Diagnostic accuracy of computed tomography coronary angiography in patients with high heart rates: a systematic review

Thesis prepared by

Gordon T.W. Mander

in fulfilment of the degree of Master of Clinical Science,

Joanna Briggs Institute, Faculty of Health Science

University of Adelaide

17th October 2019

Supervisors:

A/Prof Zachary Munn, PhD
Associate Professor
The Joanna Briggs Institute
School of Public Health
University of Adelaide

Dr Karen Dobeli, PhD
Radiographer Consultant, CT Team Leader
Royal Brisbane and Women's Hospital
Metro North Health Service
Queensland Health

Table of Contents

Table of Figures ..... ***vi***

Tables ..... ***vii***

Thesis Declaration ..... ***viii***

Acknowledgements and Dedication ..... ***ix***

List of Abbreviations ..... ***x***

Abstract ..... ***xi***

Introduction ..... **xi**

Methods ..... **xi**

Results ..... **xi**

Conclusion ..... **xii**

Implications for Practice ..... **xii**

Implications for Future Research ..... **xii**

Chapter One: Preamble ***1***

Introduction **1**

Studies of Diagnostic Test Accuracy 1

Common Diagnostic Test Accuracy Study Designs 8

Systematic Reviews **10**

Systematic Reviews of Diagnostic Test Accuracy 11

Risk of Bias and the QUADAS-2 Critical Appraisal Instrument 11

Data Analysis 15

Chapter Two: Background and Rationale ***17***

Background **17**

Coronary Artery Disease **17**

Significant Coronary Artery Stenosis 18

Computed Tomography Coronary Angiography **18**

Background 18

Performance of CTCA 19

Importance of Heart Rate in CTCA **20**

Temporal Resolution 20
Models of CT Scanners 20
Reference Standard - Invasive Coronary Angiography 22
Rationale 23
Chapter Three: Methods 25
Introduction 25
Review Question 25
Inclusion Criteria 25
Types of Studies 25
Participants 25
Index Test 26
Reference Standard 26
Diagnosis of Interest 26
Search Strategy 27
Assessment of Methodological Quality 28
QUADAS-2 Risk of Bias Signalling Questions 28
QUADAS-2 Concerns Regarding Applicability 31
Data Collection 32
Data Synthesis 32
Narrative Synthesis 32
Meta-Analysis 32
Summary of Findings 34
Chapter Four: Results 35
Introduction 35
Search Results 35
Characteristics of Included Studies 37
Participants 38
Index Test 40
Reference Standard 41
Diagnosis of Interest 41
Assessment of Methodological Quality 42
Risk of Bias 42

Concerns Regarding Applicability 44
Potential Conflicts of Interest 45
Findings of Review 45
Narrative Synthesis 45
Meta-Analysis 51
Summary of Findings 72
Chapter Five: Discussion 74
Introduction 74
Interpretation of Results 74
Comparison to Existing Literature 76
Vessel-Level Analysis 77
Diagnostic Threshold 79
Non-Diagnostic Segments 79
Subgroup Analyses 80
Heart Rate 80
Scanner Make and Model 80
Limitations of the Review 81
Heterogeneity 82
Publication Bias 83
Deviation from Protocol 83
Chapter Six: Conclusions 84
Summary of Findings 84
Implications for Practice 85
Implications for Research 86
Future Research 86
Conclusion 87
References 88
Appendices i

Appendix One: Search Strategy i

PubMed Search String i

CINAHL Search String i

Embase Search String ii

Scopus Search String ii

Appendix Two: Details of Studies Excluded at Full Text i

Appendix Three: Data Extracted from Primary Studies i

Part i - Data Extracted for Participants i

Part ii - Data Extracted for Index Test (CT Coronary Angiography) ii

Part iii - Data Extracted for Reference Standard (Coronary Angiography) iii

Part iv - Data Extracted for Diagnosis of Interest (Coronary Artery Stenosis) iv

Table of Figures

Figure 1: Example ROC with high level accuracy 6
Figure 2: Example ROC with modest accuracy 6
Figure 3: Example ROC showing an uninformative test 7
Figure 4: Example ROC with similar AUC but differing sensitivity and specificity 8
Figure 5: Diagnostic case-control study design 9
Figure 6: Observational cohort study design 10
Figure 7: Search results 36
Figure 8: QUADAS-2 Risk of bias assessment 42
Figure 9: QUADAS-2 Risk of bias signalling questions for each included study 44
Figure 10: QUADAS-2 Concerns regarding applicability of included studies 45
Figure 11: Paired forest plot of diagnostic accuracy at per-patient level 52
Figure 12: Paired forest plot of diagnostic accuracy at per-vessel level 55
Figure 13: Paired forest plot of diagnostic accuracy at per-segment level 57
Figure 14: Sensitivity analysis at segment level with indeterminate segments treated as positive 60
Figure 15: Sensitivity analysis at segment level with indeterminate segments excluded from analysis 62
Figure 16: Range of heart rates for high heart rate groups in each study 64
Figure 17: Paired forest plot subgroup analysis of minimum heart rate thresholds 65
Figure 18: Per-segment subgroup analysis of minimum heart rate thresholds 67
Figure 19: Per-patient subgroup analysis of scanner make and model 69
Figure 20: Segment-level subgroup analysis of individual manufacturer scanners 71
Figure 21: Hypothetical scenario based on review findings 75

Tables

Table 1: 2x2 Table of diagnostic test accuracy 3
Table 2: Heart rates for participants included in diagnostic accuracy assessment.....39
Table 3: Reported sensitivity and specificity by each included study 50
Table 4: Number included in analysis with disease prevalence 51
Table 5: Statistical weight for each study reporting data at patient level 53
Table 6: Statistical weight for each study reporting accuracy at vessel level 54
Table 7: Statistical weight for each study reporting accuracy at segment level 56
Table 8: Statistical weight for each study reporting accuracy at segment level (non-evaluable segments treated as positive).....59
Table 9: Statistical weight for each included study reporting accuracy at segment level (non-evaluable treated as excluded).....61
Table 10: GRADE summary of findings table 73
Table 11: Previous systematic reviews reporting diagnostic accuracy in CTCA 77

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Gordon T.W. Mander

Date: 17/10/2019

Acknowledgements and Dedication

I gratefully acknowledge the assistance and support I received from my primary Supervisor, Associate Professor Zachary Munn, and the assistance from all staff at the Joanna Briggs Institute.

I also wish to gratefully acknowledge the extensive time commitment and ongoing support from my Co-Supervisor Dr Karen Dobeli.

Acknowledgement should also go to the assistance received from Ms. Caitlin Steffensen as the second reviewer responsible for co-reviewing the literature search and critical appraisal for the systematic review.

I also wish to acknowledge the assistance of Jane Smith in copyediting the thesis prior to submission in accordance with guidelines for the editing of research theses by professional editors and the Australian standards for editing practice.

The Australian Society of Medical Imaging and Radiation Therapy (ASMIRT) is acknowledged for financial assistance in the form of a 2018 ASMIRT Research Scholarship Grant ($10,000).

Finally, I wish to acknowledge the endless support from my beautiful and patient wife Allison and my gorgeous sons Cameron and Rory.

This thesis is dedicated to the loving memory of my mother, Alison Margaret Mander (1951-2012): Teacher, academic, and lifelong student.

List of Abbreviations

AHA	American Heart Association
AUC	Area Under the Curve
bpm	beats per minute
CAD	Coronary Artery Disease
CI	Confidence Interval
CT	Computed Tomography
CTCA	Computed Tomography Coronary Angiography
DLP	Dose Length Product (milliGray centimetres)
DSCT	Dual-Source Computed Tomography
ECG	Electrocardiogram
FN	False Negative
FP	False Positive
GRADE	Grades of Recommendation, Assessment, Development and Evaluation
HR	Heart Rate
ICA	Invasive Coronary Angiography
NPV	Negative Predictive Value
PPV	Positive Predictive Value
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2 [revised tool]
RCT	Randomised Controlled Trial
ROC	Receiver Operating Characteristic
SCCT	Society of Cardiovascular Computed Tomography
SROC	Summary Receiver Operating Characteristic
STARD	Standards for Reporting Diagnostic Accuracy Studies
TN	True Negative
TP	True Positive

Abstract

Introduction

An understanding of the diagnostic accuracy of computed tomography coronary angiography (CTCA) is critical for clinicians and guideline developers to determine the appropriate use and position of the scan in the diagnostic pathway. Current imaging guidelines, which are based on evidence from non-contemporary CT technology, only recommend the routine use of CTCA in patients with low heart rates (HR). The aim of this systematic review was to investigate the diagnostic accuracy of CT coronary angiography using state-of-the-art scanner technology, in comparison with invasive coronary angiography, for patients with high HRs.

Methods

Methods for the systematic review were determined a priori, based on a previously published protocol. A systematic search of PubMed, CINAHL, Embase and Scopus was performed as well as a search of unpublished sources and reference lists. Titles and abstracts were screened by two independent reviewers. Full-text screening was then performed on all studies that met the criteria for inclusion in the systematic review at the title and abstract level. Studies were included that described diagnostic accuracy metrics in patients with high HR. Studies that did not compare CTCA to invasive coronary angiography were excluded. Only current generation scanners with greater than 128 detectors were included. Included studies underwent critical appraisal using the QUADAS-2 tool. All critically appraised studies were then included in the final review regardless of methodological quality.

Data extraction was then undertaken and the results were collated and analysed through narrative synthesis and a diagnostic test accuracy meta-analysis.

Results

Twelve studies were included in the systematic review; 11 of these studies were also included in a diagnostic test accuracy meta-analysis. Meta-analysis indicated high level pooled sensitivity 99% (95% CI: 97%,100%) in CTCA at high HR. Pooled specificity was lower

at 79% (95% CI: 72%, 85%). Diagnostic accuracy performed better at artery level (pooled sensitivity 96% (95% CI: 93%, 97%) and pooled specificity 93% (95% CI: 90%, 96%)); and segment level (pooled sensitivity 91% (95% CI: 88%, 93%) and pooled specificity 96% (95% CI 95%, 98%)). The prevalence of clinically significant coronary artery disease was high in each of the included studies.

There were insufficient data to effectively evaluate the accuracy of CTCA at individual HRs. No significant difference was evident between different CT makes and models included in the review in terms of diagnostic accuracy.

Conclusion

Diagnostic sensitivity of CTCA is high at elevated HRs. Consequently, CTCA can still be performed when standard HR control is contraindicated or ineffective as it is an effective test to rule out coronary artery disease. However, the modest results for sensitivity indicate a positive result should be assessed with caution.

Implications for Practice

CT scanning is still appropriate in patients with high HRs when contraindications to HR lowering medications exist or are ineffective and when the CTCA is performed in order to rule out rather than quantify coronary artery stenosis.

Implications for Future Research

Further research is required to better understand the effect high HRs have on important patient outcomes, such as over-testing and anxiety related to false negative results.

Chapter One: Preamble

Introduction

A detailed understanding of diagnostic test accuracy is vital for diagnosticians, guideline developers and researchers alike. The diagnostic capability of a test should directly inform the appropriateness of the use of that test in clinical practice.[1] Specifically, diagnostic test accuracy studies can be used to establish the accuracy of a new test, compare accuracy between tests, or rank them based on their characteristic ability to rule in or out a condition of interest. Diagnostic test accuracy is a particularly important consideration in medical imaging studies.[2] The value of a particular test to visualise or otherwise detect a diagnosis of interest is a key element of imaging. Whilst imaging studies have a number of uses, their primary purpose is to aid in diagnosis.

Primary studies of diagnostic test accuracy use a distinctive study design. They are generally prospective designs and follow consecutive cohorts.

This chapter describes commonly encountered diagnostic test accuracy study designs, and describes the basic measurements and biases associated with primary studies of diagnostic accuracy. The purpose is to inform the reader of the methodology that is encountered in the main text of the systematic review. The importance of systematic review and meta-analysis is then discussed, particularly with respect to how systematic reviews of diagnostic test accuracy studies differ from systematic reviews of treatment effects.[3]

Studies of Diagnostic Test Accuracy

Diagnosis is the categorisation of a patient's signs and symptoms based upon predefined criteria. A diagnosis is often assisted by diagnostic testing. For a test to be useful, it must reliably provide a result typically representative in the diagnosis of interest. The reliability of a test to provide an appropriate result is referred to as diagnostic test accuracy.

In order to identify the diagnostic accuracy of a test, the test of interest, or index test, must be compared against an agreed-upon reference standard. The reference standard is a test or outcome sign considered most likely to identify the diagnosis of interest. All clinical tests must be compared against a reference standard in order to ascertain the accuracy of the test in practice. As a consequence, the index test can never achieve greater accuracy than the reference standard.

It should be noted that diagnostic test accuracy studies are focused solely on identifying the validity of an index test at detecting a diagnosis of interest.[1] These studies do not directly describe clinical pathways to outcomes such as the significance of a test result on treatment effectiveness.[4]

Measures of a disease state are usually considered dichotomously; that is, the importance of the outcome relates to the presence or absence of disease. Whilst in clinical usage this is not always the case, for two reasons. Certain disease states are not strictly dichotomous, as presence of disease often sits on a spectrum of severity, from non-significant to severe. Secondly, a test may signify a range of results, not simply positive and negative. Therefore a threshold or cut-off value is used to indicate precisely whether or not a particular test has value in determining a diagnosis of interest.

The value of a diagnostic imaging test can be described both in terms of how well the test is able to determine pathology when it exists, as well as how well the test can determine an absence of pathology. Diagnostic test accuracy studies are those that specifically describe the key metrics associated with diagnostic accuracy. That is, true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). These four parameters make up the fundamental building blocks for all calculations used for the various metrics of diagnostic accuracy and are usually portrayed in a 2x2 table (as shown in Table 1 below).[5]

The four variables of diagnostic accuracy can be defined as:

- **True Positives**

  True positives are cases where the index test has appropriately determined the

presence of the condition according to a predetermined significance threshold compared with a reference standard.

- **False Positives**

  False positives are cases where the index test has incorrectly defined the condition as being present when the condition does not exist according to a predetermined significance threshold compared with a reference standard.

- **True Negatives**

  True negatives are cases where the index test has appropriately determined the absence of a condition when it does not exist according to a predetermined significance threshold compared with a reference standard.

- **False Negatives**

  False negatives are cases where the index test has incorrectly defined the absence of a condition when it exists according to a predetermined significance threshold compared with a reference standard.

Table 1: 2x2 Table of diagnostic test accuracy

	Condition Positive	**Condition Negative**
Test Result Positive	True Positive (TP)	False Positive (FP)
Test Result Negative	False Negative (FN)	True Negative (TN)

Diagnostic accuracy results are commonly presented in terms of sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV), which are determined from formulae that express various combinations of the TP, FN, FP and TN.[5] These are described below.

Sensitivity

Sensitivity can be defined as the true positive rate. That is, the number of positive results recorded by the index test as a proportion of the positive results recorded by the reference standard.

$$Sensitivity = \frac{TP}{TP+FN}$$

A test with a high sensitivity is very effective at determining when the target condition is present (there are not many false negatives).

Specificity

Specificity is defined as the true negative rate. That is, the number of negative results recorded by the index test as a proportion of the negative results recorded by the reference standard.

$$Specificity = \frac{TN}{TN+FP}$$

A test with a high specificity is very effective at determining when the target condition is absent (there are not many false positives).

Positive Predictive Value

Positive predictive value (PPV) is the accuracy of the index test at predicting a true positive result. That is, the number of positive results recorded by the index test as a proportion of the positive results recorded by the index test and reference standard.

$$PPV = \frac{TP}{TP+FP}$$

A test with a high PPV is very effective at correctly identifying when the target condition is positive, that is, confirming disease is present.

Negative Predictive Value

Negative predictive value (NPV) is the accuracy of the index test at predicting a true negative result. That is, the number of negative results recorded by the index test as a proportion of the (true and false) positive results recorded by the index test.

$$NPV = \frac{TP}{TP+FN}$$

A test with a high NPV is very effective at correctly identifying when the target condition is negative, that is, ruling out disease.

An overall estimate of the accuracy of a test is also possible. Whilst overall accuracy is simply calculated as half of the summed value of the sensitivity and specificity, this value adds little information and is not applicable beyond the immediate cohort. A more appropriate assessment of overall diagnostic accuracy of a test is the area under the curve (AUC) measurement of a receiver operating characteristic (ROC) curve.[2]

Receiver Operating Characteristic Curve

An ROC curve is used to graphically display the effective sensitivity for each specificity for every disease threshold. More specifically, the ROC is a plot of the true positive rate (or sensitivity) compared to the true negative rate (equal to 1 minus the specificity).[2] To be useful a test must be able to produce a sensitivity and specificity of more than 50%. That is, it needs to be more effective than a random guess. A 45^{o} angled dotted line running through the middle of the graph represents the relative diagnostic threshold used. As the diagnostic threshold is increased from the lower left to the upper right portion of the graph, the bias moves in favour of FP to FN outcomes. The degree to which the test is useful is visually represented by how close the curve reaches toward the top left corner of the graph. An example of a ROC curve for a highly accurate test is presented in Figure 1. Figure 2 displays a ROC curve with more modest diagnostic accuracy, and Figure 3 shows a ROC curve for an ineffective test.

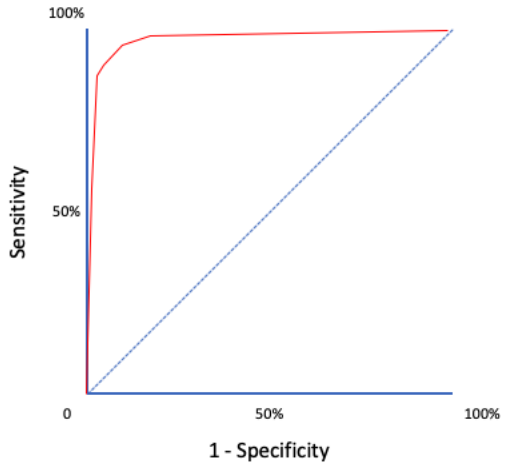

Figure 1: Example ROC with high level accuracy

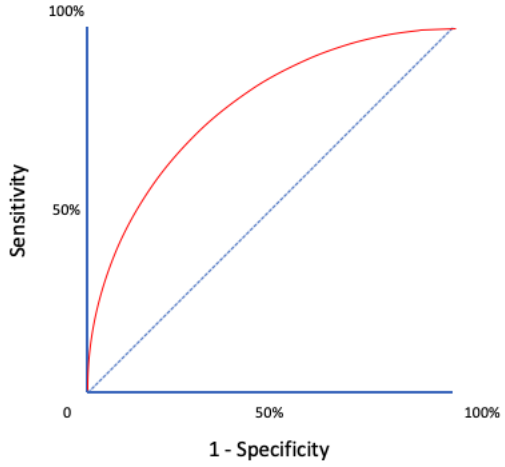

Figure 2: Example ROC with modest accuracy

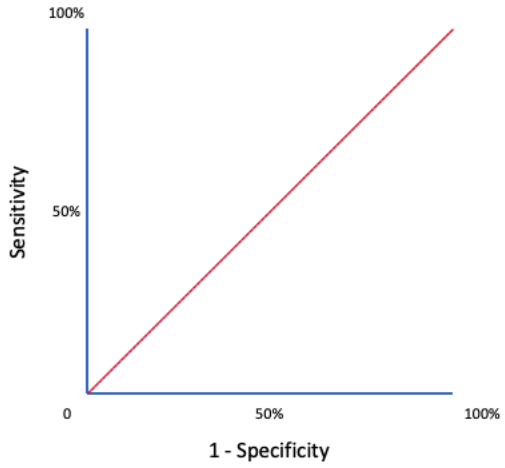

Figure 3: Example ROC showing an uninformative test

Area Under the Curve

AUC is a measure of the overall accuracy of a diagnostic test. The AUC describes the measurement of the area under the ROC curve. For tests with a high sensitivity and specificity across different diagnostic thresholds, there is a greater area between this line and the line of test ineffectiveness.

Whilst the AUC has value in the overall power of a test, the clinical significance of this metric is limited in isolation as it does not describe the dichotomous relationship of diagnostic accuracy.[2] Figure 4 describes two alternative tests (red and green lines) with similar AUC. The green line represents a test with a high sensitivity but has relatively low specificity. The red line represents a test with high specificity but low sensitivity. The appropriate use of these tests will differ depending on the pre-test probability and the implications of a FP or FN for the diagnosis of interest. However, an AUC measurement does not provide this information. Therefore, the assessment of a test's importance should not rest solely on the AUC value.

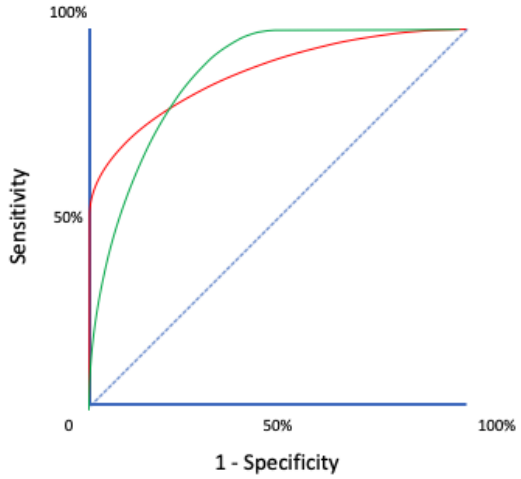

Figure 4: Example ROC with similar AUC but differing sensitivity and specificity

The appeal of the ROC plot is that it describes the accuracy of a test irrespective of a specific disease threshold. This is a particularly important value particularly in diagnostic fields such as radiology, where continuous or subjective expert observer judgements are the norm.[2]

Other Diagnostic Accuracy Metrics

Many other calculations are used in studies of diagnostic test accuracy; for example, diagnostic odds ratio, likelihood ratio, and Youden's Index.[6] However, only the metrics explained above were encountered in the review. Consequently, descriptions of other diagnostic test accuracy metrics are not provided here.

Common Diagnostic Test Accuracy Study Designs

Studies of diagnostic accuracy can be broadly categorised as one of two designs: case-control studies or consecutive cohort designs. Less commonly, diagnostic accuracy may be measured via a randomised control trial.[3]

Case-Control Study

Diagnostic case-control studies (sometimes called two-gate designs) are study designs where a group of participants with the diagnosis of interest (known cases of disease) are compared to a group of participants without the diagnosis of interest (controls). Because

case-control studies are able to provide similar numbers of participants with and without the disease, they may overestimate diagnostic accuracy due to the artificially high prevalence of the diagnosis of interest in the data set.[7] Additionally, these studies usually contain a high risk of bias if observers are not appropriately blinded to the results. Figure 5 depicts a common two-gate design.

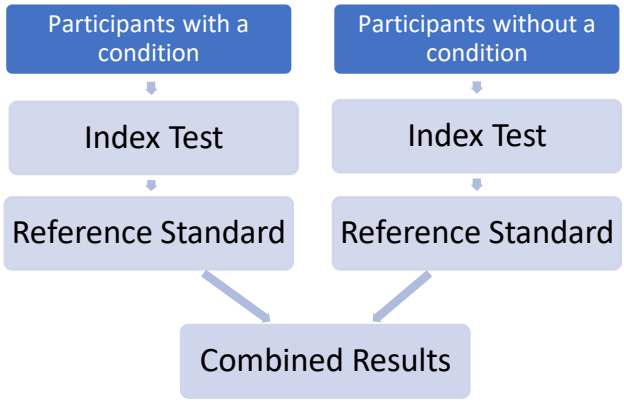

Figure 5: Diagnostic case-control study design

Consecutive Cohort

The majority of diagnostic accuracy studies are consecutive cohort studies.[8] This is a single-gate design, where participants receive the index and reference test (Figure 6). In theory, this is a more robust design as it better reflects the true prevalence of the disease in the population, giving a more accurate representation of diagnostic accuracy in the broader population. This is generally a prospective design but may be performed retrospectively. Retrospective designs tend to be at higher risk of spectrum bias, however, as the exclusion criteria are not determined *a priori,* and it is not always clear if patients were randomly selected. Further explanation of spectrum bias is provided later in the chapter.

In prospective consecutive cohort designs, the participants may receive the index or reference test first, so long as the results from one test cannot influence the results of the other. It is also vital that the tests are performed within a similar time period, which will vary depending on how likely it is that the diagnosis of interest will change in the interim.

