Validation of surrogate outcomes: application to biomarkers of atherosclerosis

Maryam Khan

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in
Partial fulfillment of the requirements for the MSc degree in Epidemiology

Epidemiology and Community Medicine
Faculty of Medicine
University of Ottawa

Abstract

Statement of the problem: Many methods for surrogate outcome validation require individual patient data which is often inaccessible by clinical trialists. Methods: A review was performed to identify statistical methods for surrogate outcome validation that may be implemented using summary data from published clinical trials. The methods were used to evaluate carotid intima-media thickness (CIMT) as a surrogate outcome for cardiovascular events in a systematic review of randomized trials of interventions for atherosclerosis. Results: the review of methods identified five procedures. At two or more years of follow-up, there was a marginally significant association of CIMT with myocardial infarction and a statistically significant association with cardiovascular mortality. At ≥ four years of follow-up, a statistically significant, negative relationship was observed between CIMT and stroke. Conclusions: CIMT may be a valid surrogate outcome for myocardial infarction and cardiovascular mortality. Additional data is needed to evaluate CIMT in specific drug classes.

Acknowledgements

I want to express my deepest gratitude to Dr George A Wells, who I am most fortunate to have had as a thesis supervisor. It would have been impossible for me to complete this thesis without Dr Wells' perfect patience, guidance and generous support that allowed me to overcome the various theoretical and methodological challenges I faced when working on this thesis. Dr Wells' availability for his students is remarkable and I am especially grateful for the time he spent on all aspects of my studies. It was a huge support for me to know that I could contact Dr Wells up to the last minute before a deadline, regardless of whether he was in his office and despite his busy schedule. I also want to sincerely thank Dr Wells for providing me with a study space at the University of Ottawa Heart Institute that was most conducive to working on my thesis, especially the private office I was given in the final phases of my thesis. I am indebted to Dr Wells for his wisdom and gentle style of providing advice that allowed me to clarify my future goals, and his enthusiastic support that has given me the confidence to pursue them.

I want to express my heartfelt thanks to Anne Gray for the ease and efficiency with which I was able to obtain thesis meeting times with Dr Wells. I am indebted to Anne for managing the resources in my study space so that I always had the supplies I needed when working on my thesis.

I am indebted to Fay Draper for supporting me through all aspects of my graduate studies, from the very beginning when I was applying to the program, until the very end when I submitted my thesis.

I want to sincerely thank Dr Rob Beanlands for readily providing his cardiovascular expertise so that I could properly address the clinical considerations in this thesis. I am especially grateful for his insightful review of this work which strengthened the findings and improved their clinical relevance.

I am especially thankful to Lee-Anne Ufholz for her dedication in training me on how to perform electronic literature searches.

I want to sincerely thank Kathryn Williams for the time she devoted towards teaching me to use the SAS software, especially the extra evenings during which she stayed late at the Heart Institute to train me to become more proficient at using the software.

I want to express my sincere gratitude to Lily Chen for demystifying statistical technicalities and teaching me how to efficiently think through statistical considerations.

I want to sincerely thank my sisters, Fatima, for her encouragement in my academic endeavors, and Zainab, for being a constant source of wisdom. Most importantly, I want to thank my parents. It is their innumerable sacrifices that have given me the very best educational opportunities. Without their endless support, I would not have reached this point in my academic career.

Table of contents

ABSTRACT ..... **II**

ACKNOWLEDGEMENTS ..... **III**

TABLE OF CONTENTS ..... **V**

LIST OF TABLES ..... **X**

LIST OF FIGURES ..... **XI**

1 INTRODUCTION 1

1.1 Statement of the problem *1*

1.2 Objectives *3*

1.3 Relevance of the research *3*

1.4 Outline *4*

2 BACKGROUND 8

2.1 Surrogate outcomes *8*

2.1.1 Properties of surrogate outcomes 8

2.1.2 Surrogate outcomes and the disease process 8

2.2 Statistical Methods for the Evaluation of Surrogate Endpoints *13*

2.2.1 Single-trial methods 13

2.2.1.1 The Prentice criteria 13

2.2.1.2 Freedman's proportion of treatment effect explained by a surrogate 14

2.2.1.3 Relative Effect and Adjusted Association 15

2.2.2 Meta-analytic techniques 17

2.2.2.1 Trial and individual level association between surrogate and clinical outcome for normally distributed outcomes 18

2.3 Surrogate Validation Schemas *21*

2.3.1 The evidence based medicine (EBM) users' guide 21

2.3.2: The National Institutes of Health (NIH)-derived guide 22

2.4 Surrogate Outcomes in Cardiovascular Disease Research *23*

2.4.1 The importance of surrogate outcomes in the treatment of atherosclerosis 23

2.4.2 The vascular biology of atherosclerosis 24

2.4.2.1 Composition and function of the healthy arterial wall 24

The intima 24

The media 25

The adventitia 25

2.4.2.2 Pathogenesis of atherosclerosis 27

Endothelial dysfunction 27

Lipoprotein entry and modification 27
Recruitment of leukocytes 27
Recruitment of smooth muscle cells 28
2.4.2.3 Atherosclerosis and thrombus formation 29
Direction of growth of the atherosclerotic lesion 31
2.5 Vascular Imaging in Atherosclerosis and Associated Surrogate Outcomes 31
2.5.1. Quantitative coronary angiography 32
2.5.1.1. Description of technique 32
2.5.1.2 Surrogate outcomes provided by QCA 32
2.5.1.3 Advantages and limitations of QCA 32
2.5.2. B-mode ultrasound 33
2.5.2.1. Description of technique 33
2.5.2.2 Surrogate outcomes provided by b-mode ultrasound 33
2.5.2.3 Advantages and limitations of B-mode ultrasound 34
2.5.3. Intravascular ultrasound (IVUS) 35
2.5.3.1. Description of technique 35
2.5.3.2 Surrogate outcomes provided by IVUS 35
2.5.3.3 Advantages and limitations of IVUS 36
3 REVIEW OF TRIAL-LEVEL STATISTICAL METHODS FOR SURROGATE OUTCOME EVALUATION 37
3.1 Overview 37
3.2 Objective 37
3.3 Methods 37
3.3.1 Criteria for selecting methods 37
3.3.3.1 Types of studies 37
3.3.2 Search methods for identification of studies 38
3.3.3 Data collection and analysis 39
3.3.3.1 Study selection 39
3.3.3.2 Data extraction 40
3.4 Results 40
3.4.1 Results of the search 40
3.4.2 Description of included analytical procedures 43
3.4.2.1 Significance testing for study wise agreement 43
Contribution to the procedure 43
Description of the procedure 43
Assumptions 44
Study-wise percent concordance as a measure to evaluate surrogacy 44
3.4.2.2 Rank correlation for the correlation coefficient 44
Contribution to the procedure 44

Description of the procedure 44
Assumptions 45
The spearman rank correlation coefficient as a measure of association 45
3.4.2.3 Graphical analysis 46
Contribution to the procedure 46
Description of the procedure 46
Assumptions 47
3.4.2.4 Regression analysis for R^2, and STE 47
Contribution to the procedure 47
Description of procedure 47
Assumptions 49
The coefficient of determination R^2 as a measure of association 50
Surrogate threshold effect (STE) as a measure of association 51
3.4.2.5 Mixed models for θ and Δ 52
Contribution to the procedure 52
Description of the procedure 52
Assumptions 55
β as a measure of association- unordered treatment arms 56
Δ as a measure of association 56
3.5 Discussion *58*
4 SURROGATE OUTCOME EVALUATION FOR CAROTID INTIMA-MEDIA THICKNESS 60
4.1 Overview *60*
4.2 Objective *60*
4.3 Methods *60*
4.3.1 Criteria for selecting studies for the review 60
4.3.1.1 Types of studies 60
4.3.1.2 Types of participants 61
4.3.1.3 Types of interventions 61
4.3.1.4 Types of outcome measures 61
Primary outcome measures 61
Secondary outcome measures 62
4.3.2 Search methods for identification of studies 62
4.3.2.1 Electronic searches 62
4.3.2.2 Searching other resources 63
4.3.3 Data collection and analysis 64
4.3.3.1 Selection of studies 64
4.3.3.2 Data collection 65
4.3.3.3 Assessment of risk of bias in included studies 66
4.3.3.4 Measures of treatment effect 67

4.3.3.5 Assessment of heterogeneity 67
4.3.3.6 Data synthesis 67
4.3.3.7 Studies with multiple treatment groups 68
4.3.3.8 Dealing with missing data 68
4.3.3.9 Sensitivity analysis 68
4.4 Results *71*
4.4.1 Results of the search 71
4.4.1.1 Description of included studies 74
Design 74
Sample sizes 74
Setting 74
Participants 76
Interventions 76
Outcomes 77
4.4.1.2 Description of excluded studies 78
4.4.1.3 Description on ongoing studies 78
4.4.2 Risk of bias in included studies 80
4.4.2.1 Sequence generation 81
4.2.2.2 Allocation concealment 81
4.4.2.3 Blinding 82
CIMT 82
Clinical outcomes 82
4.4.2.4 Incomplete outcome data 83
CIMT 83
Clinical outcomes 85
4.4.2.5 Selective reporting 86
4.4.2.6 Other potential sources of bias 86
4.4.3 Effects of interventions 87
4.4.3.1 Non-drug class specific 87
All-cause mortality 87
Myocardial infarction 96
Stroke 102
Cardiovascular mortality 107
4.4.3.2 Drug class specific comparisons 113
All-cause mortality 113
Myocardial infarction 118
4.4.3.2 Sensitivity analysis 120
4.5 Discussion *121*
5 DISCUSSION 129

5.1 Summary *129*

5.2 The role of the findings in relation to current literature *130*

5.3 Recommendations for future work *131*

REFERENCES **134**

APPENDIX A: SEARCH STRATEGY FOR TRIAL-LEVEL METHODS FOR SURROGATE OUTCOME EVALUATION **144**

APPENDIX B: SEARCH STRATEGY FOR RANDOMIZED TRIALS EVALUATING INTERVENTION EFFECTS ON CIMT AND CLINICAL OUTCOMES **145**

APPENDIX C: CHARACTERISTICS OF INCLUDED STUDIES **149**

List of tables

Table 2.1: Different types of surrogate and clinical outcomes

Table 4.1 Summary characteristics of included studies

Table 4.2 Description of excluded studies

Table 4.3: Description of ongoing studies

Table 4.4: Summary of results for non-drug class specific comparisons

Table 4.5: Estimates of treatment effect for CIMT and all-cause mortality

Table 4.6: Estimates of treatment effect for CIMT and myocardial infarction

Table 4.7: Estimates of treatment effect for CIMT and stroke

Table 4.8: Estimates of treatment effect for CIMT and cardiovascular mortality

Table 4.9: Results of applying transformations for the weighted regression for studies with 2 or more years of follow-up

Table C.1: Detailed tables for characteristics of included studies

List of figures

Figure 2.1: The surrogate outcome is on the causal pathway of disease progression to the clinical outcome and the effect of treatment on changes in the surrogate outcome precedes the effect of treatment on changes in the clinical outcome.

Figure 2.2: Possible relationships between the surrogate outcome, clinical outcome and the intervention under consideration.

Figure 2.3: Diagram of the arterial wall.

Figure 2.4: Schematic diagram for the progression of atherosclerotic plaque

Figure 2.5: Atherosclerosis progression to a fibrofatty atheroma

Figure 2.6: Necrotic core of a fibrofatty atheroma

Figure 2.7: B-mode ultrasound image of the carotid intima-media thickness.

Figure 3.1: Flow diagram for selection of studies reporting surrogate outcome validation procedures

Figure 4.1: Flow diagram for selection of studies of interventions for atherosclerosis.

Figure 4.2: Risk of bias summary figure

Figure 4.3: Risk of bias graph

Figure 4.4: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality at 2 or more years of follow-up

Figure 4.5: STE for 12 month mean difference in CIMT and all-cause mortality at 2 or more years of follow-up.

Figure 4.6: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality at 4 or more years of follow-up

Figure 4.7: STE for 12 month mean difference in CIMT and all-cause mortality at 4 or more years of follow-up

Figure 4.8: Graphical analysis for 12 month mean difference in CIMT and myocardial infarction at 2 or more years of follow-up

Figure 4.9: STE for 12 month mean difference in CIMT and myocardial infarction at 2 or more years of follow-up.

Figure 4.10: Graphical analysis for 12 month mean difference in CIMT and myocardial infarction at 4 or more years of follow-up

Figure 4.11: STE for 12 month mean difference in CIMT and myocardial infarction at 4 or more years of follow-up.

Figure 4.12: Graphical analysis for 12 month mean difference in CIMT and stroke at 2 or more years of follow-up

Figure 4.13: STE for 12 month mean difference in CIMT and stroke at 2 or more years of follow-up.

Figure 4.14: Graphical analysis for 12 month mean difference in CIMT and stroke at 4 or more years of follow-up

Figure 4.15: STE for 12 month mean difference in CIMT and stroke at 4 or more years of follow-up.

Figure 4.16: Graphical analysis for 12 month mean difference in CIMT and cardiovascular mortality at 2 or more years of follow-up.

Figure 4.17: STE for 12 month mean difference in CIMT and cardiovascular mortality at 2 or more years of follow-up.

Figure 4.18: Graphical analysis for 12 month mean difference in CIMT and cardiovascular mortality at 4 or more years of follow-up.

Figure 4.19: STE for 12 month mean difference in CIMT and cardiovascular mortality at 4 or more years of follow-up.

Figure 4.20: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality for hormone therapy versus placebo.

Figure 4.21: STE for 12 month mean difference in CIMT and all-cause mortality for hormone therapy vs. placebo.

