass A, Vachenauer R, Gaemperli O, Ganzel E, et al. Accuracy of 64-slice computed tomography for the preoperative determination of coronary artery disease in patients with chronic aortic regurgitation. Am J Cardio 2007;100:701-6.

Schuijf JD, Bundschu G, Jukema JW, Lamberg HJ, van der Hoeven BL, de Roos A, et al. Diagnostic accuracy of 64-slice multislice computed tomography in the noninvasive evaluation of significant coronary artery disease. Eur J Radiol 2008;68:145-8.

Shabestari AB, Sadi A, Akhlaghebpoor S, Aradi M, Baharoudi J, Pajouh MD, et al. Diagnostic performance of 64-channel multislice computed tomography in assessment of significant coronary artery disease in symptomatic subjects. Am J Cardio 2007;99:1566-71.

Stolzmann P, Scheffel H, Leschka S, Plass A, Baumerter S, Marincak B, et al. Influence of calcifications on diagnostic accuracy of coronary CT angiography using prospective ECG triggering. AJR Am J Roentgenol 2008;191:1684-9.

Tsy XJ, Lee T, Lee W, Tsao CR, Tsai WL, Chen MC, et al. Use of 64-detector row computed tomography before catheter coronary arteriography to select early versus late invasive treatment for patients with low-risk acute coronary syndrome. J Comput Assist Tomogr 2007;31:258-64.

Turkovatian A, Biykoglu SF, Buyukbayraktar F, Cilbas T, Cumhur T, Duruc E. Clinical value of 64-row CT coronary calcium volumetric analysis in patients with suspected coronary disease. Acta Radiol 2008;49:490-8.

Ullmer OR, Gjonnaes A, Ekebom A, Dahlin T, Stranden E, Sandback G. Noninvasive coronary angiography with 64-channel multidetector computed tomography in patients with acute coronary syndrome. Acta Radiol 2008;49:1140-4.

Watts MJ, Hesse B, Green CE, Greenberg NL, Manning M, Chauudy E, et al. Detection of coronary artery stenosis using 64-channel computed tomography with multi-segment reconstruction. Am J Cardio 2007;99:175-81.

Weustink AC, Meijboom WB, Mollet NR, Otsuka M, Pugliese F, van Mieghem C, et al. Diagnostic accuracy of 64-slice computed tomographic coronary angiography in patients with non-ST-elevation acute coronary syndrome. Radiololgia Medica 2010;115:341-53.

Xia S, Arazo T, Hikita H, Miyazaki S, Takahashi Y, Kuhara T, et al. Incremental value of combining 64-slice computed tomographic angiography with stress nuclear myocardial perfusion imaging to improve noninvasive detection of coronary artery disease. J Nucl Cardio 2010;17:19-26.

XieM, Ben-Nakhi A, Shukkur AM, Sin T, Al-Rashid A. Accuracy of 64-detector row computed tomography in the diagnosis of coronary artery disease. Med Prin Pract 2009;18:923-8.

Xu L, Yang L, Fan Z, Yu W, Li B, Zhang Z. Diagnostic performance of 320-detector CT coronary angiography in patients with atrial fibrillation: preliminary results. Eur Radiol 2011;21:936-40.

Yang L, Zhang Z, Fan Z, XU C, Zhao L, Yu W, et al. 64-DMCT coronary angiography of patients with atrial fibrillation: Influence of heart rate on image quality and efficacy in evaluation of coronary artery disease. J Cardiovasc Comput Tomogr 2008;12:173-81.

Zhang Li, Wu SY, Wang J, Zhang ZL, Jiang SS, et al. Diagnostic accuracy of dual-source CT coronary angiography: The effect of average heart rate, heart rate variability, and cine score on a clinical comparison. J Cardiovasc Comput Tomogr 2010;5:717-20.

Simel DL, Feussner JR, DeLong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. Med Decis Making 1987:17:104-18.

Siegel ME, Litaker J, Goldenberg M, Kuhl H, Sahn DJ, etc. Diagnostic performance of 320-detector coronary computed tomography angiography: comparison with intravascular ultrasound and angiography. Can J Cardio 2005;29:641-7.

van Herwerden JM, Heeck MW, Schuijf JD, Jukema JW, Boogaard MA, van der Wall EE, et al. Diagnostic accuracy of 64-slice multislice computed tomographic coronary angiography in patients with an intermediate pretest likelihood for coronary artery disease. Am J Cardio 2009;5:1052-7.

Xu L, Yang L, Fan Z, Yu W, Li B, Zhang Z. Diagnostic performance of 320-detector CT coronary angiography in patients with atrial fibrillation: preliminary results. Eur Radiol 2011;21:936-40.