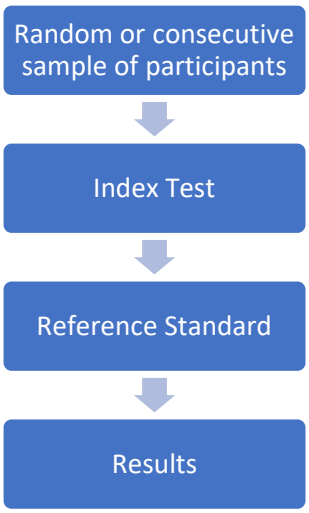

Figure 6: Observational cohort study design

Randomised Controlled Trials

Experimental or randomised controlled trial (RCT) designs are the "gold" standard for assessing effect of treatment outcomes.[9] RCTs have also been used to determine diagnostic accuracy.[4] RCTs are useful in helping to identify important outcomes resulting from the use of a test, such as treatment and cost; they are not a preferred design for assessing diagnostic accuracy. Whilst a randomised or consecutive sample of participants is recommended in diagnostic accuracy designs, randomisation of groups into various arms is not generally relevant when assessing diagnostic accuracy.[10] Furthermore, if diagnostic accuracy is assessed as a secondary outcome, the design may introduce spectrum bias relating to an artificially inflated prevalence of the diagnosis of interest.

Systematic Reviews

Systematic reviews have gained significant popularity since their inception in the 1970s.[11] A systematic review differs from a traditional literature review in a number of key ways.

Firstly, a traditional (or narrative) literature review tends not to include an exhaustive search strategy to identify all material on a particular topic. Rather, it is used to give a general overview narrative and allows a more relaxed approach to "cherry picking" included studies.

Secondly, narrative literature reviews do not critically appraise the evidence they include, as stringent inclusion criteria are generally not applied. This gives rise to potential bias in any conclusions made from summarising the results.

Thirdly, the rigour of the search process and the critical appraisal in a systematic review mean that quantitative analysis is often (although not always) possible. Meta-analysis is a way of quantitatively summarising all available data on a particular review question. All meta-analyses are systematic in nature, although not all systematic reviews necessarily contain a meta-analysis.

Owing to the robust and rigorous design, well performed systematic reviews of randomised controlled trials are often considered the highest quality of evidence available, although this is only when assessing the effectiveness of treatment strategies.[12] Systematic reviews are also used to assess questions other than effectiveness. For example, systematic reviews are frequently conducted to investigate cost effectiveness, prevalence, incidence, aetiology, risk, prognosis, and research methodology as well as reviews conducted on qualitative (or experiential) evidence.[13] Organisations such as Cochrane and the Joanna Briggs Institute now have several methodological guidance resources and tools to assist reviewers in conducting systematic reviews based on the investigated methodology of choice.

Systematic Reviews of Diagnostic Test Accuracy

A systematic review of diagnostic test accuracy is a specific type of review methodology that specifically aims to answer questions regarding the accuracy of a test based on all known (published and unpublished) data.[3,14] The data are reported specifically in terms of the diagnostic accuracy of the tests, that is, sensitivity and specificity. As such, a systematic review reports and summarises key metrics of diagnostic accuracy.

Risk of Bias and the QUADAS-2 Critical Appraisal Instrument

Robust and independent critical appraisal is a cornerstone of the rigour of systematic reviews. Critical appraisal is vital to ensure included studies are valid, consistent and reliable, and that significant sources of bias are accounted for. Critical appraisal of a

systematic review is generally performed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool,[10] more recently refined as the QUADAS-2 critical appraisal tool.[15] QUADAS-2 is the current research standard for reporting risk of bias in diagnostic accuracy reviews as it is crafted to assess all expected types of bias present in studies of diagnostic accuracy.[16]

QUADAS-2 uses four key domains and several associated signalling questions to assess risk of bias and applicability of each study relative to the systematic review question and associated criteria. The four key domains describe patient selection, the index test, the reference standard, and the flow and timing of patients who receive the index test and reference standard. The signalling questions are able to be customised to reflect the particular question *a priori*.

The tool has come under some criticism for its deliberately qualitative nature,[17] and in its current form does not adequately assess comparative diagnostic accuracy studies where more than one index test is assessed.[18,19] However, it remains the instrument of choice for assessing risk of bias in studies of diagnostic test accuracy.

Sources of Bias in Diagnostic Accuracy Studies.

In order to minimise bias and maintain transparency, studies of diagnostic test accuracy should follow the STARD (Standards for Reporting Diagnostic Accuracy Studies) statement.[8]

Potential for bias exists in all studies and is inherent in many different forms.[20] Bias may be overt or unintended. Studies of diagnostic test accuracy are more susceptible to certain types of bias.[21] The QUADAS-2 critical appraisal tool is designed specifically to flag these types of bias within included studies.[15] These have been described in detail below.

Bias associated with study participants

Spectrum bias

Spectrum bias exists where the study participants do not reflect the true spectrum of patients who would undergo testing for the target condition in clinical practice.[22] The source

of bias may be deliberate, by actively excluding participants with either a high or low pre-test probability of having the target condition, or unintentional, such as by selecting a convenience sample of patients. A convenience sample might include patients who have received both index and reference standards tests as part of routine clinical investigations, but might not include patients who were not clinically required to have the test or standard. In the latter example, spectrum bias could exist where the reference test was particularly invasive or unpalatable, and therefore only patients with high clinical suspicion of the disease are recruited to the study.

The effect of spectrum bias will depend on the group of participants who have been excluded. Where high-risk patients are excluded, there will likely be less disease in the cohort and therefore the sensitivity of the test may be overestimated. Where low risk patients are excluded, the disease burden in the patient would more likely be high and there would therefore be an overestimation of the specificity of the test when extrapolated to the general clinical population.

Selection bias

Selection bias exists where participants are assigned under a pre-known determinant, as in a case-control study.[23] That is, they are not assigned consecutively or randomly to receive the index and reference tests. A case-control study design will usually overestimate the accuracy of the index test, as known or predefined cases may be easier for a test to identify or exclude.

Bias associated with the index test

Information bias

Information bias occurs when more (or less) information is available to interpreters of the index test than would otherwise exist in routine practice.[23]

Bias associated with the reference standard

Misclassification bias

Misclassification bias occurs where there is an incorrect assignment of a piece of data.[8] In the case of a diagnostic accuracy study, this will occur where participants have been classified incorrectly with a diagnosis of interest. This will occur where the reference standard is a poor indicator of the presence or absence of disease. The accuracy of the results of the index test will vary depending on whether the reference standard is biased toward overestimating or underestimating the true diagnosis of interest.

Verification bias

Verification relates to how the patients' index test results were corroborated. Verification bias exists where there is a discrepancy in the verification of a group of participants within the study. Verification bias can be further categorised as partial or differential bias. Partial verification bias is usually associated with the conduct of the reference test, whereby a non-random group of participants do not undergo the reference test. The result of partial verification bias is generally an overestimation of sensitivity.[24]

Differential verification bias can occur when a non-random group of participants undergoes an additional or alternative reference test. This often leads to an overestimation of the accuracy, particularly when the need for a different reference standard is immediately based on the index test result.

Bias associated with flow and timing in study

Incorporation bias

Incorporation bias occurs when the index test informs the position of the reference test.[25] This is most common where the reference standard is a clinical diagnosis such as in a discharge summary, meaning that the interpretation of the index test will have comprised (at least) part of the reference standard. This will usually lead to an overestimation of the accuracy of the index test.

Disease progression bias

Disease progression bias exists when there is a change in the participant's condition between the performance of the index test and the reference test.[23] This will usually lead to overestimation of the accuracy of the index test results, as some conditions may become more pronounced as they progress and therefore easier to assess.

Information bias

Information bias exists where the reference test is interpreted with knowledge of the results of the index test or vice versa.[23] For this reason, in tests where data requires subjective interpretation, for example in radiology, diagnostic test accuracy studies should be performed with blinding of the interpreters to the results of the index test.

Bias associated with analysis of data

Excluded data bias

Excluded data bias exists where uninterpretable or equivocal test results are excluded from the analysis.[23] This will usually lead to an overestimation of the accuracy of the index test as test results that are more difficult to interpret are more likely to be those that are falsely classified.

Data Analysis

Forest Plots

Meta-analyses commonly contain forest plots. Forest plots are graphical representations of point estimates and associated confidence intervals for each study that reports a particular outcome in a review. The summary estimates also include a characteristic box that indicates how much weight the particular result carries in the overall estimate for the review.

A summary estimate of the effect size is usually calculated and presented at the bottom of the forest plot diagram. Relevant statistics – such as the specific statistical weight each study finding provides to the overall estimate, as well as indicators of the inherent statistical heterogeneity – also accompany the forest plot.

Paired Forest Plots

Forest plots in diagnostic test accuracy systematic reviews differ from traditional forest plots (i.e. those used in studies of treatment effectiveness). Whereas traditional forest plots are usually interested in comparing relative risk or odds ratio for a particular outcome, meta-analyses of diagnostic test accuracy use paired forest plots of sensitivity and specificity for each included study.[26] Pooling of these results is possible but is not always appropriate.

Pooled Sensitivity and Specificity

Meta-analytical pooling of sensitivity and specificity data from primary studies has been performed using various methods.[27,28] In meta-analysis of diagnostic test accuracy studies, data should only be pooled where it is believed there is little heterogeneity in the conduct of included studies, and the same diagnostic threshold is used in each study. Where thresholds vary, diagnostic accuracy meta-analysis is still possible, but should be performed via a Summary Receiver Operating Characteristic curve.[14]

Summary Receiver Operating Characteristic

Summary Receiver Operating Characteristic (SROC) curves are meta-analysed ROC curves. SROC curves are also commonly referred to as Hierarchical Summary Receiver Operating Characteristic Curves (HSROC) as they use statistical weighting to evaluate the strength of a finding from the primary study in ROC space.[29]

SROC curves are valuable when primary studies report diagnostic accuracy data with differing diagnostic thresholds. The SROC allows for these data to be collated in ROC space.

The purpose of this chapter is to familiarise the reader with the key components of studies of diagnostic accuracy as well as systematic reviews of diagnostic test accuracy, and to show how these components differ from traditional systematic reviews of effectiveness. The following chapters describe the conduct and findings of the systematic review.

Chapter Two: Background and Rationale

Background

Several tests are used in the classification and diagnosis of coronary artery disease (CAD). These range from common tests such as blood pressure tests, electrocardiogram (ECG), and blood tests, to more complicated medical imaging examinations. The accuracy of such tests is an important consideration for guideline developers and clinicians alike, to determine the appropriate position for each diagnostic test in the assessment of CAD.

This chapter discusses pertinent background information regarding the current use of computed tomography coronary angiography (CTCA), how the technology works, and the limitations and developments of this method as they relate to patients with elevated heart rates (HR).

Coronary Artery Disease

CAD, also known as ischaemic heart disease, is narrowing of the coronary arteries, which are responsible for blood supply to the myocardium. CAD is caused by a build-up of fatty deposits within the lumen of the artery, which may harden over time.

CAD is the leading cause of death in Australia and worldwide, in both males and females.[30] According to the Australian Bureau of Statistics, CAD was responsible for 11.6% of recorded deaths in 2017.[30] Clinically significant CAD is likely to lead to acute coronary syndrome (ACS), an umbrella clinical diagnosis which may be sub-categorised as myocardial infarction or unstable angina.[31]

Myocardial infarction occurs when there is occlusion or a very high-grade stenotic region that restricts flow, causing damage to the myocardium. The extent of the damage depends on which segment of the artery contains disease and how much the flow is impacted distal to this segment. Where there is transient or incomplete obstruction of flow to the myocardium, unstable angina is likely to occur.[32] This is severe chest pain associated with a narrowing, usually considered to be greater than 50% reduction of the luminal diameter,

that does not cause immediate damage to the myocardium.[33] Myocardial infarction can be considered as ST-segment elevated myocardial infarction (STEMI) or non-ST segment elevated myocardial infarction (NSTEMI).[34] The latter is a diagnosis made in the absence of significant (>2mm) change in ECG with continuous raised troponin levels in presence of chest pain with or without the use of confirmatory anatomical or functional imaging tests.[35]

Significant Coronary Artery Stenosis

The 50% intraluminal narrowing threshold has been considered the diagnostic threshold of clinical significance of CAD for several decades.[33] This threshold has been increasingly scrutinised more recently, with suggestions that it should be correlated with functional testing also.[36] As the purpose of this review is to define clinically significant CAD, and 50% is still routinely used in practice, this is the measure that will be used. Further comment as to the value or limitation of 50% as an appropriate cut-off for clinically significant CAD is beyond the scope of the current thesis.

The American Heart Association (AHA) segmental model of the coronary tree was established in 1975 by Austin and colleagues[37] and is still in use. The Society of Cardiovascular Computed Tomography[38] suggest the use of this model and include a slight 17-segment adaptation to the model for use in interpretation of images.

Computed Tomography Coronary Angiography

Background

CTCA is a type of CT examination that assesses the coronary artery tree. Studies investigating the feasibility of coronary artery angiography in CT began to emerge from the turn of the 21st century, precipitated by the introduction of helical CT, which was a significant advance in CT technology.[39] There was significant growth of CTCA in the following decade as scanner technology continued to improve. Of particular importance was the development of multislice CT scanners that could generate multiple slices of the heart within one rotation. Early multislice scanners consisted of two to four rows of detectors. They were superseded by models with 16, 32 and 64 detector row configurations. Today's systems routinely use greater than or equal to 128 detectors.

Since the advent of high accuracy results from CT scanner technology from about 2007, the importance of CTCA was further considered. CTCA is now regularly used as a frontline examination for patients with chest pain as a gatekeeper for invasive coronary angiography.[40]

Performance of CTCA

ECG gating

ECG gating is used to reduce motion artefacts. To adequately image the heart, the scanner is required to ascertain a 180-degree snapshot of data. Therefore it is vital that this is performed at a time point when there is very little motion of the coronary artery tree.[41]

Retrospective gating

The original multidetector scanners used to perform CTCA routinely employed a retrospective gating technique.[41] This technique involves initiating the exposure at a certain time point, acquiring data over more than one heartbeat and recording the ECG trace during the acquisition. The recorded ECG trace is then retrospectively analysed and images are reconstructed at a certain point between the R peaks (typically at the end diastolic period, as this is usually the part of the heart cycle with least motion). The value of this technique is that it offers a range of time positions to be reconstructed. For example, by using retrospective gating, functional analysis of the heart is possible. However, this comes at the expense of a considerable radiation dose to the patient.

Prospective gating

Prospective gating (or triggering) allows the scanner to prospectively assess the ECG trace in the beats immediately prior to the planned acquisition and prospectively turn the exposure on for the minimum amount of time needed to generate an image at the required part of the heart cycle. As a result, radiation exposure to the patient is much smaller than that received from retrospectively gated studies. Temporal padding, where the x-ray exposure is extended beyond the minimum required, is often included to allow images to be reconstructed either side of the cardiac phase selected.[42] This may be only a small temporal

window surrounding the planned end diastolic phase or it may be extended to include mid-systole. Padding provides additional data, which increases the chance of finding a point within the heart cycle to reconstruct motion-free images, particularly in patients with higher and more erratic HRs.

Importance of Heart Rate in CTCA

HR has historically been a limiting factor in the performance of CTCA.[43] HR effects the quality of the scan in two ways. If the HR is elevated, it is likely to result in motion artefact, as the quiescent phase of the myocardium will be shorter than the time needed to acquire the required number of rotational projections.

The second situation in which HR affects the quality of CTCA is where there is significant HR irregularity. This may be caused by ectopy or by a non-sinus rhythm.[43]

Temporal Resolution

Scanner temporal resolution is a term used to describe the time required to effectively image a moving object, such as the coronary arteries. The temporal resolution of a system is dependent on a number of factors. Rotation time of the scanner is an important element in defining this; however, temporal resolution is also controlled by the number of projections required to reconstruct an image, the mode of scanning (axial versus helical) and whether multi-segment reconstruction is used. Motion correction algorithms must also be considered. Therefore, the term temporal resolution is not necessarily an accurate measure of the scanner's overall ability to produce a well-defined image.

Models of CT Scanners

Emerging technology in CT is generally tightly controlled by each manufacturer; details of technical advances are therefore vendor-specific and protected by patent law. As a result, each manufacturer creates different and novel ways of tackling the issue of HR in CTCA. The specific makes, models and related technology included in this review are briefly examined below.

Canon - Aquilion ONE and Aquilion ONE ViSION

The Canon Aquilion ONE scanner was the first to market to include a 320-detector array, allowing a volume of data to be acquired in a single rotation of the x-ray tube within the gantry housing. The 0.5 mm detectors allow for up to 160 mm of anatomy to be covered per rotation. This has the advantage of being able to cover the whole heart in one rotation of the gantry, which reduces total scan acquisition time and avoids "stair step" artefact caused by mis-registration of consecutive slices due to inconsistent HR. The Aquilion ONE system is capable of a rotation time of 350 milliseconds.[44]

The Aquilion ONE ViSION Edition was the next iteration of the scanner. It uses the same scanner design as the Aquilion ONE but features a larger generator capacity and an improved rotation time of 275 milliseconds. As a reduction in rotation time improves the temporal resolution of a scanner, the ViSION edition of the Aquilion ONE has a considerable advantage over the standard version for CTCA.

GE - Revolution CT

The GE Revolution CT uses a 256 multidetector array. The minimum rotation time of the Revolution CT is 280 milliseconds.[44] The system uses a novel computational algorithm known as SnapShot Freeze™. This algorithm improves the resolution in the spatial domain by reassessing the ray-sums in the projection angles to correct for motion around vessels.[45-48] The algorithm can be applied prospectively or retrospectively to reduce artefacts from motion blur. Whilst other vendors have developed similar retrospective motion correction algorithms, SnapShot Freeze was the only motion correction technology described in diagnostic accuracy studies and was therefore the only algorithm included in the review.

Philips - Brilliance iCT

The Philips Brilliance iCT scanner contains a 256-multidector array. Due to a novel air cushion technology, the gantry rotates without the use of bearings, allowing a high rotation speed. The rotation time of the system is 270 milliseconds.[44]

Siemens - SOMATOM Definition Flash

Siemens' SOMATOM Definition Flash incorporates a dual-source (i.e. twin x-ray tube) configuration. Each tube has a corresponding 128-multidetector array. The SOMATOM Definition Flash is capable of a 280 millisecond rotation time.[44]

Reference Standard - Invasive Coronary Angiography

Invasive coronary angiography (ICA), also known as coronary catheter angiography, is a fluoroscopic procedure used to define narrowing of the coronary arteries. ICA is a well-established diagnostic technique and has long been considered the reference or "gold" standard for assessment of the degree of coronary artery stenosis.[49] Recently, however there has been some contention about the value of ICA as a reference test for CAD detection.[36] Less invasive tests such as CTCA have reduced the number of patients undergoing the invasive procedure of ICA.[40]

Furthermore, the definition of clinically significant CAD has come under increased scrutiny. This is because the level of disease-causing clinical symptoms lies on a continuum rather than at a specific measurable threshold.[50] Therefore, a true reference standard for clinically significant CAD should include the effect of the stenosis on flow, as well as the degree of luminal narrowing. However, as ICA is still the accepted reference test in majority of published papers, it has been chosen here as the reference standard. Further research describing a more effective standard is beyond the scope of the current thesis.

ICA is generally performed via an upper limb brachiostomy (Sone's) or by puncture of the common femoral artery (Judkin's) approach.[51] The added value of this test is that if a highly significant stenosis is identified, intervention can be performed immediately through angioplasty (balloon remodelling), or coronary stent insertion.

Other imaging tests regularly used in assessment of CAD include nuclear medicine myocardial perfusion imaging, stress echocardiography, and cardiac magnetic resonance imaging. Whilst studies have described varying degrees of accuracy of these tests in assessing coronary artery stenosis, invasive coronary angiography is still considered a

reference standard.[52] Therefore, only studies that make a comparison to ICA as a reference standard were included in this review.

Rationale

Guidelines currently exist for the performance and assessment of CTCA for patients with low HRs.[53,54] However, there is limited guidance on the use of CTCA in patients with high HRs. Reference to the value of different scanner hardware and software for the optimisation of diagnostic accuracy in this cohort of patients is particularly insufficient. Some studies have been performed to determine the diagnostic accuracy for this group, but they do not challenge current guidelines. This may cause patients who would otherwise be suitable candidates for CTCA to be sent for ICA instead. Therefore, there is value in a systematic review that evaluates the diagnostic accuracy of CTCA for patients with elevated HRs.

In a 2013 systematic review, Westwood and colleagues[55] described the diagnostic accuracy of CTCA in difficult-to-image patient groups. One of the subgroups classified as difficult to image were those with HRs greater than 65 bpm. However, this review described dual-source scanners specifically and did not include recent advances in technology aimed at improving temporal resolution.

A 2013 review protocol was registered and published by a collaborative known as the Collaborative Meta-Analysis of Cardiac CT (CoMe-CCT).[56] This partnership aims to meta-analyse all individual patient-related data to determine specifics of diagnostic accuracy for CTCA. A recently published initial publication was identified that described individual patient outcomes. Whilst HR was not detailed exclusively in the analysis, the authors made comment that HR was the only factor associated with non-diagnostic examinations.[57]

A 2014 systematic review by Li and colleagues[58] described the diagnostic accuracy in patients with and without HR control medications. However, the study did not specifically describe the relative HRs of each group. Furthermore, this review incorporated 64-slice scanner models, which are likely to have a lower accuracy in the detection of coronary artery stenosis for patients with elevated HRs.

No other systematic reviews were located that directly assessed the diagnostic accuracy of patients with high HRs.

The aim of this review is therefore to assess the diagnostic accuracy of patients with high HRs undergoing CTCA with current generation scanners, using ICA as a reference standard. The purpose of the review is to inform clinicians of the effect of high HRs on sensitivity and specificity and to describe the effect individual technologies may have on this. It is therefore hoped this review will inform future guidelines relating to the appropriate imaging of patients with high HRs.

Chapter Three: Methods

Introduction

This chapter describes the methodology used in the systematic review and how the data were analysed. Specifically, it describes the conduct of the systematic search strategy, extraction of results, synthesis and meta-analysis. The methods for the study were published *a priori* in a review protocol.[59]

Review Question

Based on guidance by Munn and colleagues,[13] a Participant (P), Index Test (I), Reference Standard (R), and Diagnosis of Interest (D) mnemonic was used. This focused the creation of the review question. The primary question to be answered by this review became: "What is the accuracy of current generation CTCA relative to ICA for the diagnosis of clinically significant CAD in patients with high HRs?"

A secondary review question was: "What is the diagnostic accuracy of vendor-specific scan technologies in patients with high HRs?"

Inclusion Criteria

Types of Studies

The review considered diagnostic test accuracy study types, specifically those with a cross-sectional design. Case-control studies were also considered for inclusion.

Participants

The review considered adult patients (18 years and over) with a high HR who underwent CTCA to rule out or confirm clinically significant CAD. For the purposes of the review, high HR was defined as a HR greater than 65 bpm.

The review excluded patients with coronary artery bypass grafts, a history of heart transplantation and/or intraluminal stents. Studies that exclusively studied patients with obesity, atrial fibrillation or severe CAD (greater than 75% luminal narrowing) were also excluded.

Studies that assessed anthropomorphic or other phantoms, ex-vivo, or animal studies were also excluded.

Index Test

This review considered studies that evaluated CT scanner technology utilising either single- or dual-source configuration, in prospective ECG scan acquisition mode and with a total scanner coverage of greater than or equal to 128 detector-rows. The review did not consider studies where retrospective gating was performed, or where ECG gating was not used.

Studies that investigated non-coronary examinations, such as scans of the aorta, coronary artery calcification, pulmonary veins or functional studies of the heart were excluded.

Scanners from the four major manufacturers were included. Note that whilst included studies refer to *Toshiba* scanners, Canon Medical have acquired this brand and for simplification in the review all description of Toshiba scanners will be included as Canon.

Reference Standard

This review considered studies that compared the index test (CTCA) to ICA. ICA could be performed before or after CTCA, but the two tests must have been performed within one year of each other and without interim change in a patient's symptoms or treatment. Where the interval between tests was unclear, studies were still included, but it was addressed in the methodological quality assessment.

Diagnosis of Interest

Clinically significant CAD was defined as the presence of a stenosis measuring greater than 50% of the vessel lumen diameter. The review considered diagnostic accuracy at a per-patient, per-vessel and per-segment level. Any coronary artery segmental model was included.

Search Strategy

A comprehensive search strategy was designed, which included key search terms associated with each aspect of the review question (i.e. population of interest, index test, reference standard and diagnosis of interest). This was entered into a logic grid format and then adapted as a search string. PubMed was used as the test database for refining the search. Each column of the logic grid was tested in PubMed individually to optimise search results, by testing the effect of expanding and collapsing individual search terms and considering whether the search term be applied as a medical search heading term or text word searchable term. Once this was optimised for each column of the logic grid, the sections were combined to create the final search string for the database. Search filters excluded studies performed prior to 2007 and studies published in any language other than English. Based on recommendations by Leeflang and colleagues,[60] methodological search filters were not applied. The search string was then adapted to other databases as per the individual database's preferred search terms. The detailed search strings for each database searched has been included as Appendix One. The following databases were searched in the review: CINAHL, Embase, PubMed and Scopus.