Figure 4.22: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality for statin versus placebo.

Figure 4.23: STE for 12 month mean difference in CIMT and all-cause mortality for statin vs. placebo.

Figure 4.24: Graphical analysis for 12 month mean difference in CIMT and myocardial infarction for statin versus placebo.

Figure 4.25: STE for 12 month mean difference in CIMT and myocardial infarction for statin vs placebo.

1 INTRODUCTION

1.1 Statement of the problem

Surrogate outcomes are commonly measured in clinical trials to evaluate the effectiveness of treatment. The National Institutes of Health (NIH) working group for definitions of biomarkers and surrogate outcomes has recommended the following definitions [1]:

> ***"Biological Marker (Biomarker):*** *a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.*
>
> ***Clinical endpoint:*** *A characteristic or variable that reflects how a patient feels or functions, and how long a patient survives.*
>
> ***Surrogate endpoint:*** *A biomarker intended to substitute for a clinical endpoint. A clinical investigator uses epidemiologic, therapeutic, pathophysiologic, or other scientific evidence to select a surrogate endpoint that is expected to predict clinical benefit, harm, or lack of benefit or harm"*[1].

Evidence based medicine relies on the results of clinical trials to summarize the effect of treatment on clinical or health outcomes that are relevant to the patient population. In various situations, it may not be feasible to conduct a trial that assesses the occurrence of clinical outcomes. As compared to surrogate outcomes, the assessment of clinical outcomes generally requires longer follow-up periods and larger sample sizes, placing a greater burden on patients, health care providers and health care resources. The size and length of these trials may result in substantially increased study costs and decreased patient compliance. For serious diseases with no effective treatment, the follow-up period required to measure clinical outcomes hinder the accelerated drug approval process. As compared to surrogate outcomes, some clinical outcomes of interest may be difficult and/or costly to measure reliably.

As indicated by the definition for surrogate outcomes, their use should be preceded by compelling evidence which suggests that the effect of treatment on the surrogate outcome reliably predicts the effect of treatment on the surrogate outcome. Various statistics have been proposed to evaluate the validity of surrogate outcomes. These statistics are based on quantitative models that require either single-study data or data from multiple studies. While there are differences in the methods proposed, the underlying purpose of these methods is to evaluate the strength with which a treatment related change in a particular surrogate outcome can predict treatment related change in the clinical outcome. It is expected that in situations where a surrogate outcome is not a valid substitute for the clinical outcome, due to the relationship between the disease process, treatment, surrogate and clinical outcome, statistical methods will be able to reflect this quantitatively through an appropriate statistic that shows there is very little or no strength in the relationship between change in surrogate outcome and change in clinical outcome. It is also expected that in the cases where a surrogate outcome is considered to be a valid substitute for the clinical outcome, and this is quantified using one of the relevant statistics, investigators will recognize that the relationship between the change in surrogate outcome and change in clinical outcome may or may not be present if a different drug or drug class were under consideration. It will be necessary to undergo further statistical evaluation of the surrogate outcome when a new drug is under investigation. An important limitation of the major statistical approaches for surrogate outcome evaluation is that they require access to individual patient data which cannot be easily obtained.

Surrogate outcomes measured through imaging technologies have received considerable attention in trials for the treatment of atherosclerosis, a progressive pathology which presents itself as coronary heart disease, cerebrovascular disease and peripheral arterial disease. Progression of atherosclerosis leads to major cardiovascular disease (CVD) events that include fatal and non-fatal myocardial infarction (MI), stroke, and coronary death. Carotid intima-media thickness (CIMT), measured through b-mode ultrasound, is a commonly used surrogate outcome in trials for atherosclerosis in which imaging technologies have been used to evaluate the effectiveness of treatment. The confidence in measuring CIMT to evaluate the effect of treatments is reflected in the report from the 2009 Canadian Biomarkers and Surrogate Endpoints Symposium which concluded that

despite its limitations, b-mode ultrasound provides 'established surrogate outcomes' [2]. Several reviews have discussed the use of this vascular imaging outcome [3-5]; however, a systematic evaluation of this surrogate outcome could not be identified. For this reason, it is important to perform a systematic evaluation which quantifies the strength of CIMT as a surrogate outcome and determines its validity to substitute for major CVD events.

1.2 Objectives

The objectives of the thesis are as follows:

- To identify and describe statistical methods for surrogate outcome evaluation that may be implemented using summary statistics from published clinical trials.
- To perform a systematic review of trials for atherosclerosis evaluating the effect of treatment on carotid intima media thickness.
- To evaluate the strength of carotid intima media thickness as a surrogate outcome for clinical events using the statistical methods identified.

1.3 Relevance of the research

A central aspect in implementing clinical trials is the outcome chosen to measure response to treatment. Despite the widespread use of surrogate outcomes, there are many examples in which treatments that showed beneficial effects on surrogate outcomes subsequently showed no effect, or the opposite effect, on clinical outcomes. These examples indicate that clinical trialists are using surrogate outcomes that have not been validated. Reliance on these outcomes has important consequences in cardiovascular disease because clinical events such as myocardial infarction and stroke result in significant patient harm. Given the complexity of some of the major statistical approaches for surrogate outcome evaluation, it is plausible that many clinical trialists do not have the statistical support required to evaluate surrogate outcomes that may be of interest. Even in the presence of statistical support, there is often a lack of access to individual patient data from previous studies. In these situations, the justification for the

choice of surrogate outcomes is more likely to be based on resources such as expert opinion or animal models which are not sufficient to validate a surrogate outcome.

The statistical procedures identified in this thesis provide users of surrogate outcomes with guidance on methods for surrogate outcome evaluation based on summary statistics that are easily available from published reports of clinical trials. The evaluation of CIMT illustrates a practical application of these statistical procedures and highlights some of strengths and weaknesses associated with these methods. The evaluation of CIMT also quantifies its strength as a surrogate outcome and in this way provides evidence about whether it should continue to be used as a surrogate outcome in clinical trials for atherosclerosis. While it is possible that results for CIMT based on summary data do not reflect the results that would be obtained using individual patient data, these results may be the best estimates when individual patient data is unavailable or inaccessible. The clinical trials obtained in the review of CIMT provide an indication about the current best summary data available for the evaluation of CIMT in specific drug classes and additional data that is required in order to improve the accuracy of the results describing the validity of CIMT as a surrogate outcome.

1.4 Outline

This first chapter provides an introduction to the thesis. This chapter explains the concept of surrogate outcomes and discusses advantages and limitations associated with their use in clinical trials. The chapter also explains the concept of surrogate outcome validation and some general characteristics of well known statistical methods for the validation of surrogate outcomes. Information is presented about atherosclerosis and the role of carotid intima-media thickness as a surrogate outcome, provided by b-mode ultrasound technology, in clinical trials of interventions for atherosclerosis. The need to validate this surrogate outcome is discussed. The objectives of the thesis are listed in this introductory chapter. The relevance of the thesis is described in the context of the current use of and research on surrogate outcomes with a focus on CIMT.

Chapter 2 presents a background with detailed information that expands on the content presented in the introductory chapter. The information contained in Chapter 2 is relevant

for readers interested in additional details about the introductory material and for readers who may need clarification about any aspects of the thesis work. The background contains a section on the primary properties of surrogate outcomes. A section is included to describe the various factors in the relationship between the surrogate outcome, the disease process and the intervention due to which a surrogate outcome is likely to be a good versus poor substitute for the clinical outcome. The key equations, advantages and limitations of the well known surrogate outcome validation methods are described. These methods include both single-trial and multi-trial statistical techniques. The background chapter also discusses popular surrogate validation schemas that provide guidance to users of surrogate outcomes for integrating the overall evidence about a particular surrogate outcome in determining its validity. The chapter ends with a section on surrogate outcomes in atherosclerosis. The section describes the important role of surrogate outcomes in clinical trials of atherosclerosis. The section also describes the disease process for atherosclerosis to explain the current interest in surrogate outcomes provided by imaging of the vessel walls in contrast to plasma based surrogate outcomes.

Chapter 3 presents the methodology used to perform the literature review for identification of the relevant surrogate outcome validation methods/procedures. The chapter also provides detailed results with information about each of the statistical procedures that were identified from the review. The methods section includes information on the types of reports that were considered eligible for inclusion in the review, the specific characteristics of the eligible reports and the electronic and other sources that were used to search for relevant reports. The methods section also provides information on how eligible reports of the statistical procedures were selected from the set of potentially relevant reports and the data that was extracted from these reports. For the results, a section is included to describe the results of the search process, with information on the number of potentially eligible studies at each phase of the search process, reasons for exclusion of reports and the final number of eligible reports. For each of the identified statistical procedures, a description is provided about the number of reports with information about the procedure, the specific contribution of each report to the procedure, the description of the procedure, key equations, and any measures of association.

Chapter 4 presents the methods and results for i) a systematic review of the literature to identify randomized trials that have evaluated the effectiveness of interventions on CIMT and ii) the evaluation of CIMT as a surrogate outcome through application of the statistical procedures identified from the previous chapter to data obtained from the systematic review.

The methods section of Chapter 4 describes the criteria used to identify eligible studies, with information about the populations, interventions and clinical and surrogate outcomes that were considered relevant. Information is included about the electronic and other sources to search for studies, the selection process used to identify the included studies from the set of potentially relevant studies and the data that was collected. A section is also presented with details about the methods used to evaluate the risk of bias in the included studies. In describing the methods, a data synthesis section is included to describe the application of the statistical procedures identified in the previous chapter. The measures of treatment effect for CIMT and the clinical outcomes are provided. Assessment of heterogeneity between studies, planned sensitivity analyses, and issues of multiple treatment arms and missing data are described.

The results section of Chapter 4 presents the search results with information about the number of potentially eligible studies at each step of the selection process, reasons for exclusions and the final number of included studies. A description is provided for the characteristics of the eligible studies including information about the study designs, sample sizes, participants, interventions, CIMT measurements and the number of studies that reported each of the clinical outcomes of interest. In describing the results, a brief section is also presented for excluded and ongoing studies. The results of the risk of bias assessment are described in detail. The results obtained from applying each of the relevant statistical techniques are described and estimates are presented for the relevant measures of association. The results are described separately for studies of varying follow-up durations. The results for analyses that were not limited to specific drug classes are reported separately from analyses performed in specific drug classes. The results of the sensitivity analysis are reported.

Chapter 5 contains a discussion. The chapter begins with a summary of the findings from the third and fourth chapter, followed by an explanation of the observed results. A section is included to describe the role of the findings in the context of current literature on surrogate outcome validation and the use of CIMT as a surrogate outcome in clinical trials of atherosclerosis. Recommendations for future work are discussed.

2 BACKGROUND

2.1 Surrogate outcomes

2.1.1 Properties of surrogate outcomes

The surrogate outcome literature describes the properties of surrogate outcomes using various related terminology. These descriptions indicate that surrogate outcomes are characterized by three important properties: (1) a surrogate endpoint should be a modifiable biomarker so that drugs have the potential to change the value of the surrogate endpoint [1, 6-10]; (2) changes in the surrogate outcome should be prognostic of changes in the clinical outcome [6, 7, 9, 10]; (3) the effect of treatment on changes in the surrogate endpoint should be predictive of the effect of treatment on changes in the clinical endpoint [6, 7, 9, 10]. As such, the association between change in the surrogate outcome and change in the clinical outcome is biologically plausible- the surrogate outcome lies on the causal pathway of the disease process (Figure 2.1), and the effect of treatment induced changes in the surrogate outcome occur before treatment induced changes in the clinical outcome. Because changes in the surrogate outcome should be predictive of the clinical outcome, a mere correlation between changes in the surrogate outcome and changes in the clinical outcome is insufficient to establish the validity of a biomarker for use a surrogate endpoint [6-9, 11].

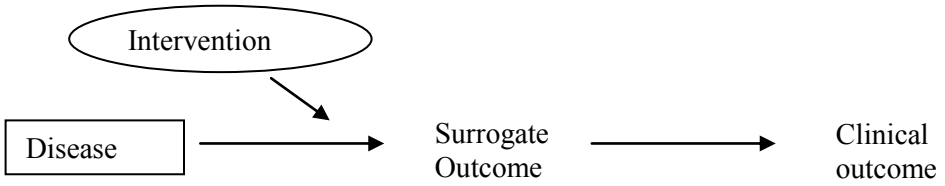

Figure 2.1: The surrogate outcome is on the causal pathway of disease progression to the clinical outcome and the effect of treatment on changes in the surrogate outcome precedes the effect of treatment on changes in the clinical outcome.

***Adapted from Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are we being mislead? Ann Intern Med 1996; 125:605-613.**

2.1.2 Surrogate outcomes and the disease process

For any given disease, the greater our understanding is about the disease process and its relationship with the surrogate outcome, clinical outcome and the treatment under

consideration, the better will be our ability to determine the suitability of a particular surrogate outcome to substitute for a clinical outcome when measuring response to treatment [9]. Due to these factors, Figure 2.2 depicts various scenarios in which a surrogate outcome may not be a good substitute for a clinical outcome.

In Figure 2.2A, there is a correlation between the surrogate and clinical outcome, but the surrogate is not on the causal pathway of disease progression to the clinical outcome. For this reason, the surrogate is not prognostic of the clinical outcome.

In Figure 2.2B, there are several causal pathways, one of which involves the surrogate outcome. The treatment under consideration only has an effect on the causal pathway in which the surrogate is present. Changes in the surrogate precede and therefore predict changes in the clinical outcome.