Yang L, Zhang Z, Fan Z, Xu C, Zhao L, Yu W, et al. 64-DMCT coronary angiography of patients with atrial fibrillation: Influence of heart rate on image quality and efficacy in evaluation of coronary artery disease. J Cardiovasc Comput Tomogr 2008;12:173-81.

Zhang Li, Wu SY, Wang J, Li B, Zhang ZL, Jiang SS, et al. Diagnostic accuracy of dual-source CT coronary angiography: The effect of average heart rate, heart rate variability, and cine score on a clinical comparison. J Cardiovasc Comput Tomogr 2010;5:717-20.

Simel DL, Feussner JR, Delong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. Med Decis Making 1987:17:104-18.

Deeks JJ, Altman DG, Diagnostic test accuracy. New York: Oxford Press, 2003:325-27.

Boissuy PM, Reitsma JB, Bruns DE, Gatschke CA, GLausing GP, Iwmg LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2005;33:41-4.

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Klovning L. The development of QUADAS: a tool for the validation of diagnostic accuracy studies. BMJ 2003;326:387-9.

Deeks JJ, Altman DG. Diagnostic test accuracy. New York: Oxford Press, 2003:325-27.

Boissuy PM, Reitsma JB, Bruns DE, Gatschke CA, GLausing GP, Iwmg LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2005;33:41-4.

Whiting P, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.

Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. Lancet 2011;377:1149-50.
## Tables

### Table 1 Analysed studies with recalculated 3×2 tables including non-evaluable results

| First author | Journal | Year | Handling of non-evaluable segments at patient level | 2×2 table | Non-evaluable results (CT) at patient level |
|--------------|---------|------|---------------------------------------------------|-----------|---------------------------------------------|
|              |         |      |                                                   | True positive | False positive | False negative | True negative | Non-evaluable (positive) | Non-evaluable (negative) |
| Achenbach S  | Eur Heart J | 2005 | Patients considered to be negative | 25          | 4          | 0         | 19          | 2             | 0             |
| Alkadhi H    | Eur Heart J | 2008 | Patients considered to be positive | 57          | 3          | 2         | 79          | 0             | 9             |
| Bonmassari R | J Cardiovasc Med | 2006 | Patients excluded | 12          | 2          | 0         | 8           | 2             | 9             |
| Boulmier D   | Arch Cardiovasc Dis | 2009 | Patients considered to be positive | 12          | 4          | 0         | 37          | 0             | 6             |
| Cademartiri F | Radiol Med | 2009 | Patients excluded | 31          | 1          | 0         | 9           | 0             | 1             |
| Carrascosa P | J Cardiovasc Comput Tomogr | 2010 | Patients considered to be positive | 26          | 3          | 0         | 18          | 0             | 3             |
| Comil JC     | Eur J Radiol | 2007 | Patients excluded | 9           | 1          | 0         | 23          | 1             | 2             |
| de Graaf FR  | Eur Heart J | 2010 | Patients considered to be positive | 35          | 3          | 0         | 22          | 2             | 2             |
| Dewey M      | Ann Intern Med | 2006 | Intention to diagnose (3×2 table) | 62          | 5          | 4         | 46          | 1             | 11            |
| Dewey M      | Circulation | 2009 | Intention to diagnose (3×2 table) | 11          | 1          | 0         | 17          | 0             | 0             |
| Garcia MJ    | JAMA      | 2006 | Patients considered to be positive | 44          | 29         | 1         | 70          | 14            | 29            |
| Gilard M     | J Am Coll Cardiol | 2006 | Patients considered to be positive | 6           | 0          | 0         | 35          | 5             | 9             |
| Hausleiter J | Eur Heart J | 2007 | Patients considered to be positive | 91          | 19         | 1         | 106         | 10            | 16            |
| Herzog BA    | Eur Heart J | 2008 | Patients considered to be positive | 16          | 0          | 0         | 10          | 2             | 2             |
| Jenkins SMM  | Q J Med   | 2011 | Patients considered to be positive, 3×2 table given | 27          | 8          | 3         | 29          | 8             | 24            |
| Leschka S    | AJR Am J Roentgenol | 2008 | Patients considered to be positive | 69          | 6          | 2         | 35          | 0             | 2             |
| Leschka S    | Heart     | 2008 | Patients considered to be positive | 35          | 1          | 1         | 33          | 0             | 4             |
| Leschka S    | Eur Radiol | 2009 | Patients considered to be positive | 14          | 1          | 0         | 19          | 0             | 1             |
| Mir-Akbari H | Acta Radiol | 2009 | Patients considered to be positive | 41          | 11         | 0         | 20          | 10            | 19            |
| Ovrehus KA   | Am J Cardiol | 2010 | Patients considered to be positive | 26          | 11         | 1         | 57          | 2             | 3             |
| Pontone G    | J Am Coll Cardiol | 2009 | Patients considered to be positive | 142         | 2          | 2         | 14          | 0             | 0             |
| Ropers U     | J Am Coll Cardiol | 2007 | Patients considered to be positive | 39          | 10         | 1         | 47          | 2             | 1             |
| Sato A       | J Nucl Cardiol | 2010 | Patients excluded | 61          | 8          | 1         | 37          | 15            | 8             |
| Scheffel H   | Am J Cardiol | 2007 | Patients considered to be positive | 12          | 0          | 0         | 35          | 1             | 2             |
| Scheffel H   | Heart     | 2008 | Patients considered to be positive | 64          | 2          | 0         | 50          | 2             | 2             |
| Ulimoen GR   | Acta Radiol | 2008 | Non-evaluable segments excluded | 32          | 6          | 4         | 6           | 10            | 0             |