In addition to the formal database search strategy, several other methods were incorporated to attempt to identify unpublished works. A search of the first three pages of Google Scholar was performed to identify any articles not picked up in the main search. The ProQuest database was also searched for published dissertations, theses or other grey literature. Search entry terms were simplified for grey literature searches to "CT", "coronary angiography" and "heart rate". All citations were downloaded into citation reference manager (EndNote X8, version 8.2, Clarivate Analytics, USA).

Additionally, key authors were contacted to suggest any additional studies they were aware of that would match the review question. Finally, the reference lists of the included studies were searched to identify possible studies.

Assessment of Methodological Quality

The QUADAS-2 tool was used to assess methodological quality in this review.[15] Selected studies were critically appraised by two independent reviewers. Any discrepancies that arose between the two reviewers' appraisals were resolved through discussion.

QUADAS-2 Risk of Bias Signalling Questions

The QUADAS-2 tool is divided up into 11 questions over four domains that interrogate the risk of bias within each study. The four risk of bias domains that the signalling questions are applied to are:

1. Patient selection
2. Index test
3. Reference standard
4. Flow of the patients through the study and timing of the tests.

QUADAS-2 is designed to be applied initially to a test group of studies and iteratively developed to ensure the signalling questions are being correctly interpreted and applied by the reviewers.[15] Therefore, the two reviewers randomly selected three included studies, and each independently assessed these for methodological quality. The results of this initial pilot were compared and any differences in results were discussed to see if any changes needed to be made to the instrument before applying to all included studies. Once both reviewers were satisfied with the application of the QUADAS-2 signalling questions, the reviewers applied the signalling questions to each study independently.

The following signalling questions and associated working definitions were agreed upon and used to assist the reviewers to determine the risk of bias and applicability scores for the systematic review.

DOMAIN 1: Patient Selection

Q1. Was a consecutive or random sample of patients enrolled?

This question is used to signal whether spectrum bias is present. Did patients referred for CTCA reflect the full cohort of patients who would undergo this test clinically? A "Yes"

should be scored if the paper describes a consecutive or random sample of participants entering the investigation. If there is concern that the participants have been entered based on their known risk of disease, a “No” should be recorded for this signalling question.

Q2. Was a case-control design avoided?

This question is used to determine if spectrum bias is present through study design. A case-control study will exaggerate the positive rate compared to the normal prevalence occurring in a clinical cohort. Any study that describes a case-control or two-gate design should receive a “No” for this signalling question.

Q3. Did the study avoid inappropriate exclusions?

This question is also used to assess spectrum bias associated with participants being excluded from the study inappropriately, i.e. the exclusion of patients with a diagnostic profile that would normally be included in clinical practice. As the reference standard is an invasive technique, it is likely that some lower risk patients will be excluded from the accuracy assessment. Studies that only include patients who have undergone (or are already scheduled to undergo) ICA for clinical reasons should receive a “No” score for this signalling question. Where it is not explicitly stated, “Unclear” should be scored for the signalling question.

DOMAIN 2: Index Test

Q4. Were the index test results interpreted without knowledge of the reference standard?

This question assesses whether the observers of the index test were blinded to the results of the reference standard. If the index test was performed prior to the reference standard, the study must directly state that blinding of the observer was performed for this signalling question to achieve a “Yes” risk of bias score. If this was not described, “Unclear” should be recorded for this signalling question.

Q5. If a threshold was used, was it prespecified?

This question interrogates whether diagnostic thresholds were predefined to reduce verification bias. For this signalling question to receive a “Yes” score, the study should

explicitly state that diagnostic thresholds were previously defined, and what the threshold was. If this is not explicitly described in the paper, an "Unclear" result should be recorded for this signalling question. It is expected that most scores will be "Yes", as greater than 50% luminal stenosis is a well-accepted standard for clinically significant disease.

DOMAIN 3: Reference Standard

Q6. Is the reference standard likely to correctly classify the target condition?

This question is used to identify if the reference standard used is appropriate to define the diagnosis of interest. As the review exclusion criteria state that only ICA is to be used as the reference standard, this signalling question will score a "Yes" risk of bias score, unless there was sufficient reason to believe that the study did not present enough information to accurately describe the performance of the reference test.

Q7. Were the reference standard results interpreted without knowledge of the results of the index test?

This question is used to assess whether bias is avoided by ensuring the observer is scoring the reference standard without knowledge of the index test. If the index test is performed prior to the reference test, this signalling question should only receive a "Yes" score if there is an explicit statement made about the appropriate blinding of the observers to the index test. If this is not included, an "Unclear" score should be entered for this signalling question.

DOMAIN 4: Flow and Timing

Q8. Was there an appropriate interval between index test and reference standard?

This question is used to assess whether interval change in a patient's condition could have affected the results of the study. That is, was there a short period (less than one month) between the performance of the two tests? If there was longer than one month between tests, the authors should specifically state that they checked there was no change in the patient's symptoms or condition between the tests, or an "unclear" risk of bias should be recorded. Where a long period exists (greater than six months), a "high" risk of bias should be recorded for this signalling question.

Q9. Did all patients receive a reference standard?

This signalling question is used to assess whether partial verification bias is likely in the conduct of the study. Where only a subgroup of study participants receives the reference standard, this may risk affecting the accuracy calculations, as typically only the more high-risk participants receive the reference standard. Where this occurs, a "No" score should be recorded for this signalling question.

Q10. Did all patients receive the same reference standard?

This signalling question is used to signal any bias associated with the varying accuracy of differing reference standards being used to define the same diagnosis of interest. As the exclusion criteria for the review explicitly state that ICA is to be used as the reference standard, this signalling question should receive a "Yes" score, unless there is high level of suspicion that the reference standard differs in its conduct significantly to that defined in the study.

Q11. Were all patients included in the analysis?

This signalling question is used to assess risk of bias associated with inappropriately excluded data. A statement that non-diagnostic segments were recoded as positives should be included in the assessment of the index test. Where this statement does not exist, the signalling question should be scored as "unclear" risk of bias. Where non-diagnostic segments have been excluded from the analysis, this should be given a "No" score for this signalling question.

QUADAS-2 Concerns Regarding Applicability

A determination of the applicability of individual studies meeting the review criteria and hence external validity was undertaken also with the QUADAS-2 instrument. Due to the strict *a priori* exclusion criteria, this section of the critical appraisal was not expected to provide much additional information.

The concerns regarding the applicability section of the QUADAS-2 review tool does not use signalling questions, but looks generally at three of the four domains: index test, reference

standard and participants. The flow and timing domain is excluded from this section of QUADAS-2.

Data Collection

Data were collected from all papers included in the review, regardless of methodological quality, and entered using Microsoft Excel (Microsoft Office for Mac 2011, Version 14.7.1) by the lead reviewer. The data extracted included specific details about the tests, populations, study methods and diagnostic accuracy outcomes at patient, vessel, and segment levels. This data was defined in the review protocol *a priori*. A full list of extracted data is included as Appendix Three.

Data Synthesis

Narrative Synthesis

All included studies were synthesised narratively to describe the outcomes of the review.

Meta-Analysis

Proportional coupled meta-analysis was performed. This was appropriate as there was no variation in the diagnostic threshold reported between studies. Based on guidance by Campbell and colleagues[13] as well as that of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy,[6] pooled meta-analysis should only be performed using forest plots if the same diagnostic threshold is used in all studies. If diagnostic thresholds vary, accuracy should be determined via a SROC curve.[3,14,23] Studies were included if enough data were present to calculate point estimates and 95% confidence intervals for sensitivity and specificity. If insufficient data were provided, study authors were contacted in an attempt to gain this information.

Meta-analysis was performed using commercially available statistical analysis software (StatsDirect Ltd, Cambridge UK). This program allows summary estimates to be calculated as a proportional meta-analysis through the DerSimonian and Laird random effects model, using the Freeman-Tukey double arcsine transformation.[61] A random effects statistical model was considered a more appropriate model than a fixed effects model for this

purpose, due to the assumption that heterogeneity is inherent between and within diagnostic accuracy studies.[62] Paired forest plot analyses were performed at patient-, vessel- and segment level. All studies that reported sufficient 2x2 table data on high HR groups were included. Where possible, data were extracted, with non-evaluable segments treated as positive. If data were presented using different algorithms or methods, data were selected for the most clinically relevant assessment. For example, where a study provided comparative data with and then without a motion correction algorithm applied, the data with the algorithm applied were selected for inclusion in the meta-analysis. Furthermore, where non-established (novel) techniques were compared to recognised techniques, the latter data were included.

Sensitivity Analysis

Sensitivity analysis was performed to determine the effect non-evaluable segments had on sensitivity and specificity summary estimates. Two additional meta-analyses were planned at the segment level: one including studies that explicitly stated that non-evaluable segments were treated as positive in the data set, and one that included only studies that excluded non-evaluable segments. Studies that did not explicitly state how non-evaluable segments were handled were excluded from both sensitivity analyses.

Subgroup Analysis

Subgroup analysis figures were calculated and created using Review Manager (RevMan v5.3, Copenhagan, Denmark) software.[63] Subgroup analysis was performed to determine the effect varying HR thresholds had on diagnostic accuracy. A second subgroup analysis was performed to identify whether there were important differences in the accuracy outcomes reported for various scanner manufacturers. Neither of these analyses provided pooled summary estimates as there was insufficient data to produce these.

Heterogeneity

Heterogeneity was assessed via visual inspection of paired forest plots as well as inspection of the I^2 inconsistency statistic (calculated via StatsDirect). Following guidance from Cronin and colleagues,[26] I^2 was interpreted as low heterogeneity (<25%), moderate heterogeneity

(25-50%) or high level heterogeneity (>75%). Cochran's Q test was not used as it has low power to detect heterogeneity.[26] For subgroup analysis, visual inspection only was used to assess heterogeneity as results were not pooled.

Publication Bias

Publication bias was assumed but not tested in this review. This is because tests used for examining publication bias in systematic reviews of treatment outcomes are generally not considered appropriate in systematic reviews of test accuracy.[64]

Summary of Findings

A summary of findings table was produced from the results following the GRADE approach for diagnostic test accuracy studies.[65] GRADE constitutes a transparent and reproducible approach to grading the strength and quality of evidence for developing guidelines and recommendations.[66] The summary of findings table was created using GRADEpro GDT software [available online: https://gradepro.org].

Chapter Four: Results

Introduction

This chapter documents the findings of the systematic review. The search strategy results, characteristics and methodological quality of included studies as well as the findings of the review are described. Findings of the review are described in a narrative synthesis and with meta-analysis. A summary of the key findings of the review is displayed.

Search Results

The initial search identified 1691 records, including 1689 from database searching and two additional records from other sources. After removal of 636 duplicates, 1055 records were screened at a title and abstract level. Eighty full-text articles were then assessed, with 12 studies subsequently included in the review (Figure 7).[45-48,67-73] The search strategy was reported as per the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.[74] Studies were excluded on full-text assessment where they did not describe diagnostic test accuracy data specifically for patient groups with high HR, or where the index and/or reference test did not meet the predetermined inclusion criteria.

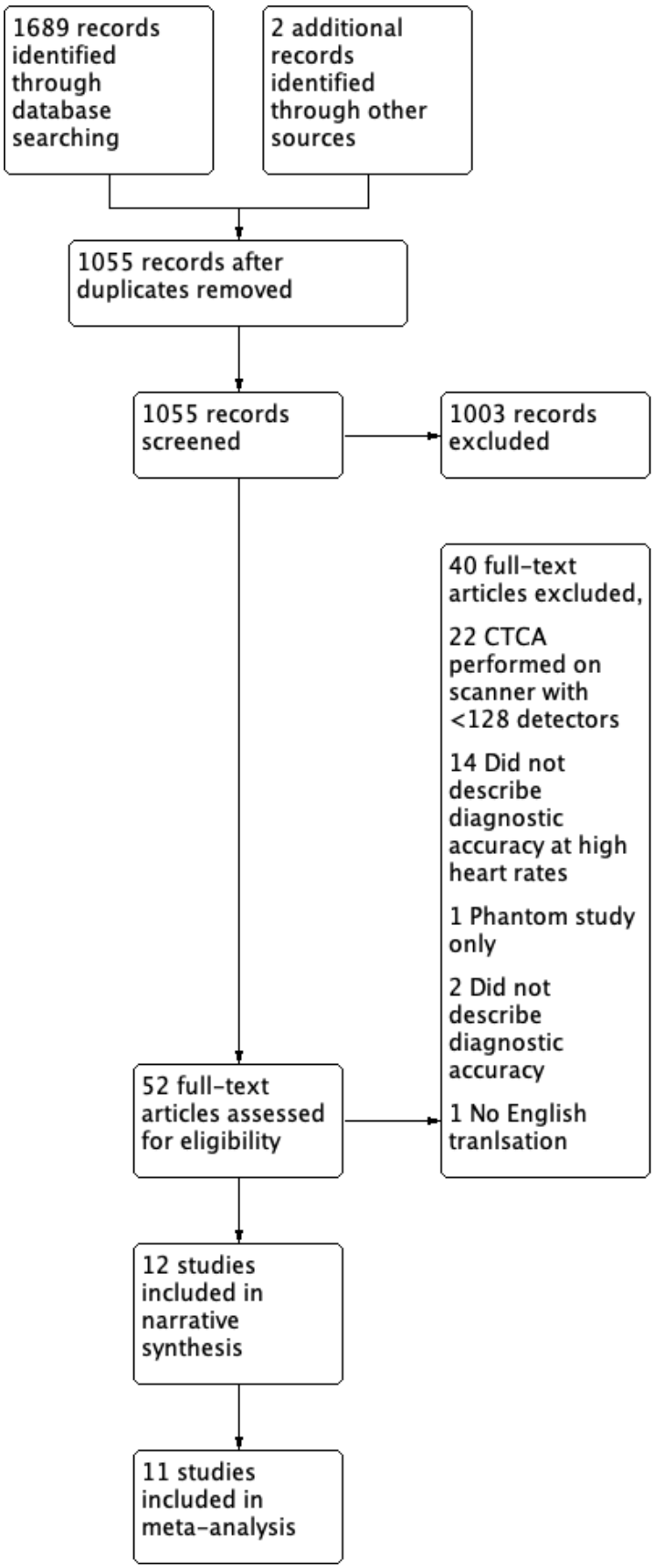

Figure 7: Search results

Appendix Two provides further details of the studies excluded on full-text examination, including the specific reason for the exclusion of each study. No studies were excluded based on methodological quality alone.

Characteristics of Included Studies

Twelve studies were included in the narrative synthesis. All included studies used a prospective cross-sectional design with consecutive patients. However, in two studies only a subgroup of patients received the reference standard.

The included studies were all from peer-reviewed journal publications. Three studies were published in *European Radiology*, two studies were published in the *International Journal of Cardiology*, and two studies in the *Journal of Computer Assisted Tomography*. The remainder were published in *Balkan Medical Journal*, *British Journal of Radiology*, *Cardiovascular Diagnosis and Therapy*, *Chinese Medical Sciences Journal*, *Clinical Radiology* and *PLoS One*. Included studies were published between 2012 and 2019. All studies were published in English language. Eight of the included studies were conducted in China, with the remainder conducted in Australia, Italy, Turkey and The Netherlands.

The primary purpose of the included studies differed, and many of the included studies included diagnostic accuracy as a secondary outcome, often for a subgroup of participants only. This was likely due to the difficulty in obtaining ethical approval to perform ICA in asymptomatic individuals due to the invasive nature of the test and the associated risk for participants.

The included studies differed in their selection of participants with respect to HR. Most studies included only patients with high HR, whereas other studies looked at diagnostic accuracy of CTCA overall and included patients with high HRs as a subgroup.

Participants

Exclusion Criteria

The *a priori* exclusion criteria differed between individual studies, but all studies excluded patients with a history of coronary artery bypass graft (CABG) and/or contraindications to CT. In some studies, it was not clear if exclusions were made prior to the commencement of the study, or during or after the study.

Heart Rate

The inclusion threshold for minimum HR differed between the included studies. The review protocol suggested in the context of CTCA, elevated HR should be defined as any HR greater than 65 bpm based on current published guidelines.[53,54] Of the included studies, only three studies used this threshold.[67,68,71] Of the remaining included studies, one study used a minimum classification threshold of 70 bpm,[70,73] four studies used a minimum classification threshold of 75 bpm,[46-48,72] and one study used 80 bpm.[45] One study was included despite using a minimum HR classification threshold of less than that described in the review protocol.[69] The study was considered by the reviewers to provide information pertinent to the review, despite the lower cut-off.

An upper inclusion cut-off point for high HR was explicitly defined in three included studies, ranging between 80 and 100 bpm. All other studies did not include a ceiling limit. However, it was possible to extract maximum range values for the majority of the included studies. Only two studies excluded this data.[45,68]

Although individual HRs were not included, studies routinely presented mean HRs, the standard deviation and HR range. These have been tabulated in Table 2 below.

Table 2: Heart rates for participants included in diagnostic accuracy assessment

Study [First Author DATE]	All Participants in Study Population (including Low HR)					High Heart Rate Group				
	No. of Participants [N]	Heart Rate [mean] (bpm)	Standard Deviation [SD]	HR Range (bpm)		No. of Participants [n]	Heart Rate [mean] (bpm)	Standard Deviation [SD]	HR Range (bpm)	
				Min.	Max.				Min.	Max.
Andreini 2018	100	N/A	N/A	N/A	N/A	40	93	±23.6	81	N/A
Gang 2012	60	73.7	±15.4	51	128	26	86.5	±15.1	73	128
Li 2013	N/A	N/A	N/A	N/A	N/A	61	75	±7.7	65	80
Liang 2017	N/A	N/A	N/A	N/A	N/A	84	82.8	±7.9	75	117
Liang 2018	N/A	N/A	N/A	N/A	N/A	64	82.5	±7.3	75	106
Liang 2019	N/A	N/A	N/A	N/A	N/A	81	83.8	±8.9	75	134
Neefjes 2013	267	65	±12	N/A	N/A	67	75	±12	65	N/A
Nerlekar 2017	107	N/A	N/A	37	80	52	69*	±8	60	80
Selçuk 2016	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Sun 2013	N/A	N/A	N/A	N/A	N/A	47	79	±9	66	100
Wang 2016	100	76.44	±13.36	39	107	60	N/A	N/A	75	107
Zhang 2016	43	69.4	±13.6	45	106	16	N/A	N/A	70	106

* result includes participants in study that were excluded from review

It was not possible to categorise specific individual HRs or HR ranges within the high HR groups. This has been reflected in the QUADAS-2 risk of bias assessment discussed below. Furthermore, the instance at which HR was measured in the study was not always described. Most studies specifically made comment that the patient's HR was recorded during the scan, whereas it was unclear in other papers. One study described HR being recorded prior to CT.[68]

Heart Rhythm

Arrhythmias (non-sinus rhythm) were an *a priori* exclusion criteria for the review where the study focused on these patients as the aim of the study. Three studies described this explicitly in their exclusion criteria, but it was not clear how this was handled in the remaining studies.

Patient Demographic

The number of patients included in each study varied, with an average of 100±61 (range 43–267). Patient age was not reported in most studies. The majority of studies (eight) were performed in institutions in China, with the remaining studies performed in Australia, Italy, the Netherlands and Turkey.

A full table of extracted data is available as Appendix Three.

Index Test

CT Scanner Make and Model

Scanners from all four major CT manufacturers met the inclusion criteria for the review. The majority of studies (four) investigated SOMATOM Definition Flash (Siemens Healthineers) and the GE Revolution CT (GE Healthcare). Two studies investigated the Aquilion ONE (Canon Medical) and another investigated the Aquilion ONE ViSION Edition model (Canon Medical). Only one included study assessed the Philips iCT (Philips Healthcare). Prospective scan modes were used in all studies; however, the specific scan parameters used varied between studies.

Reference Standard

Invasive Coronary Angiography

All included studies used ICA as the reference standard as this was an inclusion criterion in the review protocol. No weighting was given to the way ICA was performed, provided the primary study made reference to how it was performed. All included studies provided a description of how ICA was performed.

Diagnosis of Interest

Disease Significance Threshold

All studies used the American Heart Association (AHA) 15-segment model in describing vessel segments.

Diagnostic test accuracy was analysed at per-patient, per-vessel, and per-segment levels in the majority of studies. Ten studies reported patient-level sensitivity and specificity results, but only eight of these gave sufficient 2x2 table information. Nine studies reported sensitivity and specificity at the vessel level, but only five of these provided appropriate 2x2 table information. At the segment level, all 12 included studies provided sensitivity and specificity results, but only 11 reported 2x2 table data.

All included studies used the same disease significance threshold of stenosis measuring greater than 50% of the vessel lumen diameter. One of the included studies additionally described diagnostic accuracy at the level of severe CAD (>75% vessel lumen diameter). This information was not included in the review.

The prevalence of disease was high but differed between included studies (range 45%–87.5%).

Non-Diagnostic Segments

Different studies dealt differently with image quality that was uninterpretable for a particular coronary segment. Previously described processes for the assessment of uninterpretable segments suggest that these were likely to be treated as positive cases.[75]

Seven included studies treated non-diagnostic segments as positive results, three studies did not clarify how they were dealt with, and one study excluded non-diagnostic segments from the data. One of the studies that treated non-diagnostic segments as positive also reported a separate table with results where non-diagnostic segments were excluded for comparison. In this case, only the data with segments treated as positive were included in the review.

Assessment of Methodological Quality

Overall the methodological quality of the included studies was high, based on the results of the QUADAS 2 critical appraisal (Figure 8).

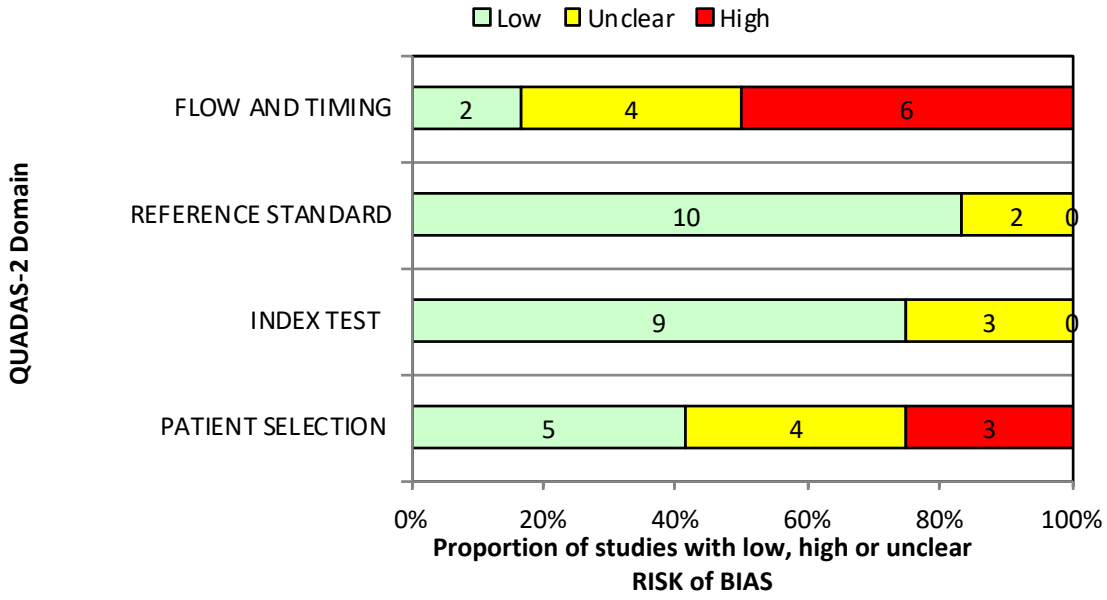

Figure 8: QUADAS-2 Risk of bias assessment

Risk of Bias

A majority of included studies were considered at high/unclear risk of bias regarding the selection of patients (signalling question Q3). This result was due to most studies only including a convenience sample of patients who had previously undergone the reference standard. The likely reason for this design was the invasiveness of the reference standard; it would have been unethical to ask patients with low or negligible risk to undergo the procedure for the study. Therefore, most studies only included those patients with a

moderate to high risk pre-test probability of CAD. This effect increases the positive rate or prevalence within the cohort, as patients who would clinically be required to undergo CTCA, but not ICA, have been excluded. It was therefore considered that this design had potential for spectrum bias, which may adversely influence the accuracy of the results.