In Figure 2.2C, there are also several causal pathways, one of which involves the surrogate outcome. However, in this scenario, the treatment has an effect on the clinical outcome via the pathway that does not include the surrogate outcome and therefore, the treatment will have no effect on the surrogate.

In Figure 2.2D, the intervention affects the clinical outcome through pathways that are both dependent and independent of the surrogate outcome. This figure provides several instructive scenarios. First, when treatment affects a clinical outcome through several disease pathways, only some of which involve the surrogate outcome, it is possible for a drug to have an overall negative effect on the clinical outcome and a positive effect on the surrogate outcome. This may occur because the treatment affects the clinical outcome to a greater extent from the pathway that does not involve the surrogate outcome, and the effect of treatment on the clinical outcome through this pathway is in the opposite direction to the effect on the clinical outcome through the pathway mediated by the surrogate outcome. Second, the true amount of change in the clinical outcome may be overestimated by considering treatment induced changes in the surrogate outcome because the treatment may have a small effect on the disease pathway that involves the surrogate outcome, and a greater effect on the disease pathway that does not involve the surrogate outcome. Third, the true amount of change in a clinical outcome may also be

overestimated if the effect of the treatment on the surrogate outcome is only of short-term duration. In this scenario, the effect of treatment on the pathway in which the surrogate outcome is present does not appreciably alter the clinical outcome and the majority of change in the clinical outcome occurs through an alternative mechanism of action related to the disease pathway.

Figure 2.2D also illustrates the situation in which treatment can affect the clinical outcome through a mechanism of action that is completely independent of the disease processes. Even if treatment affects a clinical outcome through several disease pathways, only some of which involve the surrogate outcome, and even if the effect of the treatment on the surrogate is predictive of its effect on the clinical outcome through its mechanisms of action involving the various disease pathways, the treatment may still have a harmful effect on the clinical outcome in the presence of mechanisms of action that are independent of the disease process and which have a detrimental impact on the clinical outcome.

Figure 2.1 presents the scenario in which the surrogate outcome is most likely to be a good substitute for the clinical outcome. In this scenario, there is only one disease pathway, and the surrogate outcome lies on this pathway mediating the full effect of the treatment on the clinical outcome. The drug does not have any unintended mechanisms of action independent of the disease pathway that may affect the clinical outcome in a way that is opposite to the effect of the drug on the disease pathway. Although ideal, as compared to the other scenarios, this scenario is unlikely to occur in reality.

Collectively, these scenarios illustrate the difficulties involved in determining the suitability of a surrogate outcome due to the complex relationship between the pathways of disease progression, the surrogate and clinical outcome and the treatment under consideration. Importantly, these scenarios also illustrate that the suitability of a surrogate outcome largely depends on the intervention under consideration. Different interventions have varying mechanisms of action and affect the disease pathway(s) in different ways. For this reason, a surrogate that is considered a valid substitute for a clinical outcome may fail to be valid when used in a trial evaluating a new drug for which it is unknown

whether treatment related changes in the surrogate outcome can predict treatment related changes in the clinical outcome [6, 9, 10]. Trials that use surrogate outcomes commonly evaluate new drugs, and the primary use of surrogate outcomes is in the testing of new therapeutics. It may be very difficult to know in advance of beginning the trial, whether or not the chosen surrogate outcome will be a good substitute for the clinical outcome. However, if the treatment under consideration is from the same drug class as a previous treatment in which the surrogate outcome was a valid substitute for the clinical outcome, then there is greater confidence that changes in the surrogate outcome will predict changes in the clinical outcome [1].

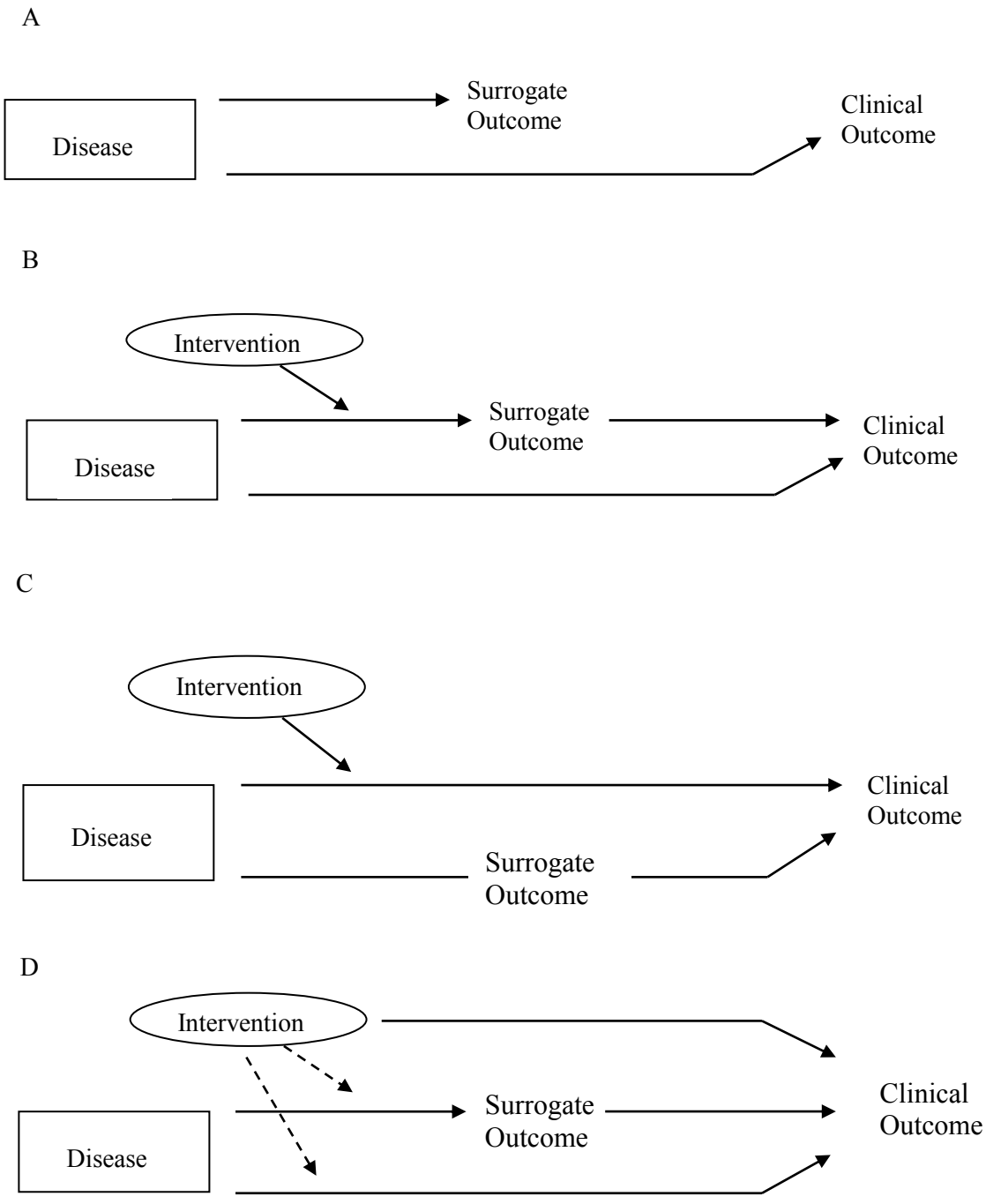

Figure 2.2: Possible relationships between the surrogate outcome, clinical outcome and the intervention under consideration.
***Adapted from Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are we being mislead? Ann Intern Med 1996; 125:605-613.**

2.2 Statistical Methods for the Evaluation of Surrogate Endpoints

A summary of the major single-trial and meta-analytic statistical approaches to the validation of surrogate outcomes is described below.

2.2.1 Single-trial methods

2.2.1.1 The Prentice criteria

Prentice in 1989 defined a surrogate endpoint as "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint" [12].
Burzykowski et al have provided an overview of the Prentice criteria and indicate that symbolically, his definition can be summarized as follows [7]:

$$f(S|Z) = f(S) \Leftrightarrow f(T|Z) = f(T) \quad [1]$$

where, S, T and Z are variables representing the surrogate outcome, clinical outcome and treatment, respectively; $f(X)$ denotes the probability distribution of random variable X; $f(X|Z)$ denotes the probability distribution of X conditional on the value of Z.

Burzykowski et al indicate that 4 operational criteria should be satisfied according to the Prentice definition of a surrogate outcome. They are as follows [7]:

$$f(S|Z) \neq f(S) \quad [2]$$

$$f(T|Z) \neq f(T) \quad [3]$$

$$f(T|S) \neq f(T) \quad [4]$$

$$f(T|S,Z) = f(T|S) \quad [5]$$

Textually, these four criteria indicate the following: 1) treatment has a statistically significant effect on the biomarker; 2) treatment has a statistically significant effect on the true clinical endpoint; 3) the biomarker has a statistically significant effect on the true clinical endpoint; and 4) the biomarker fully acquires the complete effect of treatment on the true clinical endpoint.

Burzykowski et al described various tests of significance that can be performed to verify whether a surrogate outcome, clinical outcome and treatment triplet satisfy these criteria. Although the Prentice criteria are informative, it is accepted that the fourth criteria, which requires the surrogate to fully capture the effect of treatment on the clinical outcome, is too strict and cannot be proven [7, 11, 13].

2.2.1.2 Freedman's proportion of treatment effect explained by a surrogate

In response to the limitations of the Prentice criteria, Freedman et al in 1992, proposed a new statistic, known as the proportion of treatment effect (PTE) explained by a surrogate [14]. This statistic is the ratio of the regression coefficients for the treatment indicator from two separate models, in each of which the clinical outcome is the response variable [12, 13]. Letting $PTE(T,S,Z)$ indicate the proportion of the effect of Z on T which can be explained by S, then

$$PTE(T,S,Z) = \frac{\beta - \beta_S}{\beta} = 1 - \frac{\beta_S}{\beta} \qquad [6]$$

where, β is the regression coefficient for the effect of the treatment on the clinical outcome from the model without adjustment for surrogate outcome; and β is the regression coefficient for the effect of treatment on the clinical outcome from the model with adjustment for the surrogate outcome. β and β can be obtained from the following models:

$$T_j = \alpha + \beta Z_j + \varepsilon_j \qquad [7]$$

$$T_j = \tilde{\alpha} + \beta_S Z_j + \gamma S_j + \tilde{\varepsilon}_j \qquad [8]$$

where T_j represents the change in the clinical outcome for patient j; μ and $\tilde{\mu}$ represent the intercept for the models without and with adjustment for the surrogate outcome, respectively; ε_j and $\tilde{\varepsilon}_j$ represent the random errors in the clinical outcome for patient j in the models without and with adjustment for the surrogate outcome, respectively; and γ represents the regression coefficient for the effect of the surrogate outcome on the clinical outcome from the model with adjustment for the treatment. A confidence interval for PTE can be calculated. Change in the surrogate is considered to have acceptable strength of association with change in the clinical outcome if the lower limit of this confidence interval is large [13].

Freedman et al indicated however, that if the effect of treatment on the clinical outcome is small and if only small sample sizes are available, then the confidence intervals of this statistic will be wide and there will be uncertainty in the proportion of the treatment effect that can be explained by the surrogate [7, 14]. Even in the presence of large sample sizes, the denominator in the formula for the PTE, which provides an estimate of the effect of the treatment on the clinical outcome, will generally be measured with little precision. For this reason, there will be uncertainty in the estimate of the proportion of treatment effect explained by the surrogate [7, 9]. Additionally, in the not uncommon situations in which β is less than four times its standard error, Freedman has indicated that the estimate of the proportion explained will be weak [11, 14]. It has also been shown that the value of PTE can lie outside the range of 0-1 and for this reason, it is not truly a proportion [11, 13].

2.2.1.3 Relative Effect and Adjusted Association

Buyse and Molenberghs in 1998, suggested two statistics for the validation of surrogate outcomes [7, 15]. The approach proposed by these investigators consists of validating the surrogate at the trial level and the individual level. The trial level association between change in the surrogate and change in the clinical outcome is known as the 'relative effect' (RE). This statistic represents the strength of the association between the treatment effect on the surrogate and clinical outcome at the trial level. The individual level association between the two outcomes is the adjusted association (AA). RE is the ratio of

the regression coefficients for the effect of treatment on the clinical and surrogate outcome. Letting RE (T,S,Z) indicate the effect of treatment on the clinical outcome relative to the effect of treatment on the surrogate outcome, then,

$$RE(T,S,Z) = \frac{\beta}{\alpha} \quad [9]$$

where, α is the regression coefficient for the effect of treatment on the surrogate outcome, and β is the regression coefficient for the effect of treatment on the clinical outcome. α and β can be obtained from the following models:

$$S_j = \mu_S + \alpha T_j + \varepsilon_{Sj} \quad [10]$$

$$T_j = \mu_T + \beta T_j + \varepsilon_{Tj} \quad [11]$$

where S_j represents change in the clinical outcome for patient j; μ_S and μ_T represent the intercepts in the models for the surrogate and clinical outcome, respectively; and ε_{Sj} and ε_{Tj} represent the random errors for patient j in the models for the surrogate and clinical outcome, respectively. The random errors follow a joint zero-mean normal distribution with the following variance-covariance matrix:

$$\sum = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{TT} \end{pmatrix} \quad [12]$$

Letting ρ represent the adjusted association, then,

$$\rho_{ST} = \frac{\sigma_{ST}}{\sqrt{\sigma_{SS}\sigma_{TT}}} \quad [13]$$

where, σ_{TT}, σ_{SS} and σ_{ST} are the elements of the variance-covariance matrix [12].