Naive summary 3x2 table values that do not represent the statistical model: 999, 141, 23, 881, 89, 165
| Basis for calculation                          | Mean sensitivity (95% CI) | Mean specificity (95% CI) | Area under the curve (95% CI) | Positive likelihood ratio (95% CI) | Negative likelihood ratio (95% CI) |
|-----------------------------------------------|---------------------------|---------------------------|-------------------------------|-----------------------------------|----------------------------------|
| 2×2 table (non-evaluable results excluded)    | 98.2 (96.7 to 99.1)*      | 89.2 (84.2 to 92.8)       | 0.99 (0.98 to 1.00)*          | 9.1 (6.2 to 13.3)*                | 0.02 (0.01 to 0.04)*             |
| Non-evaluable patients considered as positive | 98.3 (96.9 to 99.0)*      | 78.4 (71.6 to 84.0)       | 0.98 (0.97 to 0.99)*          | 4.5 (3.5 to 6.0)                  | 0.02 (0.01 to 0.04)*             |
| Non-evaluable patients considered as negative | 92.9 (88.8 to 95.5)       | 90.5 (86.8 to 93.2)*      | 0.96 (0.94 to 0.98)           | 9.8 (7.0 to 13.7)*                | 0.08 (0.05 to 0.13)              |
| 3×2 table (intention to diagnose approach)    | 92.7 (88.5 to 95.3)       | 79.0 (72.3 to 84.4)       | 0.93 (0.91 to 0.95)           | 4.4 (3.3 to 6.0)                  | 0.09 (0.06 to 0.15)              |

*Significantly different (P<0.05) from data obtained by use of 3×2 table.
Figures

Fig 1 The “classic” 2×2 table, calculation of sensitivity and specificity

| Index test | Reference standard |
|------------|--------------------|
| +          | -                  |
| *          | True positive      | False positive |
| False negative | False negative  |

Sensitivity = true positive/(true positive + false negative)
Specificity = true negative/(true negative + false positive)

Fig 2 Different methods of handling non-evaluable results

| Excluding non-evaluable results | Reference standard |
|---------------------------------|--------------------|
| +                               | -                  |
| *                               | True positive      | False positive |
| Non-evaluable                   | Non-evaluable (positive) | Non-evaluable (negative) |
| False negative                  | True negative      |

Sensitivity ↑, specificity ↓

| Non-evaluable results considered to be positive | Reference standard |
|------------------------------------------------|--------------------|
| +                                               | -                  |
| *                                               | True positive      | False positive |
| Non-evaluable (positive)                        | Non-evaluable (negative) |
| False negative                                  | True negative      |

Sensitivity ↑, specificity ↓

| Non-evaluable results considered to be negative | Reference standard |
|------------------------------------------------|--------------------|
| +                                               | -                  |
| *                                               | True positive      | False positive |
| Non-evaluable (positive)                        | Non-evaluable (negative) |
| False negative                                  | True negative      |

Sensitivity ↓, specificity ↓

Fig 3 3×2 table and intention to diagnose principle
Fig 4 Summary receiver operating characteristics curves for 2×2 and 3×2 tables. The graph shows summary receiver operating characteristics curves using pairs of sensitivity and specificity of the 26 studies that provided enough background information to construct 3×2 tables. The upper left curve is based on the results of the studies when excluding non-evaluable results (2×2 table), the lower right curve when including them as either false positives or false negatives according to the results of the reference standard (3×2 table with an intention to diagnose approach). Curves include a summary operating point for sensitivity and specificity on the curve and a 95% confidence contour ellipsoid.