Additionally, a significant group of studies scored a high or unclear risk for the flow and timing domain, as there was thought to be considerable risk of partial verification bias. This was because only some of the participants enrolled in the study received the reference standard.

Both spectrum bias and partial verification bias can be linked to the invasiveness of the reference standard, which makes it unable to be provided to all enrolled participants.

With regard to the performance and conduct of the reference standard itself, two studies were deemed to be of unclear risk of bias, as it was not clear that readers of the ICA (reference standard) were blinded to the results of the CTCA.

Regarding the performance and conduct of the index test (CTCA), three studies were considered to be of unclear risk of bias, as it was not evident that the readers were blinded to the results of the reference standard.

Details of the individual QUADAS-2 risk of bias scores have been tabulated below (Figure 9).

	QUADAS-2 Signalling Questions										
	Domain 1			*Domain 2*		*Domain 3*		*Domain 4*			
STUDY (first author DATE)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Andreini 2018	YES	YES	YES	YES	YES	YES	YES	UNCLEAR	NO	YES	YES
Gang 2012	YES	YES	NO	UNCLEAR	YES	YES	YES	UNCLEAR	YES	YES	YES
Li 2013	YES	YES	NO	YES	YES	YES	YES	UNCLEAR	YES	YES	UNCLEAR
Liang 2017	YES	YES	UNCLEAR	YES	YES	YES	YES	YES	YES	YES	UNCLEAR
Liang 2018	YES	YES	UNCLEAR	UNCLEAR	YES	YES	YES	YES	YES	YES	YES
Liang 2019	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES
Neefjes 2013	YES	YES	YES	YES	YES	YES	UNCLEAR	UNCLEAR	NO	YES	NO
Nerlekar 2017	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES
Selcuk 2016	YES	YES	YES	YES	YES	YES	UNCLEAR	YES	NO	YES	UNCLEAR
Sun 2013	YES	YES	UNCLEAR	YES	YES	YES	YES	UNCLEAR	YES	YES	YES
Wang 2016	YES	YES	UNCLEAR	UNCLEAR	YES	YES	YES	YES	YES	YES	NO
Zhang 2016	YES	YES	YES	YES	YES	YES	YES	UNCLEAR	NO	YES	YES

Figure 9: QUADAS-2 Risk of bias signalling questions for each included study

The individual signalling questions were then assessed for each included study and an overall risk of bias was agreed upon for each of the four domains. The proportion of studies reported with each level of risk of bias is reported in Figure 8.

Concerns Regarding Applicability

Overall, there was low concern regarding the applicability of included studies meeting the review question (Figure 10). Only one study was regarded as having high concerns in the patient selection domain.[69] This concern was due to the included study having an inclusion threshold (>60 bpm) for "high HR" patients. This minimum value was below that described in the review protocol (>65 bpm).

The flow and timing domain are not assessed in the applicability section of QUADAS-2 as a high degree of heterogeneity is expected in studies of diagnostic test accuracy.[23]

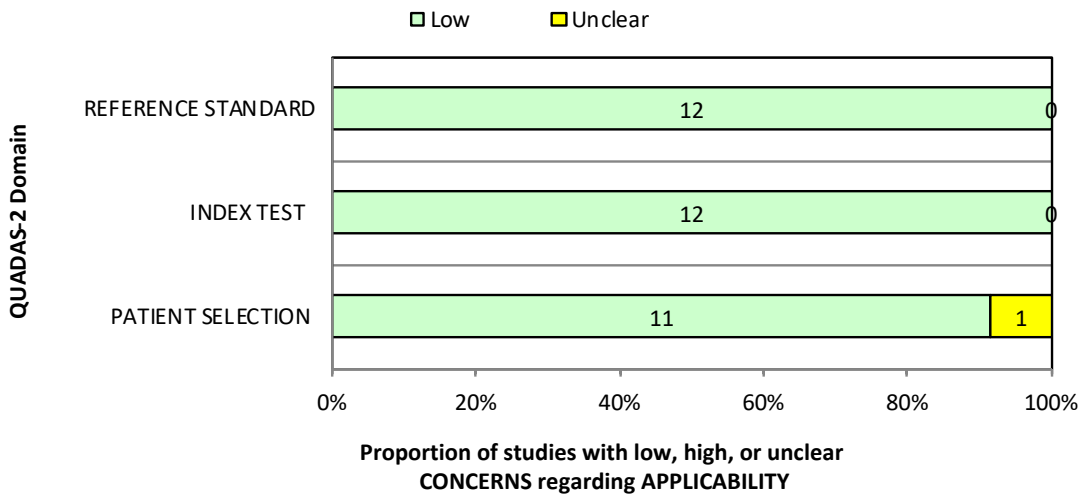

Figure 10: QUADAS-2 Concerns regarding applicability of included studies

Potential Conflicts of Interest

Of the 12 included studies, nine did not declare any conflict of interest associated with the authors. Two studies included authors who had direct affiliations with CT manufacturers. One study declared that one of the authors receives research support and is a consultant for two of the manufacturers.

Whilst conflict of interest is not a source of bias addressed in the QUADAS-2 risk of bias assessment and is not necessarily cause for concern in the systematic review, it was felt this information was important to note. This is due to the competitive nature of the high value index test and the effect negative results may have on the manufacturer's products commercially.

Findings of Review

The findings of the review have been summarised through both a narrative synthesis as well as quantitatively through meta-analysis.

Narrative Synthesis

All included studies were cross-sectional designs. All designs included consecutive samples; however, one study was a retrospective analysis. The majority of studies performed the ICA prior to CTCA as this ensured that only patients receiving both tests were being

independently assessed for diagnostic accuracy. This was likely due to the invasive nature of the reference standard.

Six primary studies were identified that directly analysed the diagnostic accuracy of patients undergoing CTCA with high HRs.[45-48,69,71] Andreini and colleagues[45] described the accuracy of CTCA using a Revolution CT (GE Healthcare) scanner in patients with high HR. The study included the use of a proprietary computer algorithm (SnapShot Freeze, GE Healthcare) to improve visualisation of the coronary arteries by reducing blur resulting from cardiac motion. Their paper directly compared the diagnostic accuracy of this method between patients with high HRs (>80 bpm) with a control group of patients with lower HRs. The authors described a sensitivity and specificity of 100% (confidence interval [CI] not reported) and 81.8% (95% CI: 65.7%, 97.9%), respectively, for patients with high HRs, compared to their finding of 91.1% (95% CI: 89%, 93.1%) and 96% (95% CI : 94.4%, 97.5%) for sensitivity and specificity, respectively, in the control group. The results from this study excluded non-diagnostic segments, in which a section was excluded from analysis if it was uninterpretable.

Liang and colleagues[46-48] described high level accuracy results in their studies of diagnostic test accuracy for patients undergoing CTCA with a Revolution CT (GE Healthcare) scanner utilising the same motion reduction algorithm technology. A 2017 preliminary study observed 84 patients with high HRs undergoing CTCA, with the algorithm employed or not as preferred by the radiologist.[48] The study found this method achieved a sensitivity of 100% (95% CI : 93.6%, 100%) and specificity of 85.7% (95% CI : 67.3%, 96.0%). The authors concluded that the incorporation of SnapShot Freeze improved the accuracy of CTCA.

A 2018 follow-up study by the same authors directly compared the impact of viewing images from a GE Revolution CT scanner with and without the motion correction algorithm applied.[47] The study found a sensitivity of 100% (95% CI: 91.8%, 100%) and specificity of 85.7% (95% CI : 63.7%, 97.0%) when the images were analysed using the SnapShot Freeze software. This is compared to sensitivity of 100% (95% CI: 91.8%, 100%) and specificity of 14.3% (95% CI: 3.1%, 36.3%) without the algorithm applied.

A newer iteration of the same algorithm was also investigated by the aforementioned authors in a recent (2019) study.[46] The results were similar to those of their 2018 study. The most recent study analysed sensitivity and specificity. Without SnapShot Freeze, sensitivity was 100.0% and specificity 22.2%; with the original algorithm applied, sensitivity was 100.0% and specificity 50.0%; and with the more recent iteration, sensitivity was 100.0% and specificity 77.8%. Confidence intervals were not reported in this study. Whilst sensitivity did not vary, the study found considerable improvement in the specificity of the scan with iterations of SnapShot Freeze applied. Moreover, an area under the curve (AUC) measurement for the receiver operating characteristic (ROC) curve resulted in 0.61, 0.75 and 0.91 for the standard reconstruction, SnapShot Freeze version 1, and SnapShot Freeze version 2, respectively.

A 2017 study by Nerlekar and colleagues[69] measured the accuracy of CTCA using a Canon (Aquilion ONE ViSION Edition) scanner. The study made a direct comparison between high HRs (between 60 and 80 bpm) and a control group (less than 60 bpm). The authors found that diagnostic accuracy was comparable between the two groups, with a sensitivity of 100% (95% CI : 90%, 100%) and specificity of 88% (95% CI: 64%, 99%) in the 60 to 80 bpm group, compared with sensitivity of 97% (95% CI: 86%, 100%) and sensitivity of 88% (95% CI: 64%, 99%) in the low HR control group.

In their 2013 study, Sun and colleagues[71] investigated the diagnostic accuracy of a Siemens (Flash) dual-source CT scanner using a prospective scan mode for patients with elevated HRs (66–100 bpm). Sensitivity and specificity results for patients with high HRs were 100% (95% CI : 88.0%, 100%), and 63.6% (95% CI: 31.6%, 87.6%), respectively.

Four other studies that reported patient-level diagnostic accuracy data were included in the review. Whilst these studies did not directly study high HRs, they contained data pertinent to the review question. A 2013 study by Li and colleagues[67] introduced a novel approach of cardiac gating, using doppler ultrasound rather than traditional ECG signal. The diagnostic accuracy of an Aquilion ONE (Canon Medical) CT system when utilising this method was compared with diagnostic accuracy from traditional ECG gated scanning. A subset of participants with elevated HRs (65-80 bpm) was included in the evaluation. For patients

with high HRs in the traditional prospective ECG-gated arm, sensitivity was 97% (95% CI: 84.7%, 99.5%) and specificity was 89.3% (95% CI: 72.8%, 96.3%).

Neefjes and colleagues[68] compared different ECG gating protocols in their 2013 study, assessing Siemens' (SOMATOM Definition Flash, Siemens Healthcare,) systems. A subgroup analysis investigated the value of the method for patients with HR greater than 65 bpm. Patients were randomised to receive either prospectively or retrospectively ECG-gated scans. These were compared to a group with HRs less than 65 bpm, with participants randomly assigned both scan protocol groups also. Of this, a subset again received a clinically mandated ICA, and diagnostic accuracy was calculated. The resulting sensitivity and specificity of the high HR group that received a prospective scan was 100% (95% CI: 93.0%, 100%) and 63% (95% CI: 35%, 85%), respectively.

Dual-source CTCA scanning was also the focus of a more recent (2016) study by Selçuk and colleagues.[70] This study, which investigated image quality and diagnostic accuracy of SOMATOM Definition Flash (Siemens Healthineers) at various HRs, gave a differing result with a sensitivity of 87.8% and specificity of 99.2%. Confidence intervals were not reported. The statistical power of this study was likely to be limited, as only a small subgroup received the reference test, and a smaller group again had HRs above a high HR threshold (70 bpm). Furthermore, the study did not report 2x2 table information and the author did not respond to a request to provide this information. Subsequently this study was excluded from meta-analysis.

Zhang and colleagues[73] reported high level of accuracy in their study of a small group of participants undergoing CTCA with a SOMATOM Definition Flash (Siemens Healthineers). A subgroup analysis was performed for patients who had a HR greater than 70 bpm. For this group, sensitivity was reported as 100% (95%CI: 73.2%, 100%) and specificity as 100% (95% CI: 19.8%, 100%).

Overall, the literature provides evidence of high sensitivity results for patients with high HRs with point estimates of sensitivity varying between 87.8% and 100%. Specificity generally performs slightly less well, at 63% to 100%.

The remaining two studies included in the review reported diagnostic accuracy at segment level only. Gang and colleagues[76] enrolled consecutive patients with high-risk CAD to receive clinically mandated CTCA and then ICA. A subgroup of enrolled patients with HR greater than 70 bpm was reported, and this data was collected for the review. Whilst the authors reported patient-level accuracy results overall, this was not reported for the high HR subgroup.

Similarly, Wang and colleagues[72] reported segment level data in their assessment of patients with various heart rates. Three subgroups of patients were compared: patients with low heart rates less than 75 bpm, patients with heart rates between 75 and 90 bpm and those with heart rates greater than 90 bpm. The authors reported sensitivity and specificity in the latter two groups as 96.0% and 93.70%; and 97.60% and 92.20%, respectively. Confidence intervals were not provided for these results.

Detailed narrative synthesis of the vessel- and segment-level accuracy has not been provided for each study as it does not directly affect the patient outcome but is summarised in Table 3 below. Further detail is provided in Appendix Three, part iii.

Table 3: Reported sensitivity and specificity by each included study

Study		Patient level		Vessel Level		Segment Level	
		Sens (%) (95%CI)	***Spec (%) (95%CI)***	***Sens (%) (95%CI)***	***Spec(%) (95%CI)***	***Sens (%) (95%CI)***	***Spec (%) (95%CI)***
Andreini 2018		100*	81.8 (65.7–97.9)*	N/A	N/A	95.2 (93.6–96.9)	98.9 (98.1–99.7)
Gang 2012		N/A	N/A	N/A	N/A	94.6 (85.13–98.88)	97 (94.38–98.62)
Li 2013		*97 (84.7–99.5)*	*89.3 (72.8, 96.3)*	*91.1(79.3–96.5)*	*96.5(93.0–98.3)*	*95.5 (90.9–97.8)*	*98.0 (96.7–98.8)*
Liang 2017		*100 (93.6–100)*	*85.7 (67.3–96.0)*	95.2 (89.2–98.4)	93.5 (89.5–96.3)	91.5 (85.8–95.5)	95.6 (94.0–96.8)
Liang 2018		100 (91.8–100)	85.7 (63.7–97.0)	*96.2 (89.3–99.2)*	*94.3 (89.9–97.3)*	*91.9 (85.2–96.2)*	*95.8 (94.1–97.2)*
Liang 2019		*100****	*85.7****	96.6***	96.6***	92.2***	97.8***
Neefjes 2013		100 (93.0–100 95)	63 (35–85)	99 (96–100)	84 (78–89)	93 (88–98)	93 (91–95)
Nerlekar 2017		100 (90–100)	88 (64–99)	*98 (91–100)*	*94 (89–97)*	*84 (76–90)*	*96 (94–97)*
Selçuk 2016		*87.8****	*88****	81.4***	95***	87.8***	99.2***
Sun 2013		*100 (88.0–100)*	*63.6 (31.6–87.6)*	90.0 (81.4–95.0)	95.2 (91.9– 97.2)	92.6 (86.1–96.4)	97.0 (95.1–98.2)
Wang 2016	(HR 70–90 bpm)	N/A	N/A	*N/A*	*N/A*	*96.00****	*93.70****
	(HR>90 bpm)	*N/A*	*N/A*	*N/A*	*N/A*	*97.60****	*92.20****
Zhang 2016		*100 (73.2–100)*	*100 (19.8–100)*	96.4 (80.0–99.8)	91.7 (76.4–97.8)	88.6 (74.6–95.7)	90.8 (84.8–94.7)

* result based on evaluable segments only (non-diagnostic segments excluded)

** unclear if indeterminate results treated as positive

*** 95% confidence interval not reported

The diagnostic test accuracy of the review was further explored through a coupled forest plot meta-analysis.

Meta-Analysis

Of the 12 included studies, 11 were included in meta-analysis. One study was excluded as there was insufficient raw (2x2 table) data available to allow data extraction.[70] The authors were contacted to request further information, but no response was given.

Overall, sensitivity and specificity values were high for all included studies. Paired forest plots for pooled sensitivity and specificity were produced, if 2x2 table data could be extracted. These were created for each of the per-patient, per-vessel and per-segment analyses (Figures 12, 13, 14, respectively). Pooled data included 450 participants, 1229 vessel analyses and 8144 vessel-segment analyses. These have been tabulated below with the associated disease prevalence at each analysis level (Table 4). The pooled sensitivity and specificity were calculated using a random effects model as described in the methods chapter.

Table 4: Number included in analysis with disease prevalence

	Total number included in the meta-analysis	Disease Prevalence (TP+FN)
Patient Level	450 patients	69.2% (315)
Vessel Level	1229 vessels	24.7% (371)
Segment Level	8144 segments	19.8% (1345)

Per-Patient Diagnostic Accuracy

The summary estimates for pooled sensitivity and specificity were 99% (95%CI: 97%, 100%) and 79% (95%CI: 72%, 85%), respectively (Figure 11).

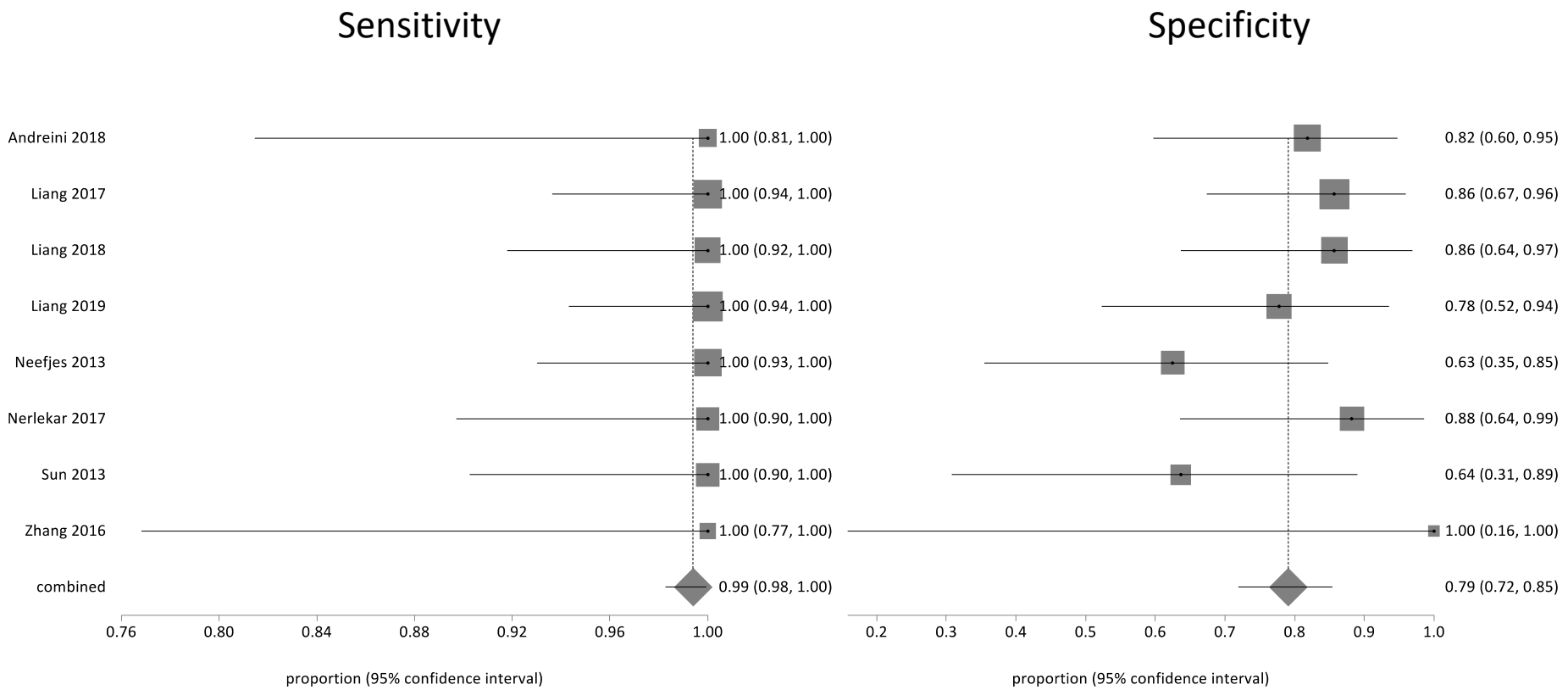

Figure 11: Paired forest plot of diagnostic accuracy at per-patient level

There was little heterogeneity visible in the sensitivity meta-analysis. The point estimates for each included study for specificity varied considerably more; however, the summary estimate for specificity in the review was within the 95% confidence interval for all included studies. However, confidence intervals were large for most of the included studies that reported specificity data. Three studies contained confidence intervals less than the 0.5 threshold, below which a diagnostic test is considered ineffective. The confidence intervals for the summary estimates remain relatively small. The percentage of statistical weight given to each study in the random effect meta-analysis are provided in Table 5. StatsDirect determined an I^2 statistic as 0% (95% CI: 0%, 56%) for both sensitivity and specificity. This indicates no statistical heterogeneity was detected.

Table 5: Statistical weight for each study reporting data at patient level

Study	Sensitivity – % Weight [Random Effects]	Specificity – % Weight [Random Effects]
Andreini 2018	5.799373	16.18705
Liang 2017	17.711599	20.503597
Liang 2018	13.636364	15.467626
Liang 2019	19.905956	13.309353
Neefjes 2013	16.144201	11.870504
Nerlekar 2017	10.815047	12.589928
Sun 2013	11.442006	8.273381
Zhang 2016	4.545455	1.798561

Per-Vessel Diagnostic Accuracy

The summary estimates for pooled sensitivity and specificity at the per-vessel level were 96% (95% CI: 93%, 97%); and 93% (95% CI : 90%, 96%), respectively (Figure 12).

There was some heterogeneity in the vessel-level coupled forest. The summary estimate for vessel-level sensitivity was within the 95% CI limits of all but one of the pooled studies. In the specificity forest plot, two studies' CIs did not cross the summary estimate point, indicating some heterogeneity. Table 6 shows the relative statistical weight associated with each included study in the analysis. StatsDirect calculated an I^2 statistic of 27.1% (95% CI: 0%, 68.8%) for sensitivity and 74.7% (95% CI: 31.5%, 86.4%) for specificity. This indicates that significant heterogeneity was detected in the model.

Table 6: Statistical weight for each study reporting accuracy at vessel level

Study	Sensitivity – % Weight [Random Effects]	Specificity – % Weight [Random Effects]
Liang 2017	17.191234	15.994305
Liang 2018	14.583256	14.983664
Liang 2019	18.953542	15.570011
Neefjes 2013	15.366133	15.026069
Nerlekar 2017	11.568324	14.221447
Sun 2013	15.994701	16.631055
Zhang 2016	6.34281	7.57345

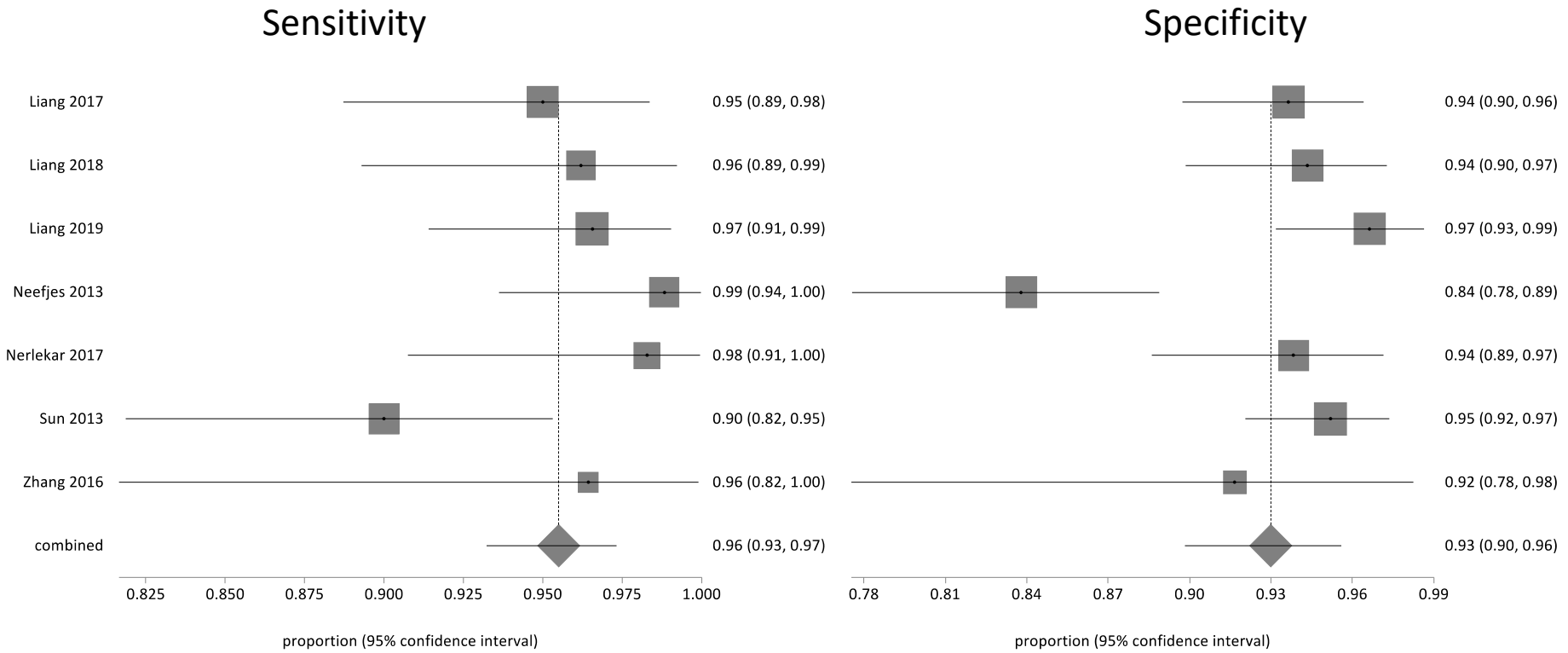

Figure 12: Paired forest plot of diagnostic accuracy at per-vessel level

Per-Segment Diagnostic Accuracy

The summary estimates for pooled sensitivity and specificity at the per-segment level were 91% (95% CI: 88%, 93%) and 96% (95% CI: 95%, 98%), respectively (Figure 13).