If RE=1, then the effect of treatment on both the surrogate and clinical outcome will be equal and the surrogate will be 'perfect at the trial level'. If AA=1, then there is a deterministic association between the surrogate and the clinical outcome, and the

surrogate will be 'perfect at the individual level'. The authors indicate that if a multiplicative relationship could be assumed in [9] and if the true value of RE is known , then the RE may be used to predict the effect of treatment on a clinical outcome based on data which only includes information about the effect of treatment on the surrogate outcome [7]. Although in practice RE needs to be estimated, the assumption of the multiplicative relationship between the effect of treatment on the clinical and surrogate outcome may be inappropriate and the can only be checked by using data from multiple trials [7].

2.2.2 Meta-analytic techniques

To address some of the limitations present in the individual level parameters, various other statistics have been developed within a meta-analytic framework. Buyse et al have proposed two well-known statistics by extending the concepts of RE and AA to the meta-analytic setting [7, 16]. In this setting, the coefficient of determination, R^2_{trial}, is the trial-level association between the surrogate and clinical outcome and is equivalent to RE. Similarly, the coefficient of determination, R^2_{indiv}, is the individual level association between the surrogate and clinical outcome and is equivalent to AA. In their book, Burzykowski, Molenberghs and Buyse have applied and modified the models used to determine R^2_{trial} and R^2_{indiv} for different types of outcomes [7] (Table 2.1).

Table 2.1: Different types of surrogate and clinical outcomes

Outcome measurement	Example of surrogate outcome	Example of clinical outcome
Binary	CD4+ counts over 500/mm^3	Tumor shrinkage
Categorical	Cholesterol levels < 200mg/dl, 200-299 mg/dl, 300 + mg/dl	Clinical response (complete response, partial response, stable disease, progressive disease)
Continuous	Log-PSA level	Depression scale
Censored continuous	Time to undetectable viral load	Time to cardiovascular death
Longitudinal/repeated measures	CD4+ counts over time	Depression scale measurements over time
Multivariate longitudinal	CD4+ and viral load over time	Dimensions of quality life over time

***Adapted from Burzykowski T, Molenberghs G, Buyse M. The evaluation of surrogate endpoints [Statistics in biology and health]. New York: Springer; 2005.**

2.2.2.1 Trial and individual level association between surrogate and clinical outcome for normally distributed outcomes

This section describes the models used to determine R^2_{trial} and R^2_{indiv} in the case of a normally distributed surrogate and clinical outcome. In the models described by Buyse et al, T_{ij} and S_{ij} are random variables that represent the clinical and surrogate outcome, respectively, for subject *j* in trial *i*. The first stage of a hierarchical two-staged model is described as follows:

$$S_{ij} = \mu_{S} + \alpha_{i} Z_{ij} + \varepsilon_{Sij} \quad [14]$$

$$T_{ij} = \mu_{T} + \beta_{i} Z_{ij} + \varepsilon_{Tij} \quad [15]$$

where, μ_{S} and μ_{T} are intercepts, the variable *Z* is the treatment indicator; α and β are the regression coefficients, for the effect of treatment Z on the surrogate and clinical outcome, respectively; and ε_{Sij} and ε_{Tij} are error terms which represent the imprecision in the estimate of the treatment effect in the models. The error terms follow a mean-zero normal distribution with covariance matrix

$$\sum \begin{pmatrix} \sigma_{S}^{2} & \sigma_{ST} \\ & \sigma_{T}^{2} \end{pmatrix} \quad [16]$$

In the second stage of the model,

$$\begin{pmatrix} \mu_{S} \\ \mu_{T} \\ \alpha_{i} \\ \beta_{i} \end{pmatrix} = \begin{pmatrix} \mu_{S} \\ \mu_{T} \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} m_{Si} \\ m_{Ti} \\ a_{i} \\ b_{i} \end{pmatrix} \quad [17]$$

where, μ_{S} and μ_{T} are fixed intercepts; m_{Si} and m_{Ti} are the random intercepts; α and β are the regression coefficients representing the fixed effects of treatment Z on the surrogate and clinical outcome, respectively; and a_i and b_i are the random effects of

treatment Z. m_{Si}, m_{Ti}, a_i, and b_i follow a zero-mean normal distribution with covariance matrix

$$D = \begin{pmatrix} d_{SS} & d_{ST} & d_{Sa} & d_{Sb} \\ & d_{TT} & d_{Ta} & d_{Tb} \\ & & d_{aa} & d_{ab} \\ & & & d_{bb} \end{pmatrix} \quad [18]$$

The random effects representation consists of combining both stages as follows:

$$S_{ij} = \mu_S + m_{Si} + \alpha_S Z_{ij} + \varepsilon_{Sij} \quad [19]$$

$$T_{ij} = \mu_T + m_{Ti} + \beta_T Z_{ij} + \varepsilon_{Tij} \quad [20]$$

To define R^2_{trial}, Buyse et al used the results obtained from the aforementioned models in the context of prediction. In this context, a new trial, i=0 contains data on the surrogate outcome, but not on the clinical outcome. To predict the effect of the treatment on the clinical outcome in i=0, $E(\beta_0 | m_{S0}, a_0)$, based on the observed effect of treatment on the surrogate outcome, S_{0j},

$$E(\beta_0 | m_{S0}, a_0) = \beta_T + \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}^T \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} \mu_{S0} - \mu_S \\ \alpha_0 - \alpha_S \end{pmatrix} \quad [21]$$

$$\mathrm{Var}(\beta_0 | m_{S0}, a_0) = d_{bb} - \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}^T \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix} \quad [22]$$

where the superscript T denotes transpose of the vector.

Then,

$$R^2_{\text{trial(f)}} = R^2_{b_i | mSi, ai} = \frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}^T \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}} \quad [23]$$

where the subscript (f) in the above formula indicates that $R^2_{\text{trial(f)}}$ was obtained from model in which the prediction of b_0 depends on m_{S0}. In the original text, the authors also provide a formula for $R^2_{\text{trial(r)}}$ based on a 'reduced' model in which the prediction of b_0 does not depend on m_{S0}.

Buyse et al defined the strength of the association between S and T, after adjustment for Z as R^2_{indiv}. To do this, they constructed the conditional distribution of T, given S and Z. For models [14]-[15], this conditional distribution is described as follows:

$$T_{ij}\left|Z_{ij},S_{ij}\right. \sim N\{\mu_{\ } - \sigma_{\ }\sigma_{\ }\mu_{\ } + \beta_{\ } - \sigma_{\ }\sigma_{\ }\alpha_{\ } Z_{ij} + \sigma_{\ }\sigma_{\ }S_{ij};\sigma_{\ r} - \sigma_{\ }\sigma_{\ }\sigma_{\ }\} \quad [24]$$

For model [19]-[20]:

$$T_{ij}\left|Z_{ij},S_{ij}\right. \sim N\{\mu_{\ } + \eta_{Ti} - \sigma_{\ }\sigma_{\ }(\mu_{\ } + \eta_{Si}) + \beta_{\ } - \nu_i - \sigma_{\ }\sigma_{\ }(\alpha_{\ } - \iota_i)]Z_{ij} \quad [25]$$

$$+ \sigma_{\ }\sigma_{\ }S_{ij};\sigma_{\ r} - \sigma_{\ }\sigma_{\ }\}$$

Then,

$$R^2_{\text{indiv}} = \Re^2_{\varepsilon_{\ i}|\varepsilon_{\ i}} = \frac{\sigma^2_{\ r}}{\sigma_{\ }\sigma_{\ r}} \quad [26]$$

A surrogate outcome is "trial-level valid" if $R^2_{\text{trial(f)}}$ (or $R^2_{\text{trial(r)}}$) are reasonably close to 1 and a surrogate outcome is "individual-level valid" if R^2_{indiv} is reasonably close to 1. A surrogate outcome is "valid" if it has both trial-level validity and individual-level validity.

The models in this section for normally distributed outcomes have also be adapted to describe another statistic, the surrogate threshold effect (STE). The STE quantifies the strength of the relationship between the surrogate outcome and its target. A full description of this parameter is beyond the scope of this chapter. The meta-analytic

approach may be an improvement over the single-trial methods for the evaluation of surrogate endpoints; however, it is computationally complex.

2.3 Surrogate Validation Schemas

Various Surrogate Validation schemas have been developed whose purpose is to provide a systematic guide that can be used to evaluate the adequacy or validity of a particular surrogate outcome.

2.3.1 The evidence based medicine (EBM) users' guide

A well-known surrogate validation schema was developed by Bucher et al [6]. This schema can be used when there is interest in evaluating the validity of a surrogate outcome for a specific class of drugs. According to this schema, a valid surrogate is one that fulfills the following three criteria: 1) there is a strong and independent, consistent association between the surrogate endpoint and the clinical endpoint; 2) there is evidence from randomized trials in other drug classes that improvement in the surrogate endpoint has consistently led to improvement in the target outcome; 3) there is evidence from randomized trials in the same drug class that improvement in the surrogate endpoint has consistently led to improvement in the target outcome.

The first of the three criteria requires evidence from observational studies in which there is a strong association between the surrogate and clinical outcome. The strength of the association is reflected through measures such as the relative risk (RR) and the odds ratio (OR). The strong relationship between the surrogate and clinical outcome is independent and consistent if the same relationship is observed after adjusting for potential confounders and is observed across multiple observational studies. However, observational studies do not evaluate the effect of treatment. For this reason, Bucher et al have developed criteria 2 and 3, which require evidence that the effect of treatment on changes in the surrogate outcome is associated with the effect of treatment on changes in the clinical outcome. Because randomization balances unknown confounders, the EBM users' guide indicates that this evidence should be derived from randomized controlled trials.

When a surrogate outcome satisfies criteria 2, the relationship between change in the surrogate outcome and change in the clinical outcome is considered to be consistent across drug classes. While it is not certain that the same relationship between the outcomes will be observed for a new class of drugs, fulfillment of this criteria provides strength to the inference that if treatment has a beneficial effect on the surrogate outcome, it will also result in a beneficial effect on the clinical outcome.

Criteria 3 is used to determine whether changes in the surrogate outcome are associated with changes in the clinical outcome for the particular class of drugs under consideration.

2.3.2: The National Institutes of Health (NIH)-derived guide

Lassere et al have recently developed another surrogate validation guide, based on definitions established by the NIH working group on surrogate outcomes [10, 11]. This evaluation tool consists of four domains: 1) 'target', 2) 'study design', 3) 'statistical strength' and 4) 'penalties'. Each of these domains contains a set of hierarchically ranked criteria. A surrogate outcome receives an additive score in each of the first 3 domains, according to the highest ranked criteria that it satisfies.

The purpose of the target domain is to rank the clinical outcome according to the degree to which it reflects how a patient "feels, functions or survives" [1]. Lassere et al considered outcomes as variables that lie on a 'variables in medicine' continuum. One end of the continuum contains disease centered variables. These variables are essentially biomarkes that are related to the disease pathway and do not directly describe how a patient feels, functions or survives. The other end of the continuum contains patient centered variables that directly describe how a patient feels, functions or survives. The target domain captures the location of the clinical outcome on this continuum. Clinical outcomes such as death and irreversible major organ morbidity receive a high score.

Similar to the EBM-users' guide, in the study design domain of the NIH-derived guide, treatment related change in the biomarker should be evaluated through randomized controlled trials and these studies receive a higher grade than observational studies.

Changes in surrogate outcomes that are consistently related to changes in the clinical outcome across different drug classes, receive the highest grade.

The third domain ranks the statistical strength of the association between change in surrogate outcome and change in clinical outcome from 'no association' to 'excellent association'.

In the fourth domain, penalty points can be obtained and are used to reduce the overall score. Penalty points can result if there is a lack of evidence, evidence that does not support surrogate validity or evidence of harm.

According to the NIH-derived guide, surrogate outcomes that receive a final score of 13-15 or 10-12 are considered level 1 and 2 surrogates, respectively. Lassere et al propose that only level 1 and level 2 surrogates are valid substitutes for clinical outcomes when measuring response to therapy. This evaluation schema incorporates a variety of perspectives into one surrogate validation guide, namely: biological, epidemiological, statistical and clinical. In this way, the NIH-derived guide may be viewed as an improvement to the EBM users' guide, which incorporated primarily a clinical perspective. The usefulness of this schema for various stakeholders has the advantage of facilitating communication between these viewpoints.

2.4 Surrogate Outcomes in Cardiovascular Disease Research

2.4.1 The importance of surrogate outcomes in the treatment of atherosclerosis

Clinical trials of drug therapy for CVD that evaluate the effect of treatment on cardiovascular events currently require a sample size of thousands of patients and a follow-up period of many years in order to have the statistical power necessary to detect a difference in outcomes between treatment arms [3]. Surrogate endpoint trials for CVD require a sample size of a much smaller number of patients and a follow-up period of several months. For instance, clinical trials evaluating the effect of treatment on lipid levels typically include a sample size of 100 patients followed-up for 3 to 12 months [17].

For this reason, comprehensive trial programmes include surrogate endpoint trials of new CVD drugs alongside ongoing clinical endpoint trials. In the last decades, many surrogate endpoint trials have evaluated the effectiveness of drugs on soluble surrogate endpoints, which are plasma-based and include LDL-c, HDL-c and triglycerides. However, based on the current understanding of the vascular biology of atherosclerosis, it is accepted that these surrogate endpoints form only a part of the complex pathways in atherosclerosis progression that lead to cardiovascular events [4].

2.4.2 The vascular biology of atherosclerosis

There is considerable complexity in the pathophysiology of atherosclerosis. Both genetic and environmental factors ultimately result in the development and progression of atherosclerosis, which biologically results from slowly progressive processes that begin with endothelial dysfunction and systemtic mediators, and lead to lipid accumulation, and migration of inflammatory cells into the arterial wall [4, 18]. This section describes the general framework for the development and progression of atherosclerosis.

2.4.2.1 Composition and function of the healthy arterial wall

The arterial wall is composed of three layers (Figure 2.3): (1) the intima, which is the layer nearest to the arterial lumen and also, therefore, closest to the blood in circulation; (2) the media, which is the middle layer; and (3) the adventitia, which is the outer layer of the arterial wall.

The intima

While often portrayed as a single layer of endothelial cells, the intima consists of endothelial cells residing on a basement membrane with various constituents of the extracellular matrix. In healthy arteries, the endothelial cells are tightly joined together and serve as a barrier between the circulating blood within the lumen and the sub-endothelial space. The endothelial cells also serve a metabolic and signalling function. In particular, the endothelium provides various molecules, either on its surface or through secretion into the circulation, which in normal arteries have a net antithrombotic effect. The endothelium is involved in the immune response when tissue injury occurs, such as

in atherosclerosis, by attracting leukocytes of the vessel wall; however, in normal arteries, the endothelial cells resist leukocyte adhesion and in this way, serve an anti-inflammatory role. The endothelial cells also secrete various substances that modulate the contraction of the smooth muscle cells that are present in the medial layer of the vessel wall. In healthy arteries, these substances have a net dilatory effect enabling relaxation of the smooth muscle cells.

The media

The medial layer of the artery is the thickest and consists primarily of smooth muscle cells and extracellular matrix. At its boundaries, the media consists of internal and external laminae which are composed of elastin. The internal and external laminae separate the medial layer from the intima and adventitia, respectively. The media serves a contractile and synthetic purpose. In response to various substances, some of which are released from the endothelium, the smooth muscle cells are stimulated to contract or relax. The smooth muscle cells also synthesize collagen, elastin and proteoglycans to form the extracellular matrix, which maintain the structural integrity of the vessel. In the case of tissue injury, the smooth muscle cells of the medial layer are capable of synthesizing various inflammatory mediators that stimulate the immune response of the endothelial cells, promote lymphocyte proliferation and propagate the inflammatory response. In diseased states such as that of atherosclerosis, these inflammatory functions of the smooth muscle cells increase.

The adventitia

The adventitia consists of blood vessels, nerves and lymphatics which provide nourishment to the artery.

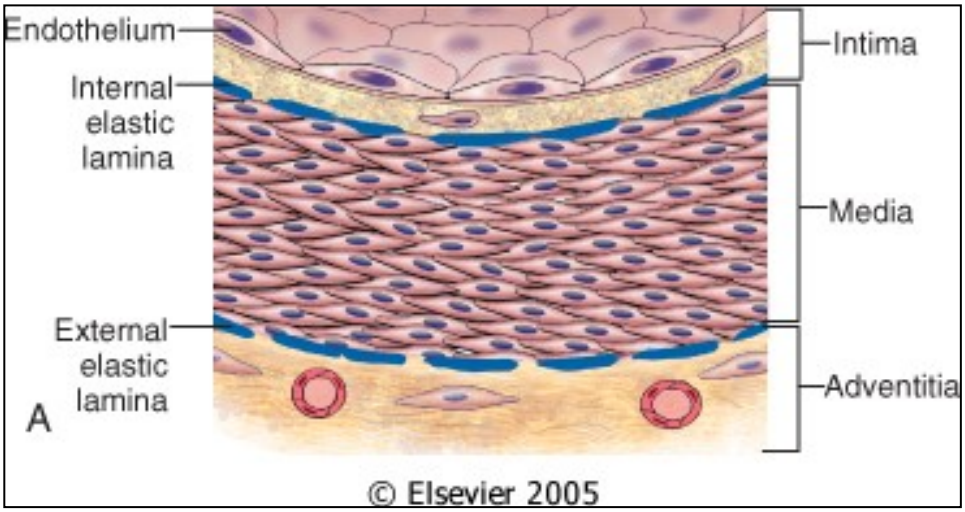

Figure 2.3: Diagram of the arterial wall.
***Adapted from: Schoen FJ, MD. Blood Vessels. In: Kumar V, Abbas AK, Fausto N, Robbins SL, Cotran RS, editors. Robbins and Cotran pathologic basis of disease. 7th ed. Philadelphia: Elsevier/Saunders; 2005.**

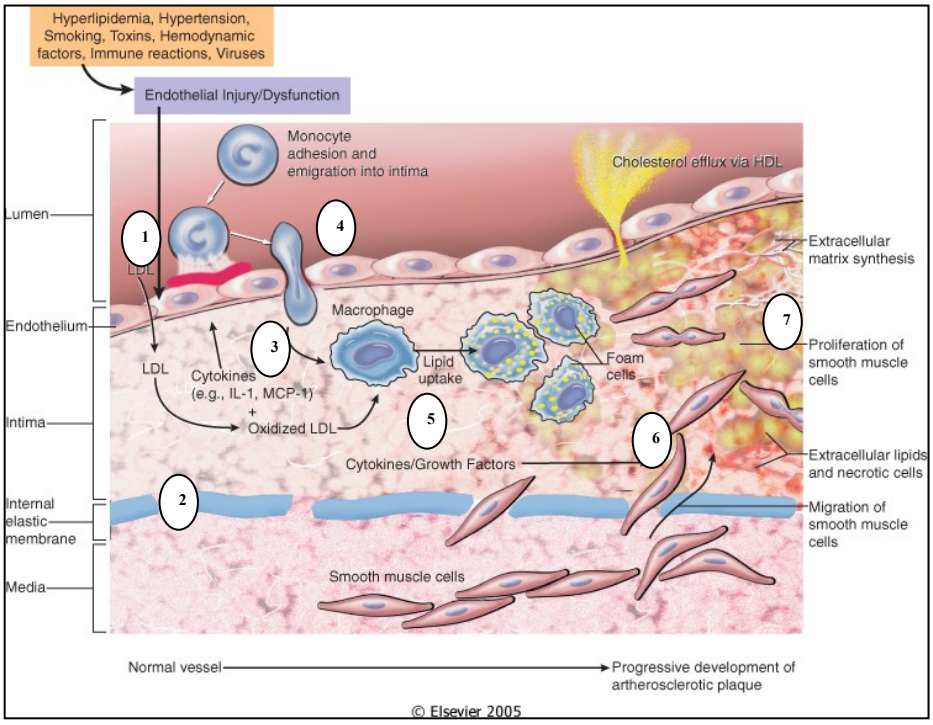

Figure 2.4: Schematic diagram for the progression of atherosclerotic plaque
***Adapted from: Schoen FJ, MD. Blood Vessels. In: Kumar V, Abbas AK, Fausto N, Robbins SL, Cotran RS, editors. Robbins and Cotran pathologic basis of disease. 7th ed. Philadelphia: Elsevier/Saunders; 2005.**

2.4.2.2 Pathogenesis of atherosclerosis

In the contemporary view, the pathogenesis of atherosclerosis is described by a concept known as the response to injury hypothesis [19]. In this view, atherosclerosis progression occurs in the following steps: 1) endothelial dysfunction, 2) lipoprotein entry and modification, 3) recruitment of leukocytes and 4) recruitment of smooth muscle cells [18].

Endothelial dysfunction

A primary step in the earliest stages of atherosclerosis is endothelial dysfunction. This occurs as a result of injury to the endothelial cells. An important factor in endothelial dysfunction is exposure to a "toxic environment" [18], represented for example, by smoking, abnormal lipid levels and diabetes. In the presence of a toxic chemical environment, the endothelium increases its production of reactive oxygen species which then alter the normal structural, metabolic and signalling functions that prevail in the healthy artery.

Lipoprotein entry and modification

Lipoproteins transport fats through the bloodstream. The levels of one type of lipoprotein, low-density lipoprotein, are positively correlated with atherosclerosis. In a state of endothelial dysfunction, the ability of the endothelial cells to prevent entry of molecules into the intima is significantly reduced. In this situation, and especially if LDL levels are high, LDL is able to enter the intima (Figure 2.4, step 1). Once inside the intima, LDL binds to components of the extracellular matrix, where it becomes trapped, and therefore, has an increased residence time. During this time, LDL undergoes various modifications which, importantly, include oxidation, and are critical to the development of atherosclerosis (Figure 2.4, step 2).

Recruitment of leukocytes

The modified LDL (mLDL) has various chemoattractant properties which attract and enable the adherence of leukocytes, primarily monocyte and t lymphocytes, to the luminal surface of the vessel wall. In response to the entry and modification of LDL,

there is also an increased expression of specific cytokines which in turn promote and increased endothelial expression of leukocyte adhesion molecules that bind the monocytes (Figure 2.4, step 3). Once these monocytes have adhered to the intima, they can penetrate into the subendothelial space by slipping through the junctions of the endothelial monolayer. The cytokines also promote expression of chemokines that direct the movement of the leukocytes into the intima. In the subendothelial space, the monocytes differentiate into macrophages (Figure 2.4, step 4) which are capable of engulfing large quantities of LDL because their receptors are unaffected by negative feedback inhibition. The lipid-laden macrophages are known as foam cells and are a primary constituent of the atherosclerotic lesions known as fatty streaks (Figure 2.4, step 5).

Recruitment of smooth muscle cells

The foam cells and injured endothelium activate platelets alongside producing various substances that stimulate the migration of smooth muscle cells into the intima. The activated platelets also stimulate migration of the smooth muscle cells into the intima (Figure 2.4, step 6). Once inside the intima, the smooth muscle cells proliferate and produce constituents of extracellular matrix (Figure 2.4, step 7). During this process of atherosclerosis progression, there is an increase in the intima thickness. Fibrofatty plaque that is present in advanced atherosclerosis consists of the accumulated smooth muscle cells, foam cells, leukocytes and a fibrous cap of extracellular matrix (Figure 2.5). The term fibrofatty plaque is used interchangeably with various other terms; namely, atheroma, atheromatous and fibroatheromatous [18]. While typical atheromas consist of abundant lipid, many fibrous plaques are composed primarily of smooth muscle cells and fibrous tissue [18].

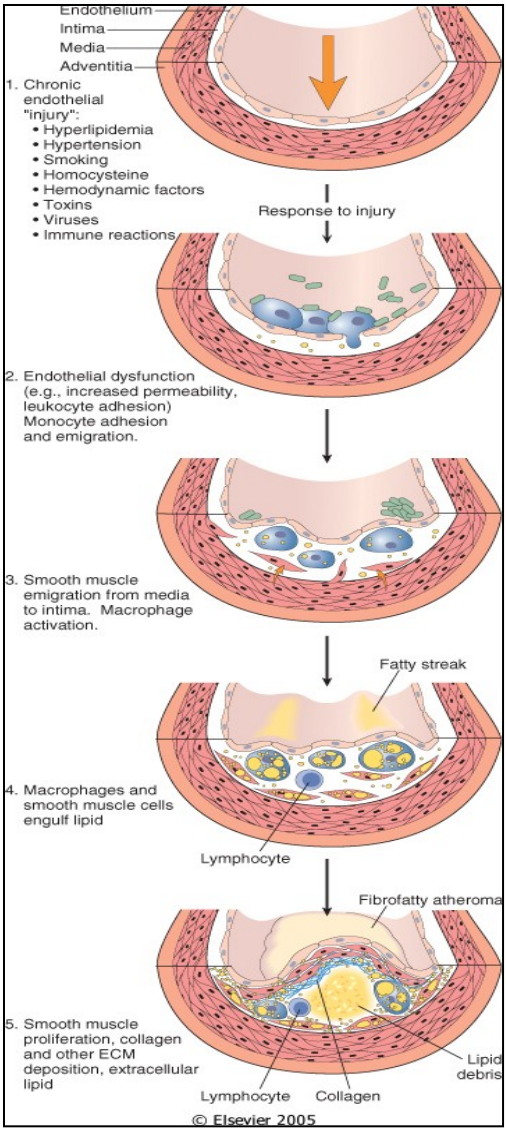

Figure 2.5: Atherosclerosis progression to a fibrofatty atheroma
***Adapted from Schoen FJ, MD. Blood Vessels. In: Kumar V, Abbas AK, Fausto N, Robbins SL, Cotran RS, editors. Robbins and Cotran pathological basis of disease 7th ed. Philadelphia: Elsevier/Saunders; 2005**

2.4.2.3 Atherosclerosis and thrombus formation

Fibrous plaques often contain a necrotic core of cell debris and foam cells (Figure 2.6) [18, 19]. This core is excessively thrombogenic because the foam cells produce tissue factor which can activate the coagulation pathway when contact is made with components in the blood. The fibrous cap and endothelium protect the atheromatous plaque and in this way prevent the thrombogenic core containing tissue factor from the circulation. Rupture or

ulceration of the fibrous plaque may expose the thrombogenic material to the blood, resulting in the formation of a thrombus which can occlude the artery and cause an infarction of the involved organ. Epidemiological studies have shown that the degree of artery narrowing viewed by angiography is not highly correlated with myocardial infarction. It has been observed that myocardial infarction has a greater association with plaque rupture and subsequent thrombosis, as compared to vessel occlusion caused by the progression of fibrous plaque. It is postulated that the tendency of plaque to rupture is partly influenced by the thickness of the fibrous cap that serves as a barrier between the foam cells containing tissue factor and the blood in circulation [18]. Although atherosclerotic plaque that contains a thick fibrous cap may cause a high degree of luminal narrowing, this type of plaque has a comparatively low susceptibility to rupture. This is known as 'stable plaque'. Conversely, 'vulnerable plaques' consist of a thinner fibrous cap, and show less luminal narrowing upon angiography [18]; however, these types of plaques are more closely associated with acute coronary events.