The proportion meta-analysis for sensitivity at the segment level indicated some heterogeneity, with 95% confidence intervals for two studies lying outside of the summary estimate for the analysis. For the specificity analysis, statistical heterogeneity was also present, with five of the included studies' 95% confidence intervals outside the summary estimate point.

StatsDirect reported an I^2 statistic of 54.5% (95% CI: 0%, 75.3%) for the sensitivity plot (indicating moderate heterogeneity) and 86.7% (95% CI: 77.9%, 91.0%) for the specificity plot (indicating high level of heterogeneity).

Table 7 shows the statistical weighting associated with each study in the analysis

Table 7: Statistical weight for each study reporting accuracy at segment level

Study	Sensitivity – % Weight [Random Effects]	Specificity – % Weight [Random Effects]
Andreini 2016	6.684278	9.250021
Gang 2012	6.273437	8.193499
Li 2013	10.563344	9.530737
Liang 2017	10.536283	9.744316
Liang 2018	9.16925	9.468878
Liang 2019	11.454812	9.739288
Neefjes 2013	9.208062	9.462344
Nerlekar 2017	9.010271	9.480153
Sun 2013	9.577042	9.15271
Wang 2016	12.173432	9.30374
Zhang 2016	5.349788	6.674315

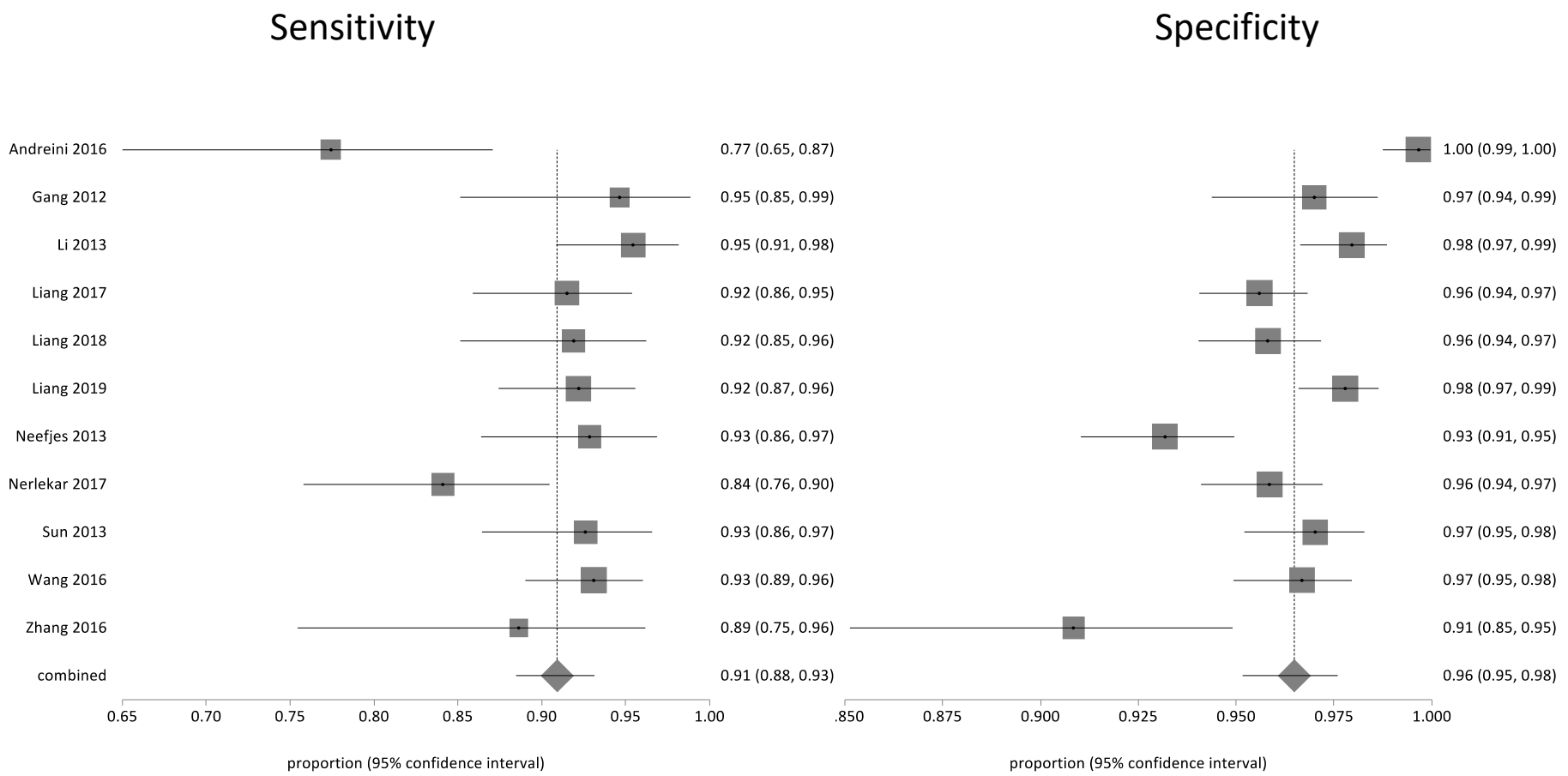

Figure 13: Paired forest plot of diagnostic accuracy at per-segment level

Visual inspection of the coupled forest plot at the patient level shows a large amount of heterogeneity in the specificity results between studies. The magnitude of this effect is reduced at the per-vessel level and further still at the per-segment level. This is likely because the severity of disease decreases with decreasing prevalence of disease at the per-vessel and per-segment level, respectively. That is, in a patient with fewer diseased segments, the severity of disease is likely to be lower. Therefore, heterogeneity in results is less likely. Disease spectrum should not be confused with disease prevalence. Prevalence does not have a direct effect on sensitivity and specificity, and is therefore an independent factor,[77] unlike PPV and NPV.

Sensitivity Analysis

A sensitivity analysis was performed to assess potential heterogeneity associated with the disparities in the reporting of unevaluable segments between studies. The overall analysis included all studies that reported 2x2 table data at high HRs. However, the decision was made to include unevaluable segments as positive when this option was available (either by recalculation by the reviewer or as reported). This was under the advice that patients who received equivocal results from motion blur, etc, would in clinical practice be treated as positive and referred for further testing.[78] A sensitivity analysis was performed to assess the effect of removing studies that excluded non-assessable segments from their analysis on the test accuracy of this review.

Figure 14 shows a forest plot at the segment level, in which only studies that reported data with non-evaluable segments were treated as positive. The resultant sensitivity and specificity are 90% (95% CI: 86%, 93%) and 96% (95% CI: 94%, 98%), compared with 92% and 96%, respectively, in the overall analysis.

The statistical weighting of the random effects model data is detailed in Table 8. The calculated I^2 statistic was 56% (95% CI: 0%, 78.2%) for sensitivity and 89.7% (95% CI: 82.2%, 93.1%) for specificity. This indicated a high level of statistical heterogeneity in the analysis.

Table 8: Statistical weight for each study reporting accuracy at segment level (non-evaluable segments treated as positive)

Study	Sensitivity – % Weight [Random Effects]	Specificity – % Weight [Random Effects]
Andreini 2016	10.334696	12.853564
Gang 2012	9.761718	11.769325
Liang 2018	13.650398	13.069378
Liang 2019	16.489835	13.332061
Neefjes 2013	13.700233	13.062977
Nerlekar 2017	13.445661	13.080418
Sun 2013	14.17114	12.756666
Zhang 2016	8.446318	10.075611

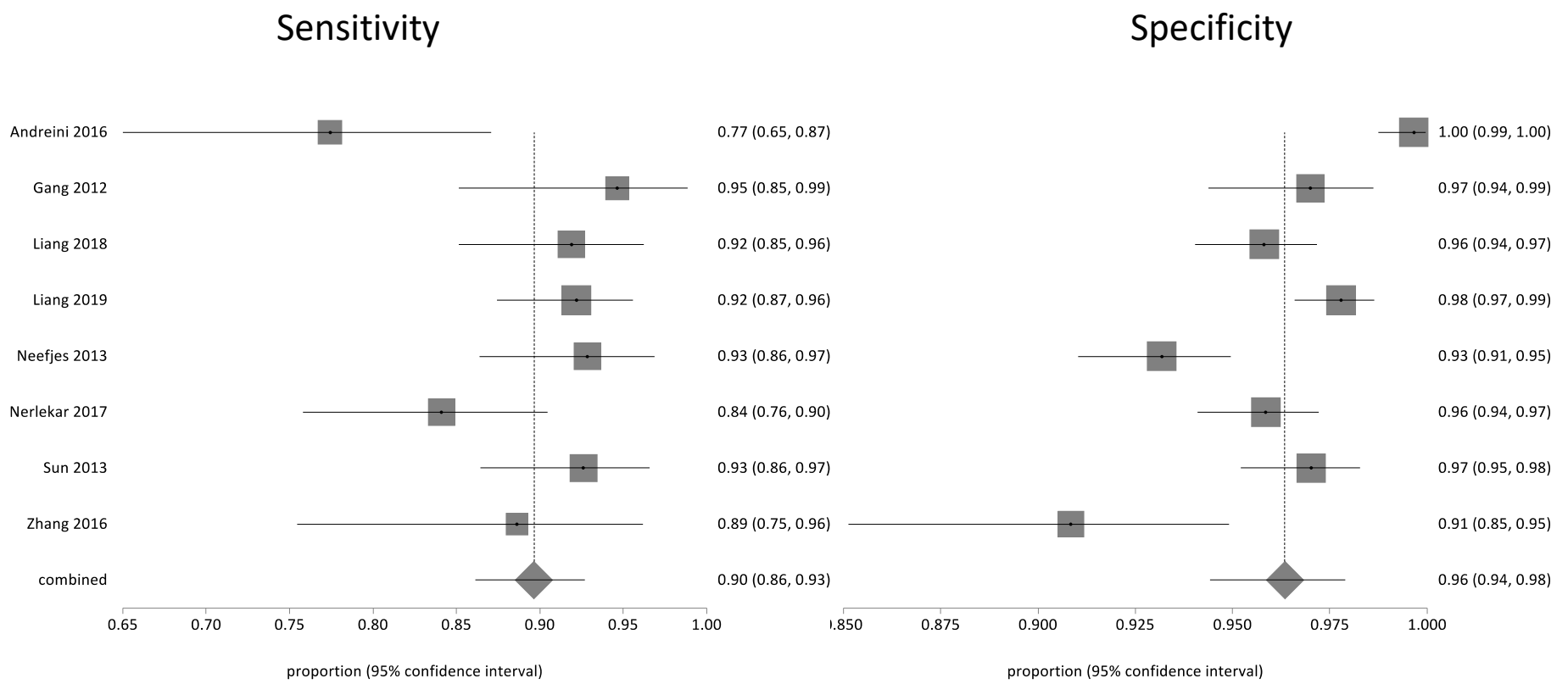

Figure 14: Sensitivity analysis at segment level with indeterminate segments treated as positive

Figure 15 shows a forest plot with studies that report accuracy with unevaluable segments excluded from their analysis. The resultant sensitivity and specificity are 93% (95% CI: 90%, 95%) and 98% (95% CI: 97%, 99%), respectively.

The statistical weighting of the random effects model data is detailed in Table 9. The calculated I^2 statistic was 0% (95% CI: 0%, 72.9%) for the sensitivity plot and 82.9% (95% CI: 0%, 92.6%) for the specificity plot. Whilst this estimate suggests no heterogeneity in the sensitivity forest plot, the associated 95% CIs indicate that there is insufficient statistical power to give a meaningful result.

Table 9: Statistical weight for each included study reporting accuracy at segment level (non-evaluable treated as excluded)

Study	Sensitivity % Weight [random effects]	Specificity % Weight [random effects]
Andreini 2016	10.691824	33.467121
Sun 2013	30.81761	32.869361
Wang 2016	58.490566	33.663518

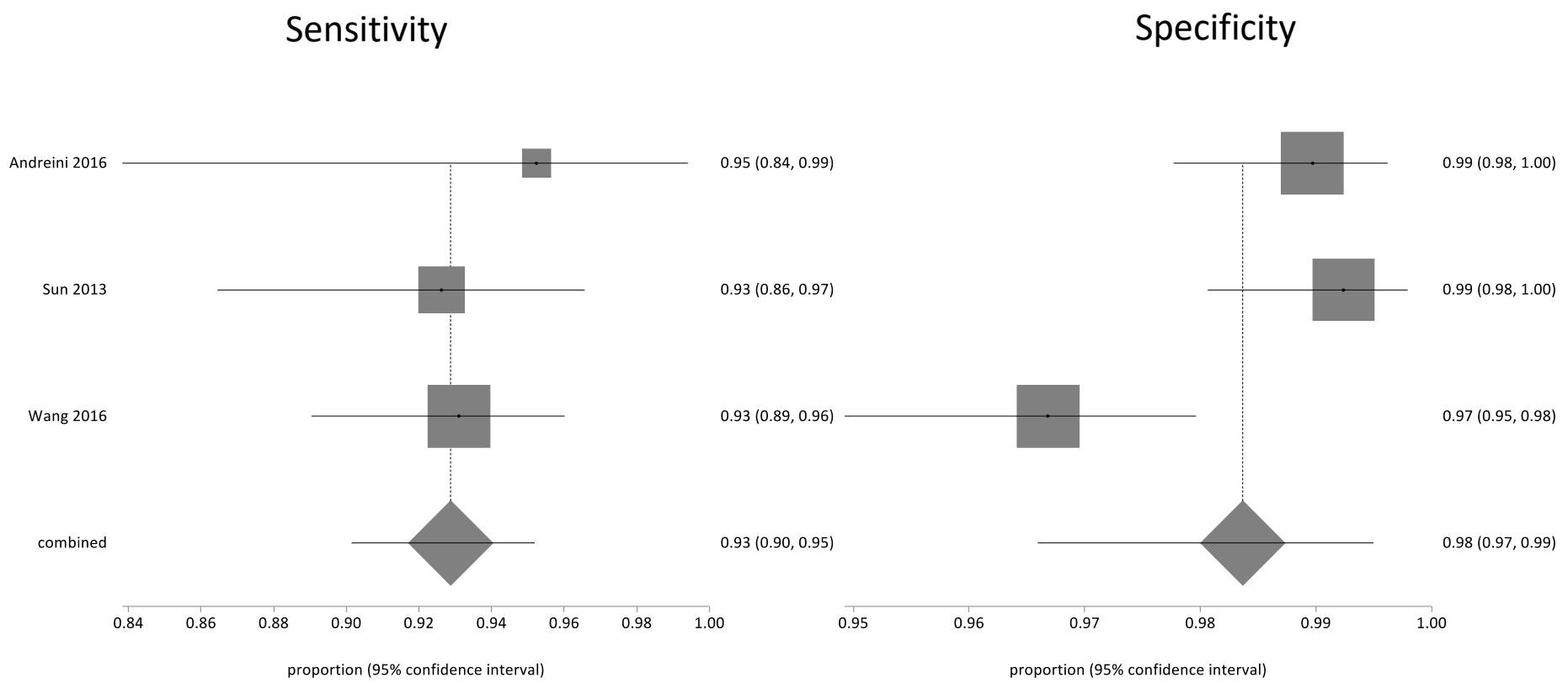

Figure 15: Sensitivity analysis at segment level with indeterminate segments excluded from analysis

Subgroup Analysis

To better understand the potential heterogeneity between the specificity values reported in the studies, a subgroup analysis was performed. Expected causes of heterogeneity were the differences in make and model of the scanner and the variation in the HR thresholds applied for each study. These two possible sources of heterogeneity were considered *a priori*. Therefore a subgroup analysis of these two factors was planned at the review protocol stage.[59]

Accuracy at different heart rate thresholds

No studies specifically reported a per-patient description of HR, preferring to group the data. Authors were contacted to provide this information; however, there was insufficient response to provide further detail. Consequently, a subgroup analysis of high HR groups at different described minimum cut-off points was performed to test for possible sources of clinical heterogeneity. Because individual HRs were not provided, it was not possible to determine the individual participants' HRs within a group. Figure 16 shows variation in the minimum (and maximum if reported) HRs for patients between each study.

A mean HR was described in almost all included studies. A subgroup analysis was therefore performed based on the minimum HR recorded for each study (Figure 17).

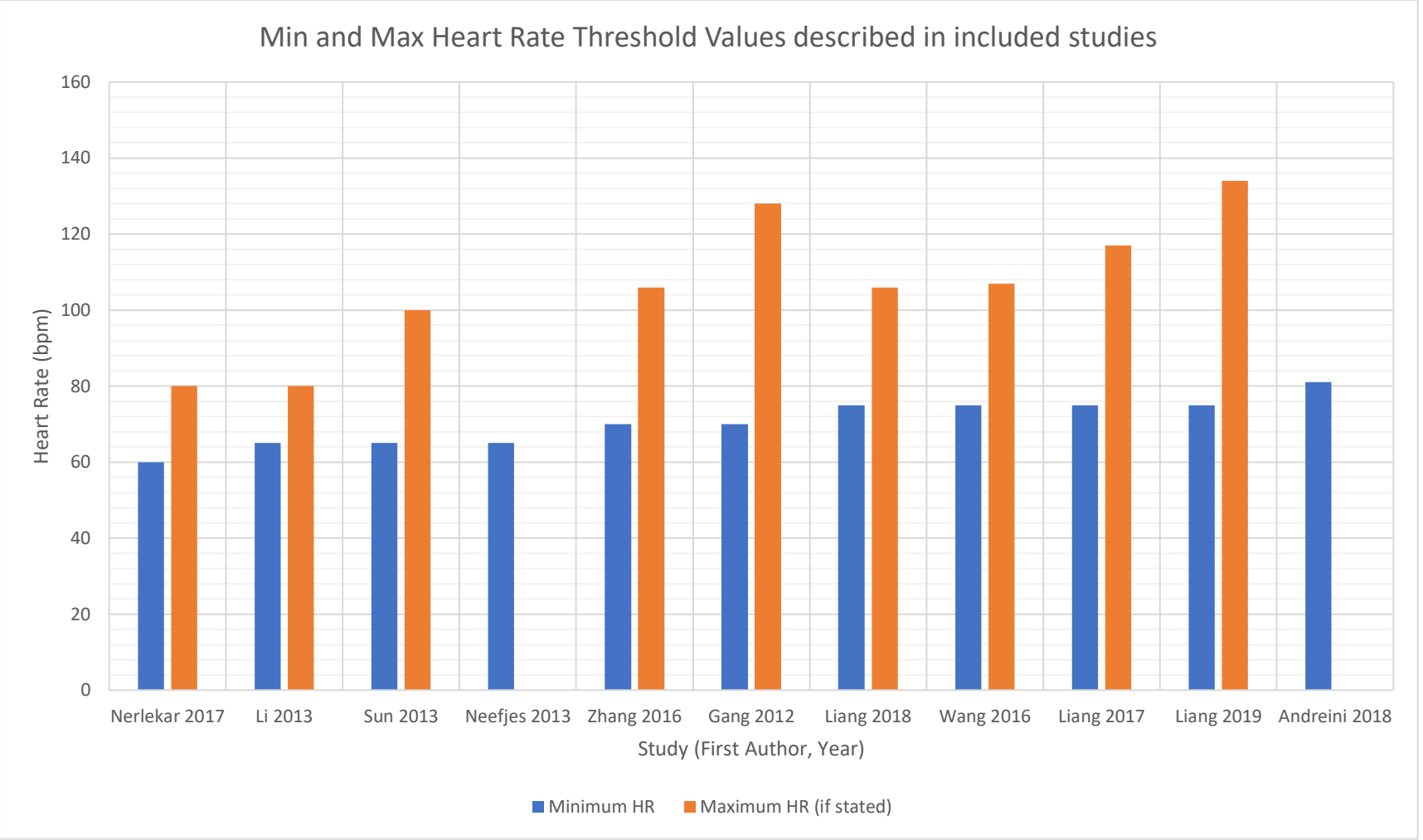

Figure 16: Range of heart rates for high heart rate groups in each study

HR >60bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Nerlekar 2017	34	2	0	15	1.00 [0.90, 1.00]	0.88 [0.64, 0.99]

HR >65bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Neefjes 2013	51	6	0	10	1.00 [0.93, 1.00]	0.63 [0.35, 0.85]
Sun 2013	36	4	0	7	1.00 [0.90, 1.00]	0.64 [0.31, 0.89]

HR >70 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Zhang 2016	14	0	0	2	1.00 [0.77, 1.00]	1.00 [0.16, 1.00]

HR >75bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Liang 2017	56	4	4	24	0.93 [0.84, 0.98]	0.86 [0.67, 0.96]
Liang 2018	43	3	0	18	1.00 [0.92, 1.00]	0.86 [0.64, 0.97]
Liang 2019	63	4	0	14	1.00 [0.94, 1.00]	0.78 [0.52, 0.94]

HR >80bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Andreini 2018	18	4	0	18	1.00 [0.81, 1.00]	0.82 [0.60, 0.95]

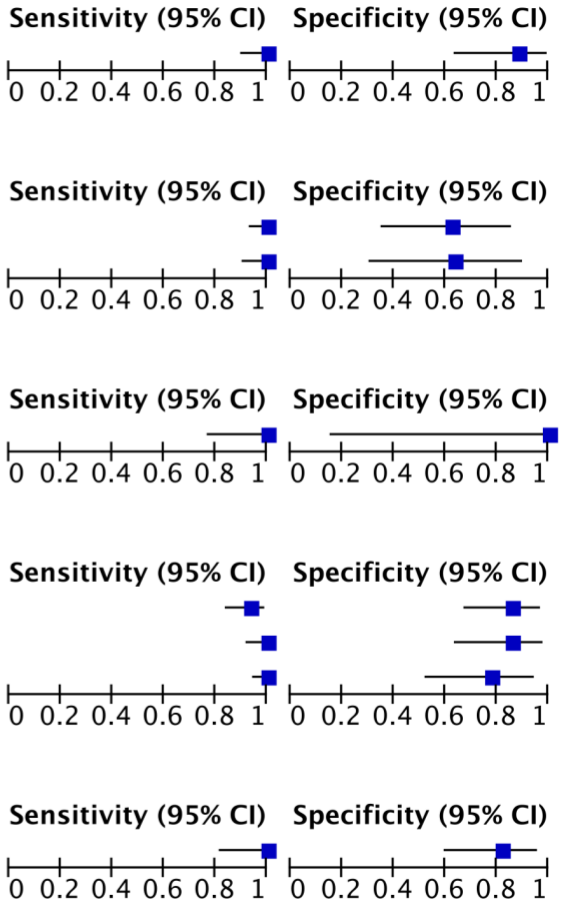

Figure 17: Paired forest plot subgroup analysis of minimum heart rate thresholds

This analysis contained an insufficient sample size to achieve a definitive result. For this reason, it was felt that subgroup results should not be pooled. Therefore, heterogeneity statistics were not included in the subgroup analysis, and assessment of heterogeneity was possible through visual inspection only.