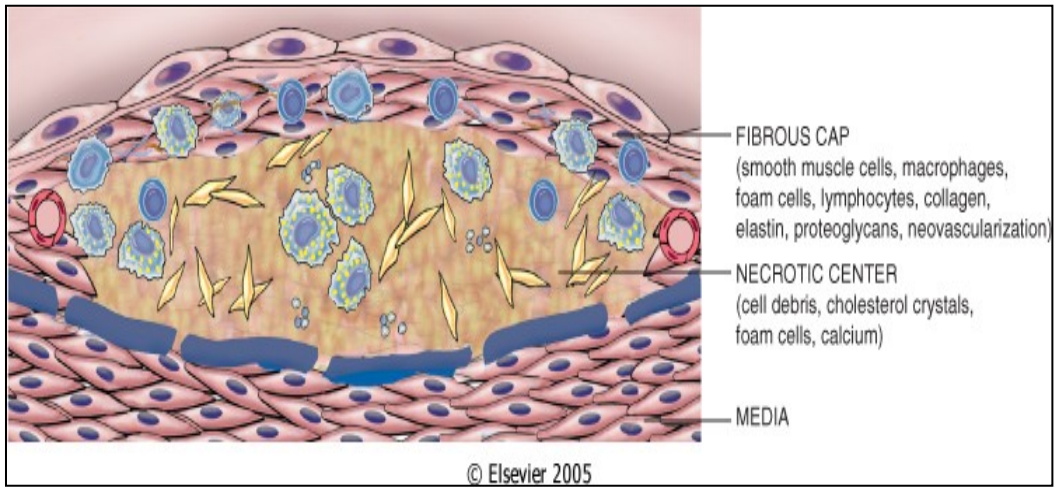

Figure 2.6 Necrotic core of a fibrofatty atheroma
***Adapted from Schoen FJ, MD. Blood Vessels. In: Kumar V, Abbas AK, Fausto N, Robbins SL, Cotran RS, editors. Robbins and Cotran pathologic basis of disease. Philadelphia: Elsevier/Saunders: 2005**

Direction of growth of the atherosclerotic lesion

As the presence of atherosclerotic plaque progresses, the arteries initially expand outward, to accommodate the plaque [18]. This outward growth is termed *positive remodelling* or *compensatory enlargement* and likely involves the turnover of extracellular matrix molecules so that the arterial lumen can increase in size. Luminal stenosis tends to occur only after the plaque burden is greater than approximately 40 percent of the cross-sectional area of the artery.

2.5 Vascular Imaging in Atherosclerosis and Associated Surrogate Outcomes

There has been a growing trend in the imaging of atherosclerosis to evaluate treatment related changes in surrogate outcomes. Vascular imaging directly evaluates atherosclerosis progression and also, therefore, displays the net effect of genetic and environmental factors on the arterial wall [4]. Clinical trials have been using various imaging modalities to evaluate the effect of new drugs on change in atherosclerotic disease status, most often measured as a continuous variable [4]. In these imaging trials, the expectation is that treatments which stop or slow the progression of, or result in a regression of, atherosclerosis will lead to a reduction in major cardiovascular events. Among the various imaging modalities that have been used to evaluate the effectiveness of treatments on atherosclerosis, three well-known technologies include quantitative coronary angiography (QCA), b-mode ultrasound and intravascular ultrasound (IVUS) [4]. A report from the 2009 Canadian Biomarkers and Surrogate Endpoints Symposium indicated that surrogate outcomes provided by QCA, b-mode ultrasound and IVUS are "established surrogate outcomes" [2]. This section provides a description of each of these technologies and the surrogate outcomes that they measure, along with each technology's strengths and limitations.

2.5.1. Quantitative coronary angiography

2.5.1.1. Description of technique

Often considered as the gold standard for diagnosis of CHD, QCA has also been performed in clinical trials to evaluate the effect of treatment on the progression of atherosclerosis. QCA was the first vascular imaging technique to be used in clinical practice and requires a patient to undergo cardiac catheterization. In this procedure, a catheter is guided through an appropriately selected vessel, usually the femoral artery, and into the coronary arteries [3, 20-22]. A radio-opaque contrast agent is injected into the coronary arteries and X-ray images are then recorded either on film or on a digital detector. QCA provides a two-dimension (silhouette) view of the arterial lumen [2]. In clinical trials, treatment related changes in atherosclerotic lesions can be evaluated by performing serial measurements along a coronary artery or segment [21, 22]. For images obtained before and after treatment, a side-by-side analysis of the images is performed to determine the change in atherosclerotic lesions.

2.5.1.2 Surrogate outcomes provided by QCA

Surrogate outcomes that have been measured to evaluate the effect of treatment include the following:

- Change in percent stenosis, where percent stenosis is determined in relation to a nearby "normal" reference segment [3].
- Change in the mean of the minimum lumen diameter (MLD), which is also known as the minimum obstructive diameter (MOD) [20].
- Global change scores, determined by an expert panel who visually evaluates angiographic changes over time and assigns a score for the overall degree of change [20].

2.5.1.3 Advantages and limitations of QCA

The advantage of performing QCA is that it provides a two-dimensional image of the arterial lumen and in this way provides information about the degree of stenosis. However, this technique presents with several important limitations. Importantly, QCA is

invasive, involves exposure to X-ray radiation and may lead to cardiovascular events from procedural complications [20]. In addition, QCA only provides information on the vessel lumen, and it is now accepted that atherosclerosis is primarily a disease of the vessel wall, rather than the vessel lumen [20]. As mentioned earlier, during the progression of atherosclerotic lesions, the artery initially expands via positive remodelling and stenosis only occurs in the advanced stages of atherosclerosis. As such, QCA is unable to detect new non-stenotic lesions and is also unable to detect advanced lesions that are present on the full length of an artery [20, 22]. Angiography poorly detects vulnerable plaque, which is associated more closely with cardiovascular events than the stable plaque which causes luminal narrowing [20]. Also, with the change in angiography from film to pixelated digital images, there has been a reduction in spatial resolution needed to detect a treatment effect [3]. This limitation is particularly emphasized for trials in which the comparator is an active control. Due to the risks associated with its invasiveness, and the exposure to x-ray radiation, QCA is not generally performed in asymptomatic individuals [20].

2.5.2. B-mode ultrasound

2.5.2.1. Description of technique

B-mode ultrasound is a non-invasive ultrasound based technique used to visualize the lumen and walls of selected arteries, which include the carotid, aorta and femoral arteries [22]. Most of the evidence on atherosclerosis that has been obtained by b-mode ultrasound is based on evaluation of the carotid arteries [3]. In this technique, an ultrasound transducer is placed over the skin above extracranial segments of each of the carotid arteries. Ultrasound images are recorded from multiple segments. The distance between arterial lumen-intima boundary, and media-adventitia boundary can be identified from an image of the carotid artery [3] (Figure 2.8).

2.5.2.2 Surrogate outcomes provided by b-mode ultrasound

Surrogate outcomes measured by b-mode ultrasound include the following:

- Change in mean carotid intima media thickness (CIMT). The mean carotid intima thickness is determined by finding the mean of multiple values from the same artery or different branches [20].
- Change in maximum carotid intima media thickness (CIMT) [3].

2.5.2.3 Advantages and limitations of B-mode ultrasound

An important advantage of B-mode ultrasound is that unlike QCA, it enables measurement of the vessel wall and the lumen. In this way, the use of b-mode ultrasound has greater sensitivity in detecting early atherosclerosis as compared to angiography [3]. Additionally, b-mode ultrasound is non-invasive, does not expose patients to x-ray radiation, is relatively simple to perform and is inexpensive. A number of limitations have been identified in the use b-mode ultrasound. While this technique is associated with high inter- and intra-observer reproducibility in well controlled research settings, there is currently a lack of consensus with regards to these protocols in the clinical settings. This limitation prevents its routine use [3]. Improvements in medical treatments have slowed the progression CIMT. For this reason, future trials using CIMT as a surrogate endpoint may require larger sample sizes, longer-follow-up periods and increased costs, reasons that may impede its use as a surrogate outcome [3].

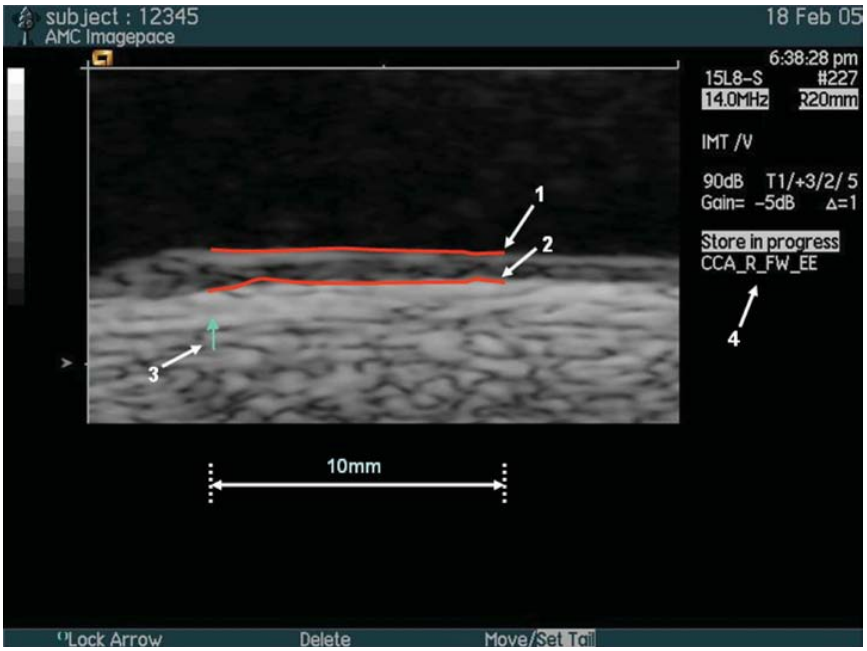

Figure 2.7: B-mode ultrasound image of the carotid intima media thickness. Line 1 indicates the lumen-intima interface and line 2 indicates the media-adventitia interface.
***Adapted from: Kastelein JJ, de Groot E. Ultrasound imaging techniques for the evaluation of cardiovascular therapies. *Eur Heart J* 2008 Apr;29(7):849-58**

2.5.3. Intravascular ultrasound (IVUS)

2.5.3.1. Description of technique

In this method, a miniaturized transducer is connected to the end of a catheter which is advanced through a selected coronary artery. With a mechanical pull-back of the catheter at a fixed rate of 0.5mm/s, the transducer rotates at 1800 rpm, allowing a 360° characterization of the vessel wall thickness through the acquisition of serial images [20]. Approximately 30 images/s can be acquired. IVUS images are analyzed either manually or with the use of semi-automated systems, by outlining the intimal lining of the vessel lumen and the external elastic membrane that separates the media from the adventitia [20].

2.5.3.2 Surrogate outcomes provided by IVUS

The following surrogate outcomes are provided by IVUS:

- Change in total atheroma volume (TAV) [3]. TAV is calculated as:

TAV =Σ(EEM_{CSA} −$LUMEN_{CSA}$) where, EEM_{CSA} is the external elastic membrane cross-sectional area, and $LUMEN_{CSA}$ is the luminal cross-sectional area cross-sections.

- Change in percent atheroma volume [3]. Percent atheroma volume is calculated as follows: PAV = [Σ (EEM_{CSA} −$LUMEN_{CSA}$) / Σ EEM_{CSA}] X 100.

2.5.3.3 Advantages and limitations of IVUS

An important advantage of IVUS as compared to QCA is that images obtained from this technique can depict positive remodelling of the artery and in this way provide information about the presence of atherosclerosis before it causes luminal narrowing [3]. In addition, IVUS has greater sensitivity as compared to QCA in the detection of stenosis [3]- IVUS can provide information about diffuse advanced lesions and unstable plaque. Unlike b-mode ultrasound, IVUS provides a direct visualization of the coronary vessels. IVUS allows the evaluation of plaque volume, which may be a better outcome measure than CIMT in evaluating the effect of therapy on changes in atherosclerosis. This is primarily because, as mentioned above, changes in CIMT occur slowly, unlike changes in plaque volume. In one study, large differences in the change in plaque volume were noted in patients who were randomized to either placebo or atorvastatin at a dose of 80mg/day for three months [2]. Although this potential advantage exists, IVUS has not been established as a superior method to b-mode ultrasound in evaluating treatment effects on atherosclerosis. The major limitation of IVUS is that it is an invasive procedure, and therefore presents with risks to patients. However, this technique is associated with a very low incidence of complications in properly selected individuals. IVUS is also very costly, requires a steep learning curve to use, and has a complicated set up when used occasionally. Additionally, IVUS measurements cannot currently evaluate lumen diameters than are smaller than 1.00 mm.

3 REVIEW OF TRIAL-LEVEL STATISTICAL METHODS FOR SURROGATE OUTCOME EVALUATION

3.1 Overview

The purpose of this chapter was to identify and describe statistical methods for surrogate outcome evaluation at the trial-level. The key feature of methods identified in this chapter is that they only require the availability of summary statistics for surrogate and clinical outcomes measured in response to therapy, and do not require individual patient data. First, a systematic review of the literature was performed to identify reports of relevant statistical methods. Second, a narrative synthesis was performed to describe the methods.

3.2 Objective

The objective of the review was:

- To identify reports of statistical methods that quantify the strength of a surrogate outcome using summary data from clinical trials.