As minimum HR levels increase there appears to be a larger spread of corresponding confidence intervals for specificity. It could be argued, therefore, that the behaviour of specificity becomes more erratic with increasing HR. This finding would agree with anecdotal evidence that, as HR increases, the image quality and hence diagnostic accuracy is reduced. An increase in HR has little influence on sensitivity, however. A specific value at which the diagnostic accuracy of current generation scanners declines could not be determined from the available data.

A segment subgroup analysis was also performed at the coronary artery segment level (Figure 18). These results show a more consistent specificity at higher HR thresholds and do not reflect the apparent inconsistency in the per-patient level assessment. This is likely due to the increased sample size providing additional statistical power to the result, therefore narrowing the confidence intervals.

HR >60 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Nerlekar 2017	90	29	17	671	0.84 [0.76, 0.90]	0.96 [0.94, 0.97]

HR>65 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Li 2013	147	15	7	718	0.95 [0.91, 0.98]	0.98 [0.97, 0.99]
Neefjes 2013	104	47	8	642	0.93 [0.86, 0.97]	0.93 [0.91, 0.95]
Sun 2013	113	16	9	521	0.93 [0.86, 0.97]	0.97 [0.95, 0.98]

HR >70 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Gang 2012	53	9	3	291	0.95 [0.85, 0.99]	0.97 [0.94, 0.99]
Zhang 2016	39	14	5	139	0.89 [0.75, 0.96]	0.91 [0.85, 0.95]

HR >75 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Liang 2017	140	40	13	869	0.92 [0.86, 0.95]	0.96 [0.94, 0.97]
Liang 2018	102	29	9	664	0.92 [0.85, 0.96]	0.96 [0.94, 0.97]
Liang 2019	113	16	9	521	0.93 [0.86, 0.97]	0.97 [0.95, 0.98]
Wang 2016	133	14	9	514	0.94 [0.88, 0.97]	0.97 [0.96, 0.99]

HR >80 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Andreini 2018	48	2	14	576	0.77 [0.65, 0.87]	1.00 [0.99, 1.00]

HR >90 bpm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Wang 2016	83	6	7	246	0.92 [0.85, 0.97]	0.98 [0.95, 0.99]

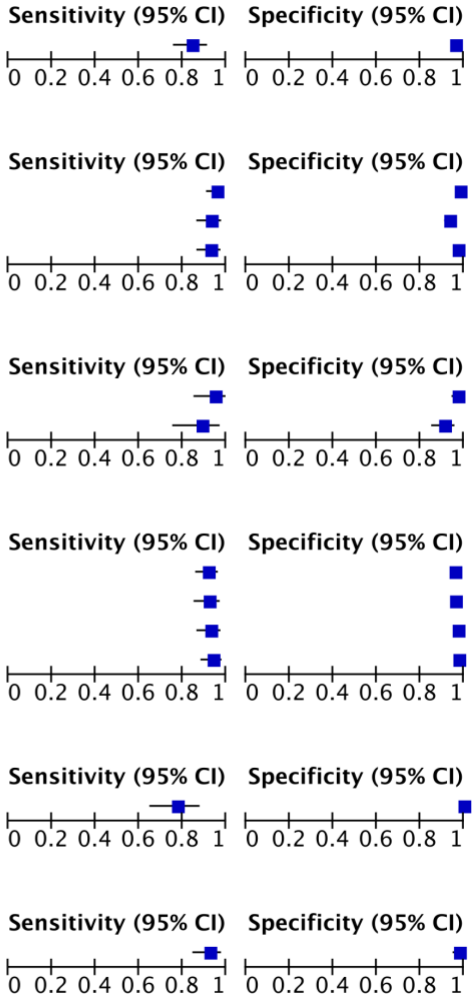

Figure 18: Per-segment subgroup analysis of minimum heart rate thresholds

Comparison of accuracy between manufacturers

Subgroup analysis was also performed to look at the accuracy for different makes and models of scanner. Figure 19 shows a subgroup analysis at patient level.

Canon scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Nerlekar 2017	34	2	0	15	Aquilion ONE Vision	1.00 [0.90, 1.00]	0.88 [0.64, 0.99]

GE scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Andreini 2018	18	4	0	18	Revolution CT	1.00 [0.81, 1.00]	0.82 [0.60, 0.95]
Liang 2017	56	4	0	24	Revolution CT	1.00 [0.94, 1.00]	0.86 [0.67, 0.96]
Liang 2018	43	3	0	18	Revolution CT	1.00 [0.92, 1.00]	0.86 [0.64, 0.97]
Liang 2019	63	4	0	14	Revolution CT	1.00 [0.94, 1.00]	0.78 [0.52, 0.94]

Siemens scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Neefjes 2013	51	6	0	10	Somatom Definition Flash	1.00 [0.93, 1.00]	0.63 [0.35, 0.85]
Sun 2013	36	4	0	7	Somatom Definition Flash	1.00 [0.90, 1.00]	0.64 [0.31, 0.89]
Zhang 2016	14	0	0	2	Somatom Definition Flash	1.00 [0.77, 1.00]	1.00 [0.16, 1.00]

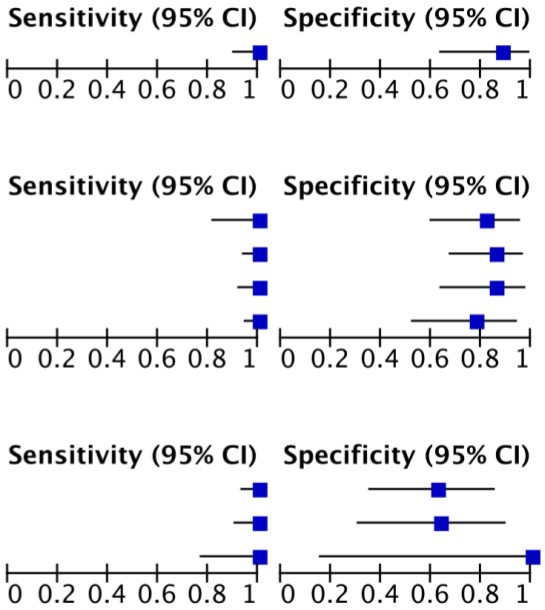

Figure 19: Per-patient subgroup analysis of scanner make and model

Visual inspection of the coupled forest plots reveals some variation in the accuracy between manufacturers. However, the estimate in the Siemens scanner subgroup is believed unlikely to represent a significant finding, as the studies in this subgroup were statistically underpowered. Furthermore, the studies in the Siemens scanner subgroup are the oldest in the analysis, and may therefore contain unseen heterogeneity. For this reason, test accuracy results according to manufacturer were not pooled prior to being compared.

The patient-level subgroup analysis did not include sufficient data to provide conclusive results. Therefore, the decision was made to run the analysis at the segment level also (Figure 20).

All four manufacturers were included in this analysis; however, there was still insufficient data to directly compare subgroups. Visual inspection of the paired forest plot suggests that there is little if any difference between the sensitivity and specificity values for each make and model at the segment level.

Canon scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Gang 2012	53	9	3	291	Aquilion ONE	0.95 [0.85, 0.99]	0.97 [0.94, 0.99]
Li 2013	147	15	7	718	Aquilion ONE	0.95 [0.91, 0.98]	0.98 [0.97, 0.99]
Nerlekar 2017	90	29	17	671	Aquilion ONE Vision	0.84 [0.76, 0.90]	0.96 [0.94, 0.97]

GE scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Andreini 2018	48	2	14	576	Revolution CT	0.77 [0.65, 0.87]	1.00 [0.99, 1.00]
Liang 2017	140	40	13	869	Revolution CT	0.92 [0.86, 0.95]	0.96 [0.94, 0.97]
Liang 2018	102	29	9	664	Revolution CT	0.92 [0.85, 0.96]	0.96 [0.94, 0.97]
Liang 2019	177	20	15	884	Revolution CT	0.92 [0.87, 0.96]	0.98 [0.97, 0.99]

Philips scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Wang 2016	216	20	16	583	Brilliance iCT	0.93 [0.89, 0.96]	0.97 [0.95, 0.98]

Siemens scanners

Study	TP	FP	FN	TN	Model of scanner	Sensitivity (95% CI)	Specificity (95% CI)
Neefjes 2013	104	47	8	642	Somatom Definition Flash	0.93 [0.86, 0.97]	0.93 [0.91, 0.95]
Sun 2013	113	16	9	521	Somatom Definition Flash	0.93 [0.86, 0.97]	0.97 [0.95, 0.98]
Zhang 2016	39	14	5	139	Somatom Definition Flash	0.89 [0.75, 0.96]	0.91 [0.85, 0.95]

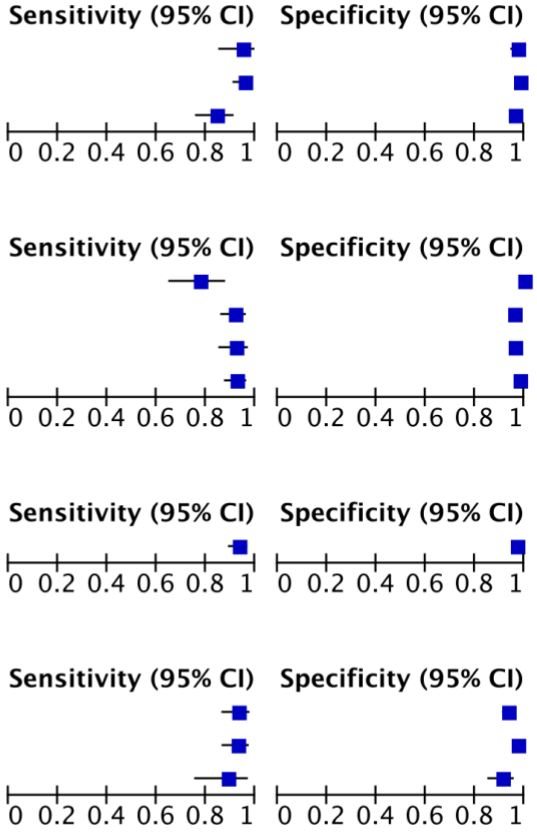

Figure 20: Segment-level subgroup analysis of individual manufacturer scanners

Summary of Findings

The quality of included evidence has been graded and summarised using the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach for diagnostic test accuracy studies (Table 10).

Table 10: GRADE summary of findings table

Pooled sensitivity: 99% (95% CI: 97%, 100%) | **Pooled specificity**: 79% (95% CI: 72%, 85%)

Test Result	Number of Results per 1000 Patients Tested (95% CI) **Prevalence 69.2%** (as calculated from included studies)	Number of Participants (Studies)	Certainty of the Evidence (GRADE)
True positives	**685** (671 to 692)	461 (11)	⨁⨁◯◯ **LOW** [a,b]
False negatives	**7** (0 to 21)		
True negatives	**213** (148 to 268)	461 (11)	⨁◯◯◯ **VERY LOW** [a,b,c,d]
False positives	**95** (40 to 160)		

CI: Confidence interval

Explanations

a. In three of 11 studies, inappropriate exclusions were made, in four of the 11 studies, not all patients were included in the analysis. In six studies it was unclear if there was interval change in the patient's condition between tests, and in four studies it was unclear how non-diagnostic coronary segments were treated in the analysis.

b. Uncertainty of directness relates to the possibility of spectrum bias within the sample due to only higher risk patients receiving the reference standard (invasive coronary angiography).

c. Significant and unexplained heterogeneity of results for specificity.

d. Large confidence intervals and no overlap between confidence intervals of studies for specificity.

Chapter Five: Discussion

Introduction

This systematic review of diagnostic test accuracy found 12 studies addressing the value of CTCA in patients with high HRs. A reasonable number of studies was expected, as HR is a factor commonly associated with poorer diagnostic test performance.[43,57]

The main results of the review were that overall sensitivity and specificity of current scan technologies were 99% (95% CI: 97%, 100%) and 79% (95%CI: 72%, 85%), respectively. This chapter further explains and interprets these results with a comparison to known estimates of CTCA accuracy based on previous systematic reviews of general populations, and provides the clinical ramifications of these findings. A discussion of the findings at a vessel and vessel segment level is also provided. Finally, the strengths and limitations of the systematic review are discussed.

Interpretation of Results

The results of this systematic review suggest that the diagnostic sensitivity of CTCA is relatively unaffected at high HRs, although there is likely an increase in the frequency of false positive results when scans are acquired at higher HRs. In practice this would increase the chance that a patient would require further assessment, such as catheter angiography. Therefore, clinicians should be aware of the post-test probability implications when considering the value of CTCA for individual patients with high HRs. This is particularly important for patients at high pre-test risk of clinically significant coronary artery stenosis.

In the context of the current review, a hypothetical cohort of 1000 patients with high HRs undergoing CTCA for clinically important coronary artery stenosis have a disease prevalence of 69.2%. In practical terms this means that 780 patients of the 1000 scanned would receive a positive test result, and 680 (95% CI: 671, 692) of these will have clinically significant coronary artery stenosis. However, 100 patients will have a false positive result (95% CI: 160, 40).

Conversely, of the 1000 hypothetical patients undergoing CTCA with elevated HR, 220 of these would receive a negative test result for clinically significant coronary stenosis. Of these patients, 213 (95% CI: 148, 268) would receive an accurate result. However, seven patients (95% CI: 21, 0) will receive a false negative result and are therefore at continued risk. Figure 21 summarises the clinical implications of this hypothetical scenario based on the systematic review findings. The graphical model follows guidance from a paper by Whiting and colleagues.[35]

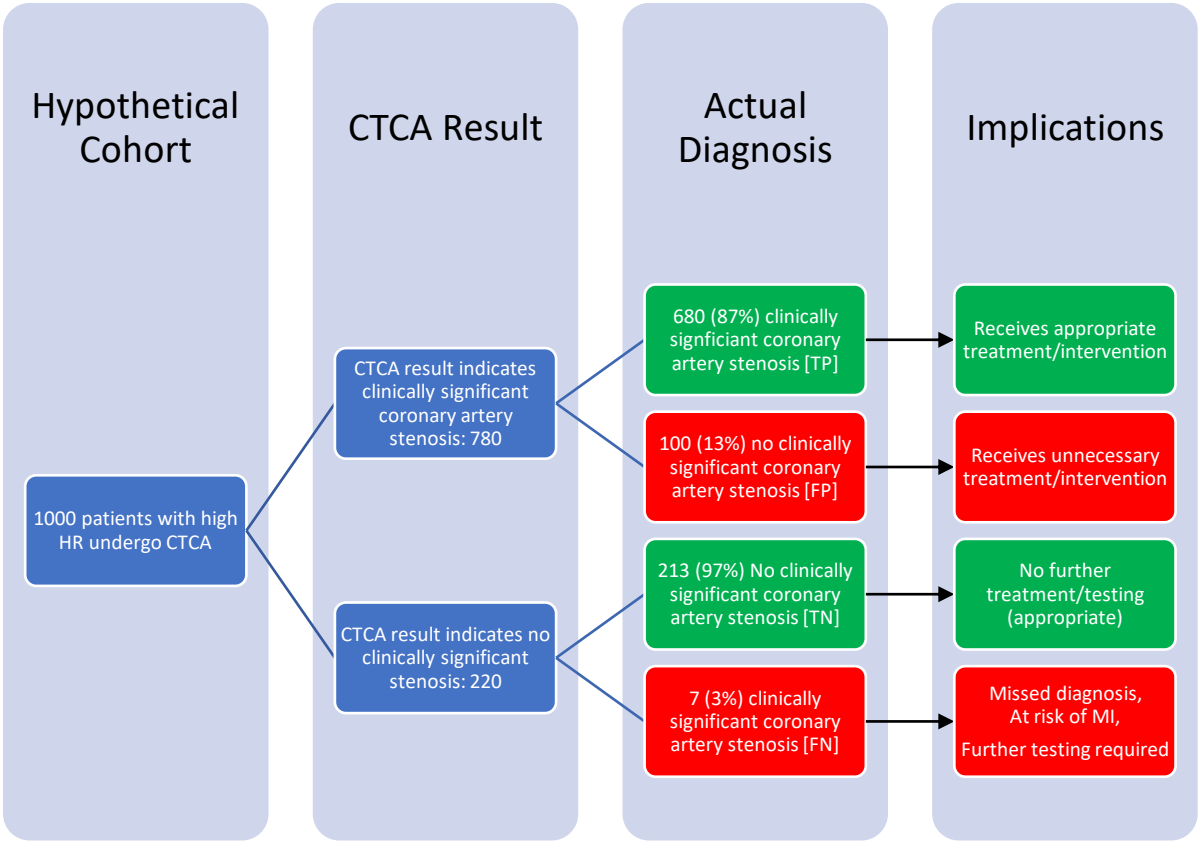

Figure 21: Hypothetical scenario based on review findings

It should be noted that the outcome implications are strongly associated with the prevalence of disease in the patient cohort being examined. This is due to the spectrum effect.[77] As disease prevalence increases, more patients will tend to be classified as false negatives, but there will be fewer false positives. Likewise, in lower prevalence cohorts, there will be more false positives but fewer false negatives. The effect of varying disease prevalence will directly affect the predictive value of the test but will not directly alter the sensitivity and specificity.[77]

If the level of prevalence commonly associated with a low-intermediate pre-test probability group is used, the results appear quite different.

It is therefore important to consider the pre-test risk of the patient undergoing CTCA as well as the position of the test in the diagnostic pathway when deciding whether CTCA is an appropriate option for patients with high HRs.

Comparison to Existing Literature

Numerous previous systematic reviews have investigated the diagnostic accuracy of CTCA. Systematic reviews have described the accuracy of CTCA both in terms of general populations as well as in difficult-to-image patient groups.[55,79]

A 2006 systematic review by Sun and Jiang[80] described the diagnostic accuracy of CTCA. They determined sensitivity and specificity were 91% (95% CI: 81%, 95%) and 86% (95% CI: 81%, 92%), respectively.

A 2008 systematic review by Janne d'Othee and colleagues[81] described the diagnostic accuracy of CTCA. The reviewers produced summary estimates of sensitivity and specificity at patient and segment levels and compared them. The per-patient summary estimate for sensitivity was 95% and specificity was 85%.

In a 2010 systematic review Ollendorf et al.[82] describe the diagnostic accuracy of CTCA. Their study produced summary estimates for patient-level sensitivity of 98% (95% CI: 96%, 99%) and specificity 85% (95%CI: 81%, 89%).

Sun and Ng[83] reported diagnostic accuracy of prospective compared with retrospective ECG gating in CTCA in their 2012 systematic review. Summary estimates were calculated as 97.7% (95% CI: 93.7%, 100%) sensitivity and 92.1% (95%CI: 87.2%, 97%) specificity.

A 2013 diagnostic accuracy meta-analysis of 320-detector row CT conducted by Li and colleagues[84] found a pooled sensitivity of 93% (95%CI: 91%, 95%) and specificity 86% (95% CI: 82%, 89%).

These review estimates of diagnostic accuracy show an increasing trend with year of publication (summarised in Table 11), which is possibly associated with advances in CT technology.

Table 11: Previous systematic reviews reporting diagnostic accuracy in CTCA

Systematic Review Authors [date]	**Reported Sensitivity % (95% CI)**	**Reported Specificity % (95% CI)**
Sun & Jiang (2006)	91 (81, 95)	86 (81, 0.92)
Janne d'Othee et al. (2008)	95*	85*
Ollendorf, Kuba & Pearson (2010)	98 (96, 99)	85 (81, 89)
Sun & Ng (2012)	98 (94, 100)	92 (87, 97)
Li et al. (2013)	93 (91, 95)	86 (82, 89)

* 95% confidence intervals not reported

Comparing these summary estimates for diagnostic accuracy to the findings of the current review sensitivity of 99% (95% CI: 97%, 100%) and specificity of 79% (95% CI: 72%, 85%), it would appear that elevated HRs are responsible for a significant decrease in specificity, especially when accounting for the inclusion of older generation scan technology in past reviews. Conversely, the sensitivity estimate in this review is equal to or greater than those reported in previous systematic reviews, which indicates that HR has little to no effect on diagnostic test sensitivity.

Vessel-Level Analysis

At the vessel level, diagnostic sensitivity remained high. However, there was a considerable increase in the specificity results (95% [95% CI: 93%, 96%]), when compared to the patient-level analysis (79% [95% CI: 72%, 85%]). This was similar to the finding at the segment level

(95% [95% CI: 92%, 98%]). This effect is probably best explained by the decreased disease prevalence in the vessel group and segment group, respectively. As described in a 2013 meta-analysis by Leeflang and colleagues,[77] an increase in disease prevalence is often associated with a decrease in specificity. However, the authors warned against suggesting calculations of sensitivity and specificity are directly affected by disease prevalence, as sensitivity is evaluated in patients with the diagnosis of interest, and specificity is calculated in those without. Therefore disease prevalence is theoretically independent of this estimate.[77]

Unlike sensitivity and specificity, the positive predictive value (PPV) and negative predictive value (NPV) calculations are directly affected by the prevalence of the diagnosis of interest in a study cohort. NPV *particularly* is an important consideration in diagnostic accuracy studies of CTCA, due to its position in the diagnostic decision pathway in clinical practice. That is, in clinical practice, false results may lead to one of two outcomes. It may either cause the patient to undergo additional unnecessary testing (in the case of a false positive), or (in the case of false negative) may miss a clinically significant coronary artery stenosis.

Missing significant CAD has the potential to leave the patient at continued risk of major adverse cardiovascular events (including unstable angina, myocardial infarction and cardiac arrest). Whilst the former situation, the false positive result, implies potential for wasted resources and some patient discomfort, the latter situation, the false negative result, is hypothetically the greater issue, as this situation can lead to patient mortality. Therefore, when considering the value of CTCA in clinical practice, the prevalence and the effect high HR may have on the NPV must be considered.

The pooled prevalence in the review was reported. The prevalence of clinically significant coronary artery stenosis is higher in the review than reported elsewhere. This is perhaps because the review reported data on a moderate- to high-risk group. It is therefore likely the reported prevalence is higher than that which would be expected in clinical practice. As mentioned, the sensitivity and specificity values should remain relatively unaffected by this finding; however, PPV and NPV will vary. The PPV of a diagnostic test is higher with higher disease prevalence. The NPV of a diagnostic test will decrease as prevalence increases.

Therefore, it is likely that the PPV described in the review is artificially inflated when compared to a clinical cohort.

Diagnostic Threshold

The diagnostic threshold used in the review was 50% stenosis. A uniform diagnostic threshold across studies decreases heterogeneity, as diagnostic threshold directly affects sensitivity and specificity values. As the threshold for positivity increases, sensitivity increases and specificity decreases. The contrary is true with a decrease in positivity. Where comparison is being made between similar studies reporting test accuracy at differing diagnostic thresholds, a SROC should be incorporated and AUC values can be calculated to provide an overall accuracy measurement for the test.

Non-Diagnostic Segments

The decision as to how non-diagnostic or non-evaluable results should be treated varied between studies. Commonly studies pragmatically treat these cases as positive to reflect clinical practices where an inconclusive result will usually be required to undergo further testing. This was the approach used for this review. However, comment should be made regarding the appropriateness of this approach. Non-evaluable results are particularly pertinent to the conduct of this review as an inconclusive result will likely result from motion blur associated with higher HR. It is therefore these segments that prove the most challenging to diagnose. Removing them from the analysis may positively bias accuracy of CTCA. However, arbitrarily including a non-diagnostic segment has limitations too, as it will generally lead to an increase in the FP rate. A systematic review by Menke and Kowalski,[85] describes diagnostic accuracy incorporating a 3x2 Bayesian meta-analysis to include unevaluable CTCAs with intention-to-diagnose. The reviewers recommended that, generally speaking, inconclusive results should not be treated as positive. Whilst further discussion of the appropriateness of Bayesian meta-analysis is beyond the scope of this research, more appropriate ways to treat non-diagnostic segments should be addressed in future research.

Subgroup Analyses

Heart Rate

The proportional effect of HR on diagnostic accuracy was not able to be fully examined in the review. This was for two key reasons. Firstly, there were insufficient data to provide conclusive evidence of poorer accuracy at higher HRs. Secondly, the HR data provided in the included studies were limited. Study authors grouped the accuracy of patients with high HRs, which did not sufficiently allow analysis of the effect of differing HRs on accuracy. With the previous caveat, visual inspection of the subgroup analysis in Figure 17 did not show any trend to decrease in sensitivity or specificity at higher HRs. This is true also of the Figure 18 segment level subgroup analysis. It is hypothesised that as the quiescent period of heart contractility approaches the temporal resolution of the scanner, a reduction in image quality in the form of motion artefact is likely to impact directly on the diagnostic accuracy of the scan. Further work is still required in order to quantify at what HR, if any, diagnostic accuracy is likely to be significantly and consistently affected. This is highly likely to be dependent on the temporal resolution of the specific scanner.