3.3 Methods

3.3.1 Criteria for selecting methods

3.3.3.1 Types of studies

Studies were selected if they fulfilled any of the following criteria:

- The article was an original methodology paper describing a meta-analysis methodology that can be applied using summary level data alone, and does not require the availability of individual patient data.
- The article was a review paper containing information on a meta-analytic method for surrogate outcome evaluation that requires summary data only.
- The article discussed or evaluated a meta-analytic method for surrogate outcome evaluation that required summary data only.
- The article was an application of a meta-analytic method to summary data for the evaluation of a surrogate outcome.

In addition, all articles had to fulfill the following criteria:

- The proposed method was applicable to the situation in which the surrogate outcome can be modelled as a continuous variable.
- The proposed method was developed for data from a prospective controlled clinical trial.
- The proposed method was transparent and feasible to implement (e.g. could be applied easily using SAS software; accessible instruction describing how the method could be implemented using the SAS software).
- an a priori decision was made to only consider for this work, the frequentist framework.

The review also included review papers as well as papers discussing/evaluating existing methods and applications because these papers could provide additional insight into existing methodologies. For instance, review papers may clarify approaches that would not otherwise be considered transparent and could provide references to additional methods that may not be identified from the review; discussion/evaluation and application papers could provide additional information on strengths and limitations of existing methods as well as information about simplified strategies that can be adopted so that meta-analytic methods based on individual patient data can be implemented using summary data only.

3.3.2 Search methods for identification of studies

The following sources were searched to identify published reports describing a statistical methodology for surrogate outcome validation:

- Chapters 7 through 9 of the book by Burzykowski et al [7]. This book is a leading text in the field of surrogate validation and the primary purpose of this text was to provide a comprehensive description of the major statistical methodologies that have been proposed to evaluate surrogate outcomes. The chapters identified were searched because they were the only chapters which were likely to describe methodologies from published studies that fulfilled the inclusion criteria listed above. The book was published in 2005 and it is expected that this books contains the majority of statistical methods proposed to this date.

- Statistics reports included in Section 2.5 and Table 2 of a report published by Lassere [11]. This report provides a detailed review of the literature related to surrogate outcome evaluation. The review identifies statistical and other literature written in this field. Section 2.5 and Table 2 of this report provide a comprehensive list of published reports of statistical models that can be used to evaluate surrogate outcomes. The methods were identified from a comprehensive search strategy that covered the time period ranging from 1930 – April 2007.
- An electronic search of MEDLINE (2007- February week 4 2010). The search strategy was constructed to identify any additional methods that may have been proposed since 2007 and that would not have been included in the work by Burzykowski et al [7], and Lassere [11]. Because methods papers are difficult to search for [23], and MEDLINE does not have an indexing term for surrogate outcomes [11], a combination two different sets of search terms (step 19 and 23 of the search strategy below) were developed in MEDLINE. The results were then combined (step 25 of the search strategies below). The search strategy was developed by modifying the strategies provided in the paper by Lassere [11] and by reviewing the text words as well as indexing terms that had been applied to key methods papers which were known to meet the eligibility criteria from background knowledge of this project. The search strategy can be found in Appendix A.
- Bibliographies of selected articles were searched, and the database SCOPUS was used to identify additional articles that had cited the articles that were selected for inclusion in the review.
- The SCOPUS database was also used to identify additional commentaries and errata for the eligible studies.

3.3.3 Data collection and analysis

3.3.3.1 Study selection

The title and abstracts of identified reports from the work of Burzykowski et al [7], Lassere [11] and MEDLINE were transferred and stored in Reference Manager (version 12). Duplicate and non-English language reports were deleted and the titles and abstracts of the identified reports were screened. Titles and abstracts that suggested fulfillment of the

eligibility criteria included combinations of statements such as: 'continuous surrogate outcome', 'bivariate', 'meta-analysis' and related terms, 'trial-level summary' and related terms, 'linear regression', and 'multiple outcomes'. Studies that clearly did not meet the inclusion criteria were excluded. The full text of the remaining studies was then evaluated for inclusion. The bibliographies of articles selected for inclusion were searched, cited reference searching was performed to identify additional relevant reports and commentaries and errata for eligible studies were retrieved.

3.3.3.2 Data extraction

The following information was extracted from the studies:

- Type of study: original methodology article; review; discussion/evaluation of existing methodology; application.
- Statistical method/model of interest to this thesis work
- Relevant statistic(s)
- Detailed description of relevant statistical models/method if provided in the article
- Focus of the article with regards to the methodology of interest.
- Relevant programming code, if available.

3.4 Results

3.4.1 Results of the search

Figure 3.1 illustrates how the studies were selected. The search strategy identified a total of 608 studies. From these studies, 17 full-text articles were eligible for inclusion in the review [24-40]. The studies consisted of original methods papers, discussions/evaluations, and applications. In total, these studies provided information about 5 different procedures for evaluating surrogate outcomes. Four additional commentaries were retrieved from the electronic search performed in the SCOPUS database to identify commentaries and errata for the studies meeting the eligibility criteria [41-44]. Searching the reference lists of identified studies yielded an additional 6 reports [45-50]. The search of the SCOPUS database for additional articles that had cited the included studies retrieved a further 2 articles [51, 52]. While the number of additional studies identified from searching reference

lists and citing references (8 in total) was relatively large in comparison to the number identified from the original search strategy, these additional studies did not provide information on any other surrogate evaluation methods that had not already been identified from the original search strategy- the two articles identified from reviewing the citing articles were applications. Further to this, the reports identified from the reference lists were published prior to 2007 and most of them were likely not included in the work of Burzykowski et al [7] and Lassere [11] because while relevant, these studies did not provide critical information required to understand the methodological aspects of the surrogate evaluation procedures contained in these reports. 4 of the 6 reports identified from the reference lists were applications. The search of the reference lists of included studies identified one important methodology paper associated with several application studies identified from the original search strategy [45]. The reference list searching also identified an early methodology that was later modified and described in one of the included articles retrieved from the original search [32, 50].

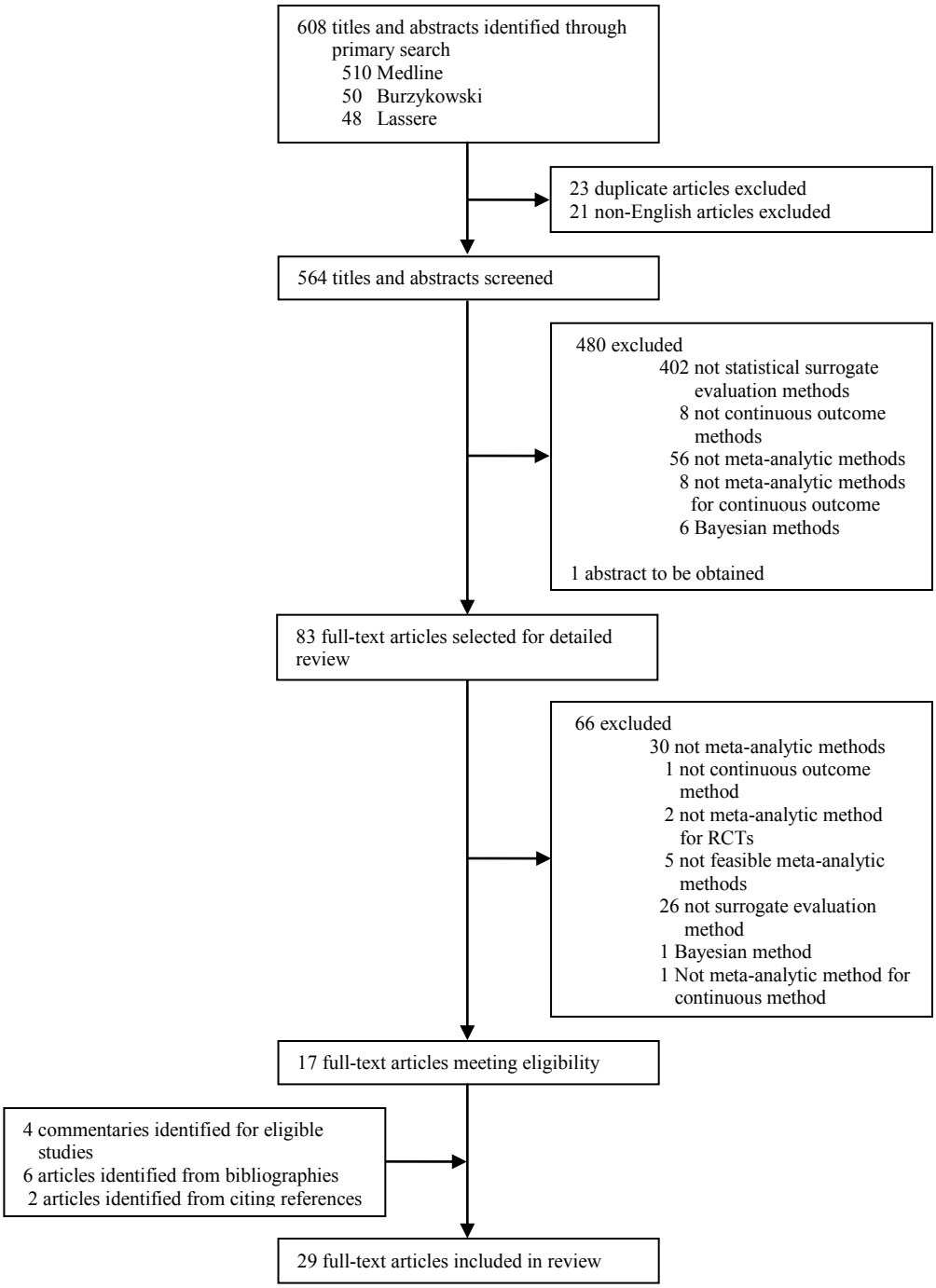

Figure 3.1: Flow diagram for selection of studies reporting surrogate outcome validation procedures

3.4.2 Description of included analytical procedures

For the analytical procedures identified from the review, information is provided for: (1) the contribution of the article identified in the review to the procedure; (2) the description of the procedure; (3) any assumptions; and (4) any measures of association for the relationship between the surrogate and clinical outcome.

3.4.2.1 Significance testing for study wise agreement

Contribution to the procedure

Begg and Leung [24] proposed the principle of study wise agreement for the evaluation of surrogate outcomes. This approach has been evaluated and discussed by Green et al[26], Hughes [29] and Johnson et al [30]. This method has been used as one of several procedures for the evaluation of surrogate outcomes in oncology by Sherrill [52], Miksad [35], and Tang [39]. Fleming applied this method for surrogate outcome evaluation towards treatments for HIV [46].

Description of the procedure

The principle of concordance as described by Begg and Leung [24] states that in the context of randomized trials, the validity of a surrogate outcome depends on the concordance between the results of a statistical analysis performed to evaluate the effect of treatment on the surrogate outcome, and the results of the analysis on the clinical outcome. The authors have not recommended a particular procedure to evaluate the effect of treatment on the surrogate and clinical outcome; however, Begg and Leung [24] have identified hypothesis testing as one method. In this approach, a significance test is chosen according to the classification of the variable (e.g. continuous or dichotomous) representing the surrogate and clinical outcome. In the application of this principle to the meta-analytic setting, for each of $i = ,..., N$ randomized trials, X_{ij} and Y_{ij} represents the observed surrogate and clinical summary outcome for treatment arm $j = $ or 2 of trial i. For the surrogate outcome, the relevant hypothesis uses a significance level of 5% to determine whether there is a significant difference between X_{i1} and X_{i2}. Similarly, the chosen

hypothesis test evaluates the significance of the differences between treatments on Y_{i1} and Y_{i2}.

Assumptions

The duration of follow-up for X_{ij} will end earlier than the follow-up period for Y_{ij}; other assumptions will vary according to the hypothesis test used to evaluate the outcomes.

Study-wise percent concordance as a measure to evaluate surrogacy

For the i^{th} trial, the directionality of the treatment effect on the surrogate outcome replicates (i.e. predicts) the effect on the clinical outcome if the hypothesis test on the surrogate outcome yields the same p-value as the significance test on the clinical outcome, where $P <$).05 or $P \geq$).05 for both outcomes. For the set of $i =$,..., N trials, percent concordance is calculated as the proportion of trials for which the p-values for the test on the surrogate and clinical outcomes are the same.

3.4.2.2 Rank correlation for the correlation coefficient

Contribution to the procedure

While rank correlation analysis has not been proposed as a formal method to evaluate surrogate outcomes, in a review of various methods, Shi et al have identified rank correlation analysis as one procedure that has been used in several applications [37]. In each of these applications, rank correlation analysis has been used as a supplementary procedure alongside other surrogate evaluation methods. Green et al [26], Tang et al [39] and Montagnani et al [42] have used rank correlation analysis as a supplementary method towards surrogate outcome evaluation in oncology.

Description of the procedure

Rank correlation analysis has been used as a non-parametric method to evaluate the correlation between treatment differences on the surrogate and clinical outcome. In the first step of the rank analysis, the original data on the outcomes is replaced by ranks. The

second step of the analysis determines the rank correlation coefficient. In the rank analysis method, for a series of $i = ,..., N$ randomized trials, X_i and Y_i are variables representing the observed differences between treatment arms on the surrogate and clinical summary outcomes, respectively. The values for X_i and Y_i in the set of N trials are then ranked in order. The ranks assigned to X_i and Y_i are represented by x_i and y_i, respectively. Where two values of X_i or Y_i are the same, the average rank is assigned to each of the tied values.

Assumptions

The duration of follow-up for X_i will end earlier than the follow-up period for Y_i.