Scanner Make and Model

The specific value of individual scan technologies on the diagnostic accuracy was not well described in this review. There was insufficient existing evidence to describe the impact individual scanners had on diagnostic accuracy at high HRs. However, from the subgroup analysis performed (Figure 19), there did not appear to be a large difference between scanner makes. Due to the high sensitivity of the test, there was no difference in the performance of each make of scanner. With regard to specificity, Siemens systems appeared to perform slightly less well in comparison to GE and Canon scanners. However, these studies were underpowered and were older than the studies describing other makes of scanner. No data were available to describe the accuracy of Philips scanners at the patient level. More research would be required to determine whether or not differences in performance of scanners from different manufacturers are clinically and statistically significant.

At the segment level (Figure 20), there was little to no difference in the reported sensitivity and specificity between manufacturers. Therefore, no recommendations can be made from this review as to the most appropriate scanner to use for scanning patients with high HRs.

More prudently, it is recommended that current generation scanners are employed wherever high HRs are to be scanned.

Limitations of the Review

There were several limitations associated with the review. Firstly, it did not directly assess the difference in diagnostic accuracy between high and low HR groups. Whilst some of the included studies made comparison to a control, or low HR, group this information was not included in the review. This was because the purpose of the review was to identify the diagnostic accuracy of high HRs outright, rather than as a comparative assessment.

Secondly, the review looked at the diagnostic accuracy associated with HR; it did not investigate the accuracy associated with HR variability. Previous studies, including a systematic review, have investigated this.[79] In clinical terms, a patient with a high HR will not have the same diagnostic accuracy if there is inconsistent beat-to-beat variation. Therefore, extrapolation of the results of this review are limited to patients in sinus rhythm.

Thirdly, the review did not investigate other difficult-to-diagnose variables, such as the presence of intraluminal stents and/or very high levels of intraluminal calcium, or obese patients. As the aim of this review was to distinguish the effect high HRs had on diagnostic accuracy, the previously mentioned difficult-to-diagnose subjects were excluded from this review. This would therefore affect the external applicability of this review to certain clinical contexts. These factors should be considered when referring patients for CTCA.

Because there were insufficient studies available, the effect individual scan technologies have on the diagnostic test accuracy of CTCA in diagnosing CAD is unknown. It is very likely that, in the performance of the index test, some heterogeneity exists between included studies. This is due to the proprietary nature of the test and the differences that exist due to

commercially patented software and hardware between manufacturers. There is also a significant risk that conflict of interest impacts on the results of the review, although the size and direction of this effect is unknown.

Finally, the review was unable to determine the effect individual HRs had on diagnostic accuracy. As included studies did not reliably describe the HRs of each included participant, and individual participants' diagnoses were not defined by index test and reference standard, the review was limited in describing the relationship between high HR and diagnostic accuracy. Furthermore, as the high HR thresholds differed between included studies, it was not always clear how much the result was the effect of HR and how much was effect of other mechanisms, such as the index test.

Some limitations regarding the search strategy and screening of papers also deserve comment. English language methodological filters were applied in database searching, which may have excluded a potentially relevant study in journals printed in languages other than English. Formal database searching was limited to PubMed, CINAHL, Embase and Scopus. This means potentially relevant results may have been missed if they were only listed in more obscure databases.

Finally, data extraction was performed by the lead author only, which may have increased the potential for error in the transcription of data.

Heterogeneity

Statistical heterogeneity is assumed in systematic reviews of diagnostic test accuracy.[23] Studies may vary by the patients studied, severity of disease, test methods or other factors.[26]

Higgins' I^2 statistic is recommended as the best placed to deal with statistical heterogeneity in diagnostic accuracy systematic reviews.[26] This was the method used in the review.

Publication Bias

Meta-analyses of treatment outcomes commonly include statistical analysis to assess possible bias associated with unpublished data. However, these test (Beggs, Egger's test and Funnel plots) are not generally appropriate in diagnostic accuracy meta-analysis as they were not developed to be used in proportional meta-analyses with paired outcomes.[64] Therefore, for the purposes of this review, publication bias was considered a likely confounder but remained unstudied.

Deviation from Protocol

One included study [69] described patients with a HR greater than 60 bpm as having high HR. This differed from the definition of high HR in this review. The study was included as it was felt it still added relevant data. The effect of heterogeneity associated with minimum high HR threshold was assessed in a predefined subgroup analysis.

Chapter Six: Conclusions

Summary of Findings

This chapter summarises the findings of the review and makes recommendations regarding implications for practice and further research.

The review focused on the appropriate use of CTCA for patients with high HRs, specifically relating to the accuracy of the test for this patient group.

This systematic review identified 12 studies for inclusion, with 11 of these included in meta-analysis. Sensitivity and specificity estimates were determined for patient, vessel and vessel-segment levels. The overall sensitivity was 99% (95% CI: 97%, 100%); and specificity was 79% (95% CI: 72%, 85%). In the vessel-level analysis, the sensitivity and specificity were higher, 96% (95% CI: 93%, 97%); and 93% (95%CI: 90%, 96%), respectively. The sensitivity and specificity were also high in the segment-level analysis, 91% (95% CI: 88%, 93%); and 96% (95% CI: 95%, 98%), respectively. There was significantly more heterogeneity present in the vessel- and segment-level forest plots than at the patient level. This was apparent through both visual inspection of comparison of confidence intervals and assessment with I^2 statistical analysis of heterogeneity. The most likely cause for this increase in heterogeneity is variation in how non-evaluable segments were reported in the analysis. The expected outcome from reporting non-evaluable data as positive is an increased number of false positive results, and therefore decreased specificity. This was borne out in the sensitivity analysis, which described a higher level specificity when studies that treated non-evaluable segments as positive results 90% (95%CI: 86%, 93%) and 96% (95%CI: 94%, 98%) were directly compared against those studies that excluded non-evaluable segments from the analysis 93% (95% CI: 90%, 95%) and 98% (95%CI: 97%, 99%). Care should be taken when interpreting these results, however, as there is likely insufficient statistical power to conclude that the difference in how positive results are handled directly leads to significant heterogeneity between the studies.

Planned subgroup analyses assessing variation in accuracy due to varying HR thresholds and scanner make and model were both inconclusive. This was due to insufficient data to provide a significant result.

The strength of findings from this systematic review were considered very low according to GRADE methodology. The strength of findings were downgraded primarily because of a significant risk of bias associated with included studies, indirectness of the applicability of some results to the review question, and suspected significant heterogeneity.

The risk of bias in the review was considered in the QUADAS-2 assessment but related primarily to the following: inappropriate exclusions from the study; inappropriate exclusion from the analysis; unclear interval progression between the index test and reference standard; and a lack of clarity regarding how non-diagnostic coronary segments were evaluated.

The directness was downgraded because of the possibility of spectrum bias within the sample, because only higher risk patients received the reference standard (ICA).

The review was downgraded for statistical heterogeneity because of the significant and unexplained heterogeneity of results relating to specificity.

Implications for Practice

Performance of CTCA without appropriate HR control medications is still not recommended as routine practice. However, CTCA in patients with high HRs is justified when appropriate HR control is unavailable or contraindicated. The review demonstrates that sensitivity of CTCA to effectively rule out 50% coronary artery stenosis remains unaffected by HR. However, it will likely increase the risk of a false positive result. For this reason, clinicians should consider the ramifications when interpreting the results of CTCA in patients with high HRs.

Whilst the review was not able to provide suitable information on which CT scanner is the most accurate to use for patients with high HRs, it is recommended that patients with high HRs are scanned using the most recent generation scan technology available in order to optimise the chance of an accurate test result.

Implications for Research

Several concerns were raised in the study design, conduct and reporting of included studies. Study designs should follow the STARD reporting guidelines and include a flow diagram to allow the reader to more easily recognise the specific study design used. Where possible, studies should use a single-gate design and avoid inappropriate exclusions from the analysis.

Regarding the reporting of studies, two concerns deserve comment. Raw 2x2 table data should be reported in all studies that evaluate diagnostic accuracy, preferably in a tabulated form. This allows transparency of the research and will allow a more complete description of the effectiveness of the test.

Explicit comment should be made in primary studies of diagnostic accuracy about how non-diagnostic thresholds were handled and how many non-diagnostic segments were included in the analysis.

There is need for further research to focus on the value of individual technologies, or to investigate proprietary algorithms for improving diagnostic accuracy for patients with high HRs.

Diagnostic accuracy studies of CTCA should include specific HRs of individual participants to better describe the target group and allow more robust and detailed meta-analysis.

Future Research

The current review focused on patients with high HRs undergoing CTCA with current generation scanners. Further research is needed to better understand how various scanners

and algorithms affect the accuracy of the test. Further research and more detailed reporting is required in order to better understand the effect of individual HRs on test accuracy.

Conclusion

A detailed understanding of the accuracy of a clinical test in practice is important to inform the most appropriate application of the test in practice. Current generation CTCA scanners are highly sensitive at diagnosing significant coronary artery stenosis even at high HRs. High HRs do not affect the sensitivity of CTCA but have a moderate effect on specificity.

References

1. Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem. 2012; 58(12):1636-43.

2. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology. 2003; 229(1):3-8.

3. Leeflang MMG, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. Syst Rev. 2013; 2(1):82.

4. Rodger M, Ramsay T, Fergusson D. Diagnostic randomized controlled trials: the final frontier. Trials. 2012; 13:137-40.

5. Swets J. Measuring the accuracy of diagnostic systems. Science. 1988; 240(4857):1285-93.

6. Šimundić A-M. Measures of diagnostic accuracy: basic definitions. EJIFCC. 2009; 19(4):203-11.

7. Rutjes AWS, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PMM. Case–control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005; 51(8):1335-41.

8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. Radiology. 2015; 277(3):826-32.

9. Sibbald B, Roland M. Understanding controlled trials: why are randomised controlled trials important? BMJ. 1998; 316(7126):201.

10. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003; 3(1):25.

11. Clarke M, Chalmers I. Reflections on the history of systematic reviews. BMJ Evid Based Med. 2018; 23(4):121-2.

12. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. J Clin Nurs. 2003; 12(1):77-84.

13. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018; 18(1):5.

14. Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, et al. Diagnostic test accuracy: methods for systematic review and meta-analysis. Int J Evid Based Healthc. 2015; 13(3):154-62.

15. Whiting PF, Rutjes AS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine. 2011; 155(8):529-36.

16. Schmidt RL, Factor RE. Understanding sources of bias in diagnostic accuracy studies. Arch Pathol Lab Med. 2013; 137(4):558-65.

17. Venazzi A, Swardfager W, Lam B, Siqueira JdO, Herrmann N, Cogo-Moreira H. Validity of the QUADAS-2 in assessing risk of bias in Alzheimer's Disease diagnostic accuracy studies. Front psychiatry. 2018; 9:221-31.

18. Wade R, Corbett M, Eastwood A. Quality assessment of comparative diagnostic accuracy studies: our experience using a modified version of the QUADAS-2 tool. Res Synth Methods. 2013; 4(3):280-6.

19. Dehmoobad Sharifabadi A, Leeflang M, Treanor L, Kraaijpoel N, Salameh J-P, Alabousi M, et al. Comparative reviews of diagnostic test accuracy in imaging research: evaluation of current practices. Eur Radiol. 2019.

20. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. Plast Reconstr Surg. 2010; 126(2):619-25.

21. Roever L. Types of bias in studies of diagnostic test accuracy. Evid Based Med Pract. 2016; 22:e113.

22. Usher-Smith JA, Sharp, SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ. 2016; 353:[Internet]. 2016 Jun 22 [cited 2019 Sep 22]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27334281.

23. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med. 2008; 149(12):889-97.

24. O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. BMJ Evid Based Med. 2018; 23(2):54-5.

25. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. CJEM. 2008; 10(2):174-5.

26. Cronin P, Kelly AM, Altaee D, Foerster B, Petrou M, Dwamena BA. How to perform a systematic review and meta-analysis of diagnostic imaging studies. Acad Radiol. 2018; 25(5):573-93.

27. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Medical Decision Making. 1993; 13(4):313-21.

28. Reitsma JB RA, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ,. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C., editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 100. The Cochrane Collaboration; 2009. Available from: http://srdta.cochrane.org/.

29. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001; 20(19):2865-84.

30. Australian Bureau of Statistics. 3303.0 – Cause of death, Australia, 2017. [Internet]: Australian Bureau of Statistics; 2017 [updated 26/09/2018; cited 24/08/2019]. Available from:
https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3303.0Main+Features12017?OpenDocument

31. Kumar A, Cannon CP. Acute coronary syndromes: diagnosis and management, part I. Mayo Clin Proc. 2009; 84(10):917-38.

32. Braunwald E. Management of unstable angina based on considerations of aetiology. Heart. 1999; 82(suppl 1):I5-I7.

33. Rosenthal RL. The 50% Coronary Stenosis. Am J Cardiol. 2015; 115(8):1162-5.

34. Basit H MA, Huecker MR. Non ST Segment Elevation (NSTEMI) Myocardial Infarction. [Internet]. Treasure Island FL: StatsPearls Publishing; 2019 [cited 4/5/2019]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK513228/

35. Whiting P, Davenport C. Understanding test accuracy research: a test consequence graphic. Diagn Progn Res. 2018; 2(1):1-5.

36. Hecht HS. Is coronary computed tomographic angiography the “gold standard” for coronary artery disease? J of Cardiovasc Comput Tomogr. 2009; 3(5):334-9.

37. Austen W, Edwards J, Frye R, Gensini G, Gott V, Griffith L, et al. A reporting system on patients evaluated for coronary artery disease. Report of the Ad Hoc Committee for Grading of Coronary Artery Disease, Council on Cardiovascular Surgery, American Heart Association. Circulation. 1975; 51(4):5-40.

38. Leipsic J, Abbara S, Achenbach S, Cury R, Earls JP, Mancini GBJ, et al. SCCT guidelines for the interpretation and reporting of coronary CT angiography: a report of the Society of Cardiovascular Computed Tomography Guidelines Committee. J Cardiovasc Comput Tomogr. 2014; 8(5):342-58.

39. Pelc NJ. Recent and future directions in CT imaging. Ann Biomed Eng. 2014; 42(2):260-8.

40. Coronary CT angiography and 5-year risk of myocardial infarction. New Engl J Med. 2018; 379(10):924-33.

41. Desjardins B, Kazerooni EA. ECG-gated cardiac CT. AJR Am J Roentgenol. 2004; 182(4):993-1010.

42. Budoff MJ. Maximizing dose reductions with cardiac CT. The International J Cardiovasc Imaging. 2009; 25(Suppl 2):279-87.

43. Weustink AC, Neefjes LA, Kyrzopoulos S, Straten Mv, Eu RN, Meijboom WB, et al. Impact of heart rate frequency and variability on radiation exposure, image quality, and diagnostic performance in dual-source spiral CT coronary angiography. Radiology. 2009; 253(3):672-80.

44. Lewis MA, Pascoal A, Keevil SF, Lewis CA. Selecting a CT scanner for cardiac imaging: the heart of the matter. Br J Radiol. 2016; 89(1065):376–80.

45. Andreini D, Mushtaq S, Pontone G, Conte E, Guglielmo M, Annoni A, et al. Diagnostic performance of coronary CT angiography carried out with a novel whole-heart coverage high-definition CT scanner in patients with high heart rate. Int J Cardiol. 2018; 257:325-31.

46. Liang J, Sun Y, Ye Z, Sun Y, Xu L, Zhou Z, et al. Second-generation motion correction algorithm improves diagnostic accuracy of single-beat coronary CT angiography in patients with increased heart rate. Eur Radiol. 2019; 29(8):4215-27.

47. Liang J, Wang H, Xu L, Dong L, Fan Z, Wang R, et al. Impact of SSF on diagnostic performance of coronary computed tomography angiography within 1 heart beat in patients with high heart rate using a 256-row detector computed tomography. J Comput Assist Tomogr. 2018; 42(1):54-61.

48. Liang J, Wang H, Xu L, Yang L, Dong L, Fan Z, et al. Diagnostic performance of 256-row detector coronary CT angiography in patients with high heart rates within a single cardiac cycle: a preliminary study. Clin Radiology. 2017; 72(8):694.e7-.e14.

49. Jiangping S, Zhe Z, Wei W, Yunhu S, Jie H, Hongyue W, et al. Assessment of coronary artery stenosis by coronary angiography. Cir: Cardiovasc Interv. 2013; 6(3):262-8.

50. Rodriguez-Granillo GA, Carrascosa P, Bruining N, Waksman R, Garcia-Garcia HM. Defining the non-vulnerable and vulnerable patients with computed tomography coronary angiography: evaluation of atherosclerotic plaque burden and composition. Eur Heart J Cardiovasc Imaging. 2016; 17(5):481-91.

51. Oldroyd K, Phadke K, Phillips R, Carson P, Clarke M, Davis J. Cardiac catheterisation by the Judkins technique as an outpatient procedure. BMJ. 1989; 298(6677):875-6.

52. Heo R, Nakazato R, Kalra D, Min JK. Noninvasive imaging in coronary artery disease. Semin Nucl Med. 2014; 44(5):398-409.

53. Abbara S, Blanke P, Maroules CD, Cheezum M, Choi AD, Han BK, et al. SCCT guidelines for the performance and acquisition of coronary computed tomographic angiography: a report of the Society of Cardiovascular Computed Tomography Guidelines Committee:

Endorsed by the North American Society for Cardiovascular Imaging (NASCI). J Cardiovasc Comput Tomogr. 2016; 10(6):435-49.

54. Harden SP, Bull RK, Bury RW, Castellano EA, Clayton B, Hamilton MC, et al. The safe practice of CT coronary angiography in adult patients in UK imaging departments. Clin Radiol. 2016; 71(8):722-8.

55. Westwood ME, Raatz HD, Misso K, Burgers L, Redekop K, Lhachimi SK, et al. Systematic review of the accuracy of dual-source cardiac CT for detection of arterial stenosis in difficult to image patient groups. Radiology. 2013; 267(2):387-95.

56. Schuetz GM, Schlattmann P, Achenbach S, Budoff M, Garcia MJ, Roehle R, et al. Individual patient data meta-analysis for the clinical assessment of coronary computed tomography angiography: protocol of the Collaborative Meta-Analysis of Cardiac CT (CoMe-CCT). Syst Rev. 2013; 2(13):1-10.

57. Haase R, Schlattmann P, Gueret P, Andreini D, Pontone G, Alkadhi H, et al. Diagnosis of obstructive coronary artery disease using computed tomography angiography in patients with stable chest pain depending on clinical probability and in clinically important subgroups: meta-analysis of individual patient data. BMJ. 2019; 365(1):1-15.

58. Li M, Zhang GM, Zhao JS, Jiang ZW, Peng ZH, Jin ZT, et al. Diagnostic performance of dual-source CT coronary angiography with and without heart rate control: Systematic review and meta-analysis. Clinical Radiol. 2014; 69(2):163-71.

59. Mander GTW, Dobeli K, Steffensen C, Munn Z. Diagnostic accuracy of computed tomography coronary angiography utilising recent advances in technology in patients with high heart rates: a systematic review protocol. JBI Database Systematic Rev Implement Rep. 2018; 17(7):1312-28.

60. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol. 2006; 59(3):234-40.

61. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7(3):177-88.

62. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. Int J Evid Based Healthc. 2015; 13(3):147-53.

63. Review Manager (RevMan) Version 5.3. Copenhagen: The Nordic Cochrane Centre The Cochrane Collaboration; 2014.

64. van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. BMC Med Res Methodol. 2014; 14(1):70.

65. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008; 336(7653):1106-10.

66. Brożek J, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. Allergy. 2009; 64(8):1109-16.

67. Li M, Yu BB, Wu JH, Xu L, Sun G. Coronary CT angiography incorporating Doppler-guided prospective ECG gating in patients with high heart rate: comparison with results of traditional prospective ECG gating. PloS One. 2013; 8(5):e63096.

68. Neefjes LA, Rossi A, Genders TSS, Nieman K, Papadopoulou SL, Dharampal AS, et al. Diagnostic accuracy of 128-slice dual-source CT coronary angiography: a randomized comparison of different acquisition protocols. Eur Radiol. 2013; 23(3):614-22.

69. Nerlekar N, Ko BS, Nasis A, Cameron JD, Leung M, Brown AJ, et al. Impact of heart rate on diagnostic accuracy of second generation 320-detector computed tomography coronary angiography. Cardiovascular Diagn Therapy. 2017; 7(3):296-304.

70. Selçuk T, Otçu H, Yüceler Z, Bilgili Ç, Bulakçı M, Savaş Y, et al. Effectiveness of using dual-source CT and the upshot it creates on both heart rate and image quality. Balkan Med J. 2016; 33(3):283-93.

71. Sun K, Han RJ, Cui LF, Zhao RP, Ma LJ, Wang LJ, et al. Feasibility and diagnostic accuracy for assessment of coronary artery stenosis of prospectively electrocardiogram-gated high-pitch spiral acquisition mode dual-source CT coronary angiography in patients with relatively higher heart rates: in comparison with catheter coronary angiography. Chin Med Sci J 2013; 27(4):213-9.

72. Wang G, Wu Y, Zhang Z, Zheng X, Zhang Y, Liang M, et al. Effect of heart rate on the diagnostic accuracy of 256-slice computed tomography angiography in the detection of coronary artery stenosis: ROC curve analysis. Exp Ther Med. 2016; 11(5):1937-42.

73. Zhang L, Wang Y, Schoepf U, Meinel F, Bayer R, Qi L, et al. Image quality, radiation dose, and diagnostic accuracy of prospectively ECG-triggered high-pitch coronary CT angiography at 70 kVp in a clinical setting: comparison with invasive coronary angiography. Eur Radiol. 2016; 26(3):797-806.

74. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6(7):e1000097.

75. Garcia MJ, Lessick J, Hoffmann MHK, CATSCAN Study Investigators. Accuracy of 16-row multidetector computed tomography for the assessment of coronary artery stenosis. JAMA. 2006; 296(4):403-11.

76. Gang S, Min L, Li L, Guo-Ying L, Lin X, Qun J, et al. Evaluation of CT coronary artery angiography with 320-row detector CT in a high-risk population. Br J Radiol. 2012; 85(1013):562-70.

77. Leeflang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. CMAJ. 2013; 185(11):E537-E44.

78. Paech DC, Weston AR. A systematic review of the clinical effectiveness of 64-slice or higher computed tomography angiography as an alternative to invasive coronary angiography in the investigation of suspected coronary artery disease. BMC Cardiovasc Disord. 2011; 11(32):1-11.

79. Vorre MM, Abdulla J. Diagnostic accuracy and radiation dose of CT coronary angiography in atrial fibrillation: systematic review and meta-analysis. Radiology. 2013; 267(2):376-86.

80. Sun Z, Jiang W. Diagnostic value of multislice computed tomography angiography in coronary artery disease: a meta-analysis. Eur J Radiol. 2006; 60(2):279-86.

81. Janne d'Othée B, Siebert U, Cury R, Jadvar H, Dunn EJ, Hoffmann U. A systematic review on diagnostic accuracy of CT-based detection of significant coronary artery disease. Eur J Radiol. 2008; 65(3):449-61.

82. Ollendorf DA, Kuba M, Pearson SD. The diagnostic performance of multi-slice coronary computed tomographic angiography: a systematic review. J Gen Intern Med. 2011; 26(3):307-16.

83. Sun Z, Ng K-H. Prospective versus retrospective ECG-gated multislice CT coronary angiography: a systematic review of radiation dose and diagnostic accuracy. Eur J Radiol. 2012; 81(2):e94-e100.

84. Li S, Ni Q, Wu H, Peng L, Dong R, Chen L, et al. Diagnostic accuracy of 320-slice computed tomography angiography for detection of coronary artery stenosis: meta-analysis. Int J Cardiol. 2013; 168(3):2699-705.

85. Menke J, Kowalski J. Diagnostic accuracy and utility of coronary CT angiography with consideration of unevaluable results: a systematic review and multivariate Bayesian random-effects meta-analysis with intention to diagnose. Eur Radiol. 2016; 26(2):451-8.