The spearman rank correlation coefficient as a measure of association

The rank correlation coefficient, r_s, (Spearman's ρ) quantifies the correlation between the ranks assigned to the observed values of the surrogate and clinical outcome. For the case in which there are no ties in the data, Spearman's ρ is given as follows:

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} \qquad [1]$$

Here, $d_i = x_i - y_i$

In the presence of ties, r_s is determined as follows:

$$r_s = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \qquad [2]$$

where $\bar{x}$ and $\bar{y}$ are the mean values of the ranks for the surrogate and clinical outcome.

3.4.2.3 Graphical analysis

Contribution to the procedure

Hughes et al [29]described a graphical approach for the evaluation of surrogate outcomes. In a review of various surrogate evaluation methods, Weir et al [40]indicate that the graphical approach may be used as a preliminary method to evaluate surrogate outcomes prior to the use of formal statistical method

Description of the procedure

In the graphical approach, the differences between treatments on the surrogate and clinical outcome are quantified and the association between these differences is evaluated while accounting for the precision of the measurements for each outcome. Hughes et al [29]indicated that the first step of such an analysis is the identification of a time-frame over which the association between the two outcomes will be evaluated. When identifying this time-frame, the duration of follow-up for the surrogate outcome should end earlier than the duration of follow-up for the clinical outcome. As an example of this time-frame, the assessment of treatment differences based on the surrogate outcome occurs at 6 months while treatment differences based on the clinical outcome are evaluated at 2 years. The second step of this process is the identification of a set of $i =$,..., N trials which contain summary data for the treatment differences on each outcome at the two follow-up durations identified in step 1. In the third step, summary estimates for treatment differences and the associated standard errors are obtained from each of the N trials for both outcomes. For the i^{th} trial, $X_i(\sigma_{\cdot\cdot i})$ and $Y_i(\sigma_{\cdot_i})$ represent the observed differences (standard errors) between the treatment arms on the surrogate and clinical summary outcomes, respectively. A graph is then created with the values of $X_i = Y_1,..., X_N$ plotted on the x-axis against the values of $Y_i = {}^{\prime}_1,..., Y_N$ on the y-axis. In this way, each point on the graph represents a randomized comparison between two treatments from a single trial. The standard errors are used to construct 95% confidence intervals for each point on the graph. In the approach recommended by Hughes et al [29], these confidence intervals should extend both vertically, to represent the precision of the Y_i values, and

horizontally, to represent the precision of the X_i values. The 95% confidence intervals are shown in both directions because it is possible that the differences between treatments on the surrogate outcome are estimated with greater or less precision than the differences on the clinical outcome. Hughes et al indicate that the data used in the graphical approach may also be subjected to a statistical analysis, but the authors have not described a particular approach.

Assumptions

The duration of follow-up for X_{ij} will end earlier than the follow-up period for Y_{ij}

3.4.2.4 Regression analysis for R^2, and STE

Contribution to the procedure

In the context of summary data obtained from prospective controlled clinical trials, regression analysis, and most commonly the weighted least squares approach, has been used extensively for the evaluation of surrogate endpoints. This review identified applications in oncology [27, 28, 30, 35, 39, 48, 49, 52], hypertension [34, 51], multiple sclerosis [38], HIV [36], postmenopausal osteoporosis [47] and cardiology [31]. Burzykowski et al [45]described a measure of association that can be used to evaluate the strength of a surrogate outcome based on the fitted regression model. Green [26], Lassere [33] and Shi [37] have evaluated and discussed the use of regression analysis for surrogate outcome evaluation. Perrone [43], Rich [44]and Montagnani [42] have provided commentaries for two of the applications. In the next section, the method of least squares is described and is followed by a description of the weighted least squares analysis.

Description of procedure

Method of least squares

In this setting, X and Y represent the differences between treatment arms on the surrogate and clinical outcome, respectively. Ordinary least squares regression models a straight line relationship between the surrogate and clinical outcome. The simple linear regression equation is given as follows

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad [3]$$

where, β_0 is the Y-intercept, or the value of Y when $X = 0$; β_1 is the slope, or the expected change in Y relative to one unit change in X; and ε is the random error. In a series of $i = 1,..., N$ controlled trials, the independent variable X_i represents the observed differences between the treatment arms on the surrogate outcome. Similarly, the dependent variable Y_i represents the observed differences between the treatment arms on the clinical outcome.

The estimates of β_0 and β_1 for which the sum of squared errors, $\sum_{i=1}^{N} \varepsilon_i$, is a minimum are represented by $\hat{\beta}_0$ and $\hat{\beta}_1$. These least squares estimates are determined according to the following formulas:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \qquad [4]$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \qquad [5]$$

where, for the set of N trials, $\bar{Y}$ is the sample mean of the observations $Y_i = Y_1,..., Y_N$; $\bar{X}$ is the sample mean of the observations $X_i = X_1,..., X_N$. The estimate of the simple linear regression equation is then given as:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad [6]$$

where, $\hat{Y}_i$ is the estimated treatment difference on the clinical outcome at X_i

Method of weighted least squares

Rather than performing the least squares analysis, investigators often perform a weighted least squares regression analysis to determine values for R^2 and *STE;* this is since the individual trials may be of varying sample sizes, leading to a large amount of variability

in the spread (or precision) of the distribution of Y for different X values. In this way, the assumption of homoscedasticity for the ordinary least squares analysis may be violated because the variance of Y varies for different values of the independent variable X. The weighted least squares analysis is similar to the general method of meta-analysis in which the overall pooled estimate of treatment effect is a weighted average. In the weighted least squares methodology, the variance, σ , of Y_i, is either known or else it is assumed that $\sigma_{\cdot} = \tau / W_i$. Here, W_i is the weight assigned to the i^{th} trial and is equal to the trial size (i.e. the number of patients in trial i). For the straight line regression case, i.e. $Y = \beta_0 + \beta_1 X + \varepsilon$, the specific weighted least-squares estimates of β and $\beta_{\cdot}$, are represented by β and $\beta_{\cdot}$, and are given by the following formula:

$$\beta_{\cdot} = \frac{\sum_{i=}^{N} W_i (X_i - \bar{X}')(Y_i - \bar{Y}')}{\sum_{i=}^{n} W_i (X_i - \bar{X}')^2} \qquad [7]$$

$$\beta_{\cdot} = \bar{Y}' - \beta_{\cdot} \bar{X}' \qquad [8]$$

where, $W_i = \frac{1}{\sigma_{\cdot}}$ when σ is known or else W_i is the sample size of the i^{th} trial when it is assumed that $\sigma_{\cdot} = \tau / W_i$.

Assumptions

Method of least squares

For any given value of X, Y is normally distributed with mean $\mu_{\cdot|X}$ and variance $\sigma_{\cdot|X}$; the Y-values are independent of each other; $\mu_{\cdot|X}$ is a straight-line function of X; the variance of Y is the same for any given value of X, i.e. $\sigma_{Y|X} \equiv \sigma$;

Method of weighted least squares

For the case in which $\sigma_{\cdot} = \tau / W_i$, it is assumed that W_i is known.

The coefficient of determination R^2 as a measure of association

Method of least squares

The coefficient of determination, R^2, determines the strength of the linear relationship for the difference between treatments on the observed surrogate summary outcome and the difference on the observed clinical summary outcome. R^2 is calculated by the following formula:

$$R^2 = \frac{SSY - SSE}{SSY} \quad [9]$$

where SSY represents the sum of the squares of deviations associated with $\bar{Y}$ and SSE is the residual sum of squares. SSY and SEE are determined according to the following formulas:

$$\text{SSY} = \sum_{i=}^{n} (Y_i - \bar{Y})^2 \quad [10]$$

$$\text{SSE} = \sum_{i=}^{n} (Y_i - \hat{Y})^2 \quad [11]$$

For the fitted least squares regression line, $\hat{Y}_i = \beta_0 + \beta_1 X_i$, [9] illustrates that R^2 may be interpreted as the proportionate reduction in the sum of the squares of vertical deviations obtained through use of the least squares regression line, rather than the naive model $\hat{Y} = \bar{Y}$, which does not account for the difference between treatments in their effect on the surrogate outcome.

Method of weighted least squares

For the weighted least squares regression analysis, R^2 is determined as follows:

$$R^2 = \frac{SSY^* - SSE^*}{SSY^*} \quad [12]$$

where,

$$SSY^* = \sum_{i=}^{N} W_i (Y_i - \bar{Y}')^2 \quad [13]$$

$$SSE^* = \sum_{i=}^{N} W_i (Y_i - \quad)^2 \qquad [14]$$

Surrogate threshold effect (STE) as a measure of association

Method of least squares

Burzykowski et al proposed the surrogate threshold effect (STE), defined as the "minimum treatment effect on the surrogate necessary to predict a non-zero effect on the true endpoint" [45]. Quantitatively, the STE is determined for a new trial, $i = \ )$, based on the fitted least squares regression line. The STE is associated with the difference between the treatment arms on the surrogate outcome, $X_i = \ Y_0$, that is needed to predict a non-zero difference in the clinical outcome. The point estimate used for the prediction of the clinical outcome is $\hat{Y}_{X_0} = \ 3 \ + \ 3 \ \ Y_0$ and the STE is determined by constructing prediction bands on a plot of the fitted least squares regression line $\hat{Y} = \ 3 \ + \ 3 \ \ Y$.

The surrogate threshold, or point beyond which predictions for the difference between treatments for the clinical outcome, $\hat{Y}_{X_0}$, are significantly greater than zero, is the point at which the lower prediction band crosses the horizontal axis. The formula of the prediction interval for $\hat{Y}_{X_0}$ is given as the following:

$$\overline{Y} + \beta \ \ X_0 - \ \overline{Y}) \pm \ _{n\text{-}2,1\text{-}\alpha \ 2} S_{Y|X} \sqrt{1 + \frac{1}{N} + \frac{\ X_0 \quad \overline{X})^2}{(N \quad 1) S_X^2}} \qquad [15]$$

where σ is replaced by its estimate, $S_{Y|X}^2$. The prediction bands are constructed by determining the prediction intervals over a range of X_0 - values . The values of X_0 used to construct the prediction bands are the values for the independent *X* variable in the set of *N* clinical trials used to fit the least squares regression line. In formula 7, the value of 1 is added under the square root sign because the value of $\hat{Y}_{X_0}$ for an individual clinical trial will have more variability than the estimator for the overall mean $\mu_{\ _1 X_0}$.

Method of weighted least squares

For the weighted least squares regression analysis, the formula for calculating the prediction interval for $\hat{Y}'_{X_0}$ is given as follows:

$$\bar{Y}'+\beta \ W_0(X_0 - \bar{\Upsilon}') \pm {}_{N\ 2,1\ \alpha\ 2} S_{Y|X} \sqrt{1+ \frac{1}{N} + \frac{'W_0(X_0 - \bar{\Upsilon}')^2}{(N -)S_X^2}} \qquad [16]$$

3.4.2.5 Mixed models for β and Δ

Contribution to the procedure

Korn et al [32] have described a mixed effects model for the evaluation of surrogate endpoints. The models proposed by Korn et al [32] are developed based on the method of Torri et al [50] which was used to evaluate surrogate endpoints in oncology. Korn et al [32] consider two different settings for the evaluation of surrogate endpoints in the context of a series of trials containing trial-level summary data on the surrogate, *X*, and clinical, *Y*, outcome. In the first setting, there is no natural ordering in the treatment arms of each trial. For a two-arm trial, these unordered treatment arms may be represented by A vs B. In the second setting, the treatment arms are ordered. For a two-arm trial, these ordered treatment arms are represented by A vs A+B, meaning B in addition to A. The models applicable to the unordered treatment arms setting are described, and followed by the ordered treatment arm setting. Freedman has provided a review and commentary to this method [25, 41].

Description of the procedure

Unordered treatment arms

In a series of randomized trials, X_{ij} and Y_{ij} is the observed surrogate and clinical summary outcome for treatment arm *j* of trial *i*. In proposing a model in which the treatment arm level effects on the surrogate outcome variable are random, rather than fixed, Korn et al [32] first present the fixed effects model proposed by Torri et al [50]:

$$X_{ij} = \mu_{\cdot} + \cdot_{ij} \qquad [17]$$

$$Y_{ij} = \alpha_{\cdot} + \beta\mu_{\cdot} + \cdot_{ij} + \delta_{\cdot} \qquad \text{for } j = \,,\ldots, h_i, \text{ and } i = \,,\ldots, k$$

where, $\mu_{\cdot}$ is the mean surrogate response for treatment arm j of trial i. e_{ij} and $\delta_{\cdot}$ represent the random error, $\alpha_{\cdot}$ represents the trial-level effect in the clinical outcome and β is the regression coefficient representing the linear relationship between the surrogate and clinical outcome while adjusting for $\alpha_{\cdot}$. g_{ij} is a random effect which represents additional variability which is not accounted for by the model.

Model [17] is generalized by modeling the differences between each arm and an arbitrarily selected reference arm. In this generalization, $X_{ij}^{(D)} = X_{ij} - X_{i1}$ and $Y_{ij}^{(D)} = \cdot_{ij} - \cdot_{i1}$. From this,

$$X_{ij}^{(D)} = \mu_{\cdot}^{(D)} + \cdot_{ij}^{(D)} \qquad [18]$$

$$Y_{ij}^{(D)} = \beta\mu_{\cdot}^{(D)} + \cdot_{ij}^{(D)} + \delta_{\cdot}^{(D)} \qquad \text{for } j = 2,\ldots, h_i, \quad i = \,,\ldots, k$$

Korn et al [32] then use model [17] and [18] as a basis to develop models in which the treatment-arm-level effects on the surrogate outcome variable are considered random, rather than fixed:

$$X_{ij} = \mu_{\cdot} + m_{ij} + \cdot_{ij} \qquad [19]$$

$$Y_{ij} = \alpha_{\cdot