Appendices

Appendix One: Search Strategy

PubMed Search String

("heart rate"[mh:noexp] OR "elevated heart rate"[tiab] OR "high heart rate"[tiab] OR "sinus tachycardia"[tiab]) AND ("cardiac computed tomography angiography"[tw] OR "Cardiac CT"[tw] OR "dual source CT"[tw] OR "DSCT"[tiab] OR "320 detector CT"[tw] OR "320 slice"[tw] OR "computed tomography coronary angiography"[tw] OR "computerized tomography coronary angiography"[tw] OR "coronary computerized tomography angiography"[tw] OR "coronary computed tomography angiography"[tw] OR "coronary CTA"[tw] OR "coronary CT angiography"[tw] OR "CT coronary angiography"[tw] OR "CTCA"[tw] OR "CCTA"[tw] OR "motion correction"[tw] OR "SSF"[tiab] OR "256-slice"[tw]) AND ("coronary angiography"[mh] OR ("coronary"[tw] AND "angiography"[tw]) OR "coronary angiography"[tw]) AND ("2008/06/13"[PDat] : "2018/06/10"[PDat] AND English[lang])

CINAHL Search String

((MH “heart rate”) OR (TI “elevated heart rate”) OR (AB “elevated heart rate”) OR (TI “high heart rate”) OR (AB “sinus tachycardia”)) AND ((TX “cardiac computed tomography angiography”) OR (TX “Cardiac CT”) OR (“dual source CT”) OR (“DSCT”) OR (TX “320 detector CT”) OR (“320 slice”) OR (“computed tomography coronary angiography”) OR (TX “computerized tomography coronary angiography”) OR (TX “coronary computerized tomography angiography”) OR (“coronary computed tomography angiography”) OR (TX “coronary CTA”) OR (TX “coronary CT angiography”) OR (TX “CT coronary angiography”) OR (“CTCA”) OR (“CCTA”) OR (TX “motion correction”) OR (TI “SSF”) OR (“256-slice”)) AND ((MH“coronary angiography”) OR (TX“coronary” AND TX“angiography”) OR (TX“coronary angiography”))

Embase Search String

(“heart rate”:ti,ab OR “elevated heart rate”:ti,ab OR “high heart rate”:ti,ab OR “sinus tachycardia”:ti,ab) AND (“cardiac computed tomography angiography”:ti,ab OR “cardiac ct”:ti,ab OR “dual source ct”:ti,ab OR “dsct”:ti,ab OR “320 detector ct”:ti,ab OR “320 slice”:ti,ab OR “computed tomography coronary angiography”:ti,ab OR “computerized tomography coronary angiography”:ti,ab OR “coronary computerized tomography angiography”:ti,ab OR “coronary computed tomography angiography”:ti,ab OR “coronary cta”:ti,ab OR “coronary ct angiography”:ti,ab OR “ct coronary angiography”:ti,ab OR “ctca”:ti,ab OR “ccta”:ti,ab OR 'motion correction':ti,ab OR 'ssf':ti,ab OR '256-slice':ti,ab) AND ('coronary':ti,ab AND 'angiography':ti,ab OR 'coronary angiography':ti,ab) AND (2008:py OR 2009:py OR 2010:py OR 2011:py OR 2012:py OR 2013:py OR 2014:py OR 2015:py OR 2016:py OR 2017:py OR 2018:py) AND [english]/lim AND [embase]/lim

Scopus Search String

((KEY ("heart rate") OR TITLE-ABS ("elevated heart rate") OR TITLE-ABS ("high heart rate") OR TITLE-ABS ("sinus tachycardia"))) AND ((TITLE-ABS ("cardiac computed tomography angiography") OR TITLE-ABS ("Cardiac CT") OR TITLE-ABS ("dual source CT") OR TITLE-ABS ("DSCT") OR TITLE-ABS ("320 detector CT") OR TITLE-ABS ("320 slice") OR TITLE-ABS ("computed tomography coronary angiography") OR TITLE-ABS ("computerized tomography coronary angiography") OR TITLE-ABS ("coronary computerized tomography angiography") OR TITLE-ABS ("coronary computed tomography angiography") OR TITLE-ABS ("coronary CTA") OR TITLE-ABS ("coronary CT angiography") OR TITLE-ABS ("CT coronary angiography") OR TITLE-ABS ("CTCA") OR TITLE-ABS ("CCTA") OR TITLE-ABS ("motion correction") OR TITLE-ABS ("SSF") OR TITLE-ABS ("256-slice"))) AND (TITLE-ABS ("coronary angiography") OR (TITLE-ABS ("coronary") AND TITLE-ABS ("angiography")) OR KEY ("coronary angiography")) AND (LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010 OR LIMIT-TO (PUBYEAR, 2009) OR LIMIT-TO (PUBYEAR, 2008)) AND (LIMIT-TO (LANGUAGE "English"))

Appendix Two: Details of Studies Excluded at Full Text

Achenbach S, Goroll T, Seltmann M, Pflederer T, Anders K, Ropers D, et al. Detection of coronary artery stenoses by low-dose, prospectively ECG-triggered, high-pitch spiral coronary CT angiography. JACC Cardiovasc Imaging. 2011; 4(4):328-37.

Alkadhi H, Scheffel H, Desbiolles L, Gaemperli O, Stolzmann P, Plass A, et al. Dual-source computed tomography coronary angiography: Influence of obesity, calcium load, and heart rate on diagnostic accuracy. Eur Heart J. 2008; 29(6):766-76.

Alkadhi H, Stolzmann P, Desbiolles L, Baumueller S, Goetti R, Plass A, et al. Low-dose, 128-slice, dual-source CT coronary angiography: accuracy and radiation dose of the high-pitch and the step-and-shoot mode. Heart. 2010; 96(12):933-8.

Andreini D, Lin FY, Rizvi A, Cho I, Heo R, Pontone G, et al. Diagnostic performance of a novel coronary CT angiography algorithm: Prospective multicenter validation of an intracycle CT motion correction algorithm for diagnostic accuracy. AJR Am J Roentgenol. 2018; 210(6):1208-15.

Baxa J, Hromádka M, Šedivý J, Štěpánková L, Moláček J, Schmidt B, et al. Regadenoson-Stress Dynamic Myocardial Perfusion Improves Diagnostic Performance of CT Angiography in Assessment of Intermediate Coronary Artery Stenosis in Asymptomatic Patients [Internet]. Biomed Res Int. 2015; Article ID 105629 [cited 2019 September 22], Available from: http://dx.doi.org/10.1155/2015/105629.

Brodoefel H, Burgstahler C, Tsiflikas I, Reimann A, Schroeder S, Claussen CD, et al. Dual-source CT: Effect of heart rate, heart rate variability, and calcification on image quality and diagnostic accuracy. Radiology. 2008; 247(2):346-55.

Chen HW, Fang XM, Hu XY, Bao J, Hu CH, Chen Y, et al. Efficacy of dual-source CT coronary angiography in evaluating coronary stenosis: Initial experience. Clin Imaging. 2010; 34(3):165-71.

Cho I, Elmore K, ó Hartaigh B, Schulman-Marcus J, Granser H, Valenti V, et al. Heart-rate dependent improvement in image quality and diagnostic accuracy of coronary computed tomographic angiography by novel intracycle motion correction algorithm. Clin Imaging. 2015; 39(3):421-6.

De Graaf FR, Schuijf JD, Van Velzen JE, Kroft LJ, De Roos A, Reiber JHC, et al. Diagnostic accuracy of 320-row multidetector computed tomography coronary angiography in the non-invasive evaluation of significant coronary artery disease. Eur Heart J. 2010; 31(15):1908-15.

Fujimura T, Miura T, Nao T, Yoshimura M, Nakashima Y, Okada M, et al. Dual-source computed tomography coronary angiography in patients with high heart rate. Heart Vessels. 2014; 29(4):443-8.

Herzog BA, Husmann L, Burkhard N, Gaemperli O, Valenta I, Tatsugami F, et al. Accuracy of low-dose computed tomography coronary angiography using prospective electrocardiogram-triggering: first clinical experience. Eur Heart J. 2008; 29(24):3037-42.

Hu XH, Zheng WL, Wang D, Xie SS, Wu R, Zhang SZ. Accuracy of high-pitch prospectively ECG-triggering CT coronary angiography for assessment of stenosis in 103 patients: Comparison with invasive coronary angiography. Clin Radiol. 2012; 67(11):1083-8.

Iyengar SS, Morgan-Hughes G, Ukoumunne O, Clayton B, Davies EJ, Nikolaou V, et al. Diagnostic accuracy of high-definition CT coronary angiography in high-risk patients. Clin Radiol. 2016; 71(2):151-8.

Kim JS, Choo KS, Jeong DW, Chun KJ, Park YH, Song SG, et al. Step-and-shoot prospectively ECG-gated vs. retrospectively ECG-gated with tube current modulation coronary CT angiography using 128-slice MDCT patients with chest pain: diagnostic performance and radiation dose. Acta radiol (Stockholm, Sweden : 1987). 2011; 52(8):860-5.

Koplay M, Erdogan H, Avci A, Sivri M, Demir K, Guler I, et al. Radiation dose and diagnostic accuracy of high-pitch dual-source coronary angiography in the evaluation of coronary artery stenoses. Diagn Interv Imaging. 2016; 97(4):461-9.

Lee H, Kim JA, Lee JS, Suh J, Paik SH, Park JS. Impact of a vendor-specific motion-correction algorithm on image quality, interpretability, and diagnostic performance of daily routine coronary CT angiography: influence of heart rate on the effect of motion-correction. Int J Cardiovasc Imaging. 2014; 30(8):1603-12.

Leschka S, Stolzmann P, Desbiolles L, Baumueller S, Goetti R, Schertler T, et al. Diagnostic accuracy of high-pitch dual-source CT for the assessment of coronary stenoses: First experience. Eur Radiol. 2009; 19(12):2896-903.

Li J, Li X, Wei M, Zhao H, Zheng M. Diagnostic accuracy of dual-source computed tomography in the detection of coronary chronic total occlusion:Comparison with invasive angiography. African Journal of Biotechnology. 2011; 10(19):3854-8.

Li ZN, Yin WH, Lu B, Yan HB, Mu CW, Gao Y, et al. Improvement of image quality and diagnostic performance by an innovative motion-correction algorithm for prospectively ECG triggered coronary CT angiography. PLoS ONE. 2015; 10(11).

Madhok R, Aggarwal A. Comparison of 128-slice dual source CT coronary angiography with invasive coronary angiography. J Clin Diagn Res. 2014; 8(6):RC08-RC11.

Mahdavi A, Mohammadzadeh A, Joodi G, Tabatabaei MR, Sheikholeslami F, Motevalli M. Diagnostic accuracy of dual-source computerized tomography coronary angiography in symptomatic patients presenting to a referral cardiovascular center during daily clinical practice. Iran J Radiol. 2016; 13(2).

Muenzel D, Noel PB, Dorn F, Dobritz M, Rummeny EJ, Huber A, et al. Step and shoot coronary CT angiography using 256-slice CT: effect of heart rate and heart rate variability on image quality. Eur Radiol. 2011; 21(11):2277-84.

Nasis A, Leung MC, Antonis PR, Cameron JD, Lehman SJ, Hope SA, et al. Diagnostic accuracy of noninvasive coronary angiography with 320-detector row computed tomography. Am J Cardiol. 2010; 106(10):1429-35.

Ochs MM, Siepen F, Fritz T, Andre F, Gitsioudis G, Korosoglou G, et al. Limits of the possible: diagnostic image quality in coronary angiography with third-generation dual-source CT. Clin Res Cardiol. 2017; 106(7):485-92.

Pan CJ, Qian N, Wang T, Tang XQ, Xue YJ. Adaptive prospective ECG-triggered sequence coronary angiography in dual-source CT without heart rate control: image quality and diagnostic performance. Exp Ther Med. 2013; 5(2):636-42.

Pontone G, Andreini D, Bertella E, Baggiano A, Mushtaq S, Loguercio M, et al. Impact of an intra-cycle motion correction algorithm on overall evaluability and diagnostic accuracy of computed tomography coronary angiography. Eur Radiol. 2016; 26(1):147-56.

Rixe J, Rolf A, Conradi G, Moellmann H, Nef H, Neumann T, et al. Detection of relevant coronary artery disease using dual-source computed tomography in a high probability patient series : comparison with invasive angiography. Circ J. 2009; 73(2):316-22.

Saad MAM, Azer HY. Dual-source CT coronary angiography: Diagnostic accuracy without the use of B blockers. Egyptian Journal of Radiology and Nuclear Medicine. 2011; 42(3-4):281-7.

Scheffel H, Alkadhi H, Leschka S, Plass A, Desbiolles L, Guber I, et al. Low-dose CT coronary angiography in the step-and-shoot mode: diagnostic performance. Heart. 2008; 94(9):1132-7.

Stolzmann P, Goetti R, Baumueller S, Plass A, Falk V, Scheffel H, et al. Prospective and retrospective ECG-gating for CT coronary angiography perform similarly accurate at low heart rates. Eur J Radiol. 2011; 79(1):85-91.

Sun K, Han RJ, Cui LF, Zhao RP, Ma LJ, Wang LJ, et al. Feasibility and diagnostic accuracy for assessment of coronary artery stenosis of prospectively electrocardiogram-gated high-pitch spiral acquisition mode dual-source CT coronary angiography in patients with relatively higher heart rates: in comparison with catheter coronary angiography. Chin Med Sci J. 2013; 27(4):213-9.

Sun ML, Lu B, Wu RZ, Johnson L, Han L, Liu G, et al. Diagnostic accuracy of dual-source CT coronary angiography with prospective ECG-triggering on different heart rate patients. Eur Radiol. 2011; 21(8):1635-42.

Tang PH, Du BJ, Fang XM, Hu XY, Qian PY, Gao QS. Submillisievert coronary CT angiography with adaptive prospective ECG-triggered sequence acquisition and iterative reconstruction

in patients with high heart rate on the dual-source CT. J Xray Sci Technol. 2016; 24(6):807-20.

Tomizawa N, Hayakawa Y, Nojo T, Nakamura S. Improving the diagnostic performance of the real world coronary computed tomography angiography including uninterpretable segments. Int J Cardiol. 2014; 176(3):975-9.

Tsiflikas I, Brodoefel H, Reimann AJ, Thomas C, Ketelsen D, Schroeder S, et al. Coronary CT angiography with dual source computed tomography in 170 patients. Eur J Radiol. 2010; 74(1):161-5.

Uehara M, Takaoka H, Kobayashi Y, Funabashi N. Diagnostic accuracy of 320-slice computed-tomography for detection of significant coronary artery stenosis in patients with various heart rates and heart rhythms compared with conventional coronary-angiography. Int J Cardiol. 2013; 167(3):809-15.

Wang R, Schoepf UJ, Wu R, Nance JW, Lv B, Yang H, et al. Diagnostic accuracy of coronary ct angiography: Comparison of filtered back projection and iterative reconstruction with different strengths. J Comput Assist Tomogr. 2014; 38(2):179-84.

Weustink AC, Mollet NR, Neefjes LA, Van Straten M, Neoh E, Kyrzopoulos S, et al. Preserved diagnostic performance of dual-source CT coronary angiography with reduced radiation exposure and cancer risk. Radiology. 2009; 252(1):53-60.

Weustink AC, Neefjes LA, Kyrzopoulos S, van Straten M, Neoh Eu R, Meijboom WB, et al. Impact of heart rate frequency and variability on radiation exposure, image quality, and diagnostic performance in dual-source spiral CT coronary angiography. Radiology. 2009; 253(3):672-80.

Xu Y, Tang L, Zhu X, Xu H, Tang J, Yang Z, et al. Comparison of dual-source CT coronary angiography and conventional coronary angiography for detecting coronary artery disease. Int J Cardiovasc Imaging. 2010; 26 Suppl 1:75-81.

Zhang LJ, Wu SY, Wang J, Lu Y, Zhang ZL, Jiang SS, et al. Diagnostic accuracy of dual-source CT coronary angiography: The effect of average heart rate, heart rate variability, and calcium score in a clinical perspective. Acta radiol (Stockholm, Sweden : 1987). 2010; 51(7):727-40.

Appendix Three: Data Extracted from Primary Studies

Part i - Data Extracted for Participants

Study	Nationality	Age [Mean±SD(Range)] (Years)	All Participants in Study Population (Including Low HR)					High Heart Rate Group					Point at which Heart Rate was Determined
			No. of Participants [N]	Heart Rate [mean] (bpm)	Standard Deviation [SD]	HR Range (bpm)		No. of Participants [n]	High Heart Rate [mean] (bpm)	Standard Deviation [SD]	HR Range (bpm)		
						Min.	Max.				Min.	Max.	
Andreini 2018	Italy		100	N/A	N/A	N/A	N/A	40	93	±23.6	81	N/A	At Acquisition
Gang 2012	China		60	73.7	±15.4	51	128	26	86.5	±15.1	73	128	At Acquisition
Li 2013	China		N/A	N/A	N/A	N/A	N/A	61	75	±7.7	65	80	At Acquisition
Liang 2017	China		N/A	N/A	N/A	N/A	N/A	84	82.8	±7.9	75	117	At Acquisition
Liang 2018	China	49.5±9.2(34-76)	N/A	N/A	N/A	N/A	N/A	64	82.5	±7.3	75	106	At Acquisition
Liang 2019	China	58.7±9.8(31-76)	N/A	N/A	N/A	N/A	N/A	81	83.8	±8.9	75	134	At Acquisition
Neefjes 2013	The Netherlands		267	65	±12	N/A	N/A	67	75	±12	65	N/A	Immediately prior to scan
Nerlekar 2017	Australia	63±10()	107	N/A	N/A	37	80	52	69*	±8	60	80	Unclear
Selçuk 2016	Turkey	57±5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	At Acquisition
Sun 2013	China		N/A	N/A	N/A	N/A	N/A	47	79	±9	66	100	Unclear
Wang 2016	China	65±11(37-87)	100	76.44	±13.36	39	107	60	N/A	N/A	75	107	Unclear
Zhang 2016	China		43	69.4	±13.6	45	106	16	N/A	N/A	70	106	Unclear

Part ii - Data Extracted for Index Test (CT Coronary Angiography)

Study	Make	Model				Scan Parameters			
			Padding		Segment	Detector Row	Rotation Time (ms)	Motion Correction	Dose-Length Product (DLP) (mGy)
			Start	End					
Andreini 2018	*GE Healthcare*	*Revolution CT*	*40%*	*80%*	*Single beat*	*256x0.625*	*280ms*	*SnapShot Freeze*	*209*
Gang 2012	*Canon Medical*	*Aquilion ONE*	*30%*	*90%*	*Multi-beat*	*320x0.5*	*350ms*		*779***
Li 2013	*Canon Medical*	*Aquilion ONE*	*30%*	*80%*	*Multi-beat*	*320x0.5*	*350ms*		*321***
Liang 2017	*GE Healthcare*	*Revolution CT*	*30%*	*60%*	*Single beat*	*256x0.625*	*280ms*	*SnapShot Freeze*	*142*
Liang 2018	*GE Healthcare*	*Revolution CT*	*30%*	*60%*	*Single beat*	*256x0.625*	*280ms*	*SnapShot Freeze*	*144*
Liang 2019	*GE Healthcare*	*Revolution CT*	*30%*	*60%*	*Single beat*	*256x0.625*	*280ms*	*SnapShot Freeze*	*70*
Neefjes 2013	*Siemens Healthineers*	*SOMATOM Definition Flash*	*55%*	*unclear*	*Single beat*	*2x64x0.6*	*280ms*		*Unclear*
Nerlekar 2017	*Canon Medical*	*Aquilion ONE ViSION Edition*	*30%*	*80%*	*Single beat*	*320x0.5*	*275ms*		*193*
Selçuk 2016	*Siemens Healthineers*	*SOMATOM Definition Flash*	*60%*	*-*	*Multi-beat*	*2x64x0.6*	*280ms*		*Unclear*
Sun 2013	*Siemens Healthineers*	*SOMATOM Definition Flash*	*20%*	*Unclear*	*Unclear*	*2x64x0.6*	*280ms*		*61*
Wang 2016	*Philips Healthcare*	*Brilliance iCT*	*Not reported*	*Not reported*	*Not reported*	*256x0.6*	*N/A*		*Not reported*
Zhang 2016	*Siemens Healthineers*	*SOMATOM Definition Flash*	*20%*	*Unclear*	*Single beat*	*2x64x0.6*	*280ms*		*Unclear*

** estimate converted to dose-length product from effective dose using the dose coefficient included in the study method.

Part iii - Data Extracted for Reference Standard (Coronary Angiography)

Study	Disease significance threshold
Andreini 2018	*≥50% coronary lumen diameter*
Gang 2012	*≥50% coronary lumen diameter*
Li 2013	*≥50% coronary lumen diameter*
Liang 2017	*≥50% coronary lumen diameter*
Liang 2018	*≥50% coronary lumen diameter*
Liang 2019	*≥50% coronary lumen diameter*
Neefjes 2013	*>50% coronary lumen diameter*
Nerlekar 2017	*>50% coronary lumen diameter*
Selçuk 2016	*>50% coronary lumen diameter*
Sun 2013	*>50% coronary lumen diameter*
Wang 2016	*≥50% coronary lumen diameter*
Zhang 2016	*≥50% coronary lumen diameter*

Part iv - Data Extracted for Diagnosis of Interest (Coronary Artery Stenosis)

Study	**Patient Level**								
	Prev (%)	**TP**	**FN**	**TN**	**FP**	**Sens (%)**	**Spec (%)**	**PPV (%)**	**NPV (%)**
Andreini 2018	*45*	*18*	*0*	*18*	*4*	*100*	*81.8*	*81.8*	*100*
Gang 2012	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Li 2013	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*97*	*89.3*	*91.4*	*96.2*
Liang 2017	*66.7*	*56*	*0*	*24*	*4*	100	85.7	93.3	100
Liang 2018	*67.2*	*43*	*3*	*0*	*18*	*100*	*85.7*	*70.5*	*100*
Liang 2019	*77.8*	*63*	*0*	*14*	*4*	100	77.8	94	100
Neefjes 2013	*76.1*	*51*	*0*	*10*	*6*	100	63	77.7	98.5
Nerlekar 2017	*67*	*34*	*0*	*15*	*2*	*100*	*88*	*95*	*100*
Selçuk 2016	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*88.8*	*81.8*	*88.8*	*81.8*
Sun 2013	*85.1*	*36*	*0*	*7*	*4*	*100*	*63.6*	*90*	*100*
*Wang 2016**	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Zhang 2016	*87.5*	*14*	*0*	*2*	*0*	100	100	100	100

* Two subgroups combined. HR 70-90 plus HR>90 groups

Study	**Artery Level**								
	Prev (%)	TP	FN	TN	FP	Sens (%)	Spec (%)	PPV (%)	NPV (%)
Andreini 2018	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Gang 2012	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Li 2013	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*91.1*	*96.5*	*85.4*	*98*
Liang 2017	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	95.2	93.5	87	97.7
Liang 2018	*30.9*	*76*	*3*	*167*	*10*	*96.2*	*94.3*	*62.9*	*99.2*
Liang 2019	*35.8*	*112*	*4*	*201*	*7*	96.6	96.6	94.1	98.1
Neefjes 2013	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	99	84	90	100
Nerlekar 2017	*28*	*57*	*1*	*137*	*9*	*98.2*	*93.4*	*76*	*97*
Selçuk 2016	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	81.4	95	88	91.9
Sun 2013	*14.5*	*81*	*9*	*277*	*14*	90	95.20	85.30	96.90
*Wang 2016**	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Zhang 2016	*43.8*	*27*	*1*	*33*	*3*	96.4	91.7	83	90.7

Study	Segment Level								
	Prev (%)	TP	FN	TN	FP	Sens (%)	Spec (%)	PPV (%)	NPV (%)
Andreini 2018	*9.7*	*46*	*16*	*576*	*2*	*95.20*	*98.90*	*86.90*	*99.60*
Gang 2012	*N/A*	*53*	*3*	*291*	*9*	*94.60*	*97*	*85.50*	*99.00*
Li 2013	*N/A*	*147*	*7*	*718*	*15*	*95.5*	*98*	*90.7*	*99*
Liang 2017	*14.4*	*140*	*13*	*869*	*40*	91.5	95.6	77.7	98.5
Liang 2018	*13.8*	*102*	*9*	*664*	*29*	*91.9*	*95.8*	*50.2*	*98.8*
Liang 2019	*17.5*	*177*	*15*	*884*	*20*	92.2	97.8	89.8	98.3
Neefjes 2013	*N/A*	*104*	*8*	*642*	*47*	93	93	69	99
Nerlekar 2017	*13*	*90*	*17*	*671*	*29*	*84*	*96*	*76*	*97*
Selçuk 2016	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	87.8	99.2	92.3	98.8
Sun 2013	*19.6*	*113*	*9*	*521*	*16*	92.60	97	87.60	98.30
*Wang 2016**	*27.8*	*216*	*16*	*583*	*20*	*93.1*	*96.7*	*93.1*	*97.3*
Zhang 2016	*22.3*	39	5	139	14	88.6	90.8	90	97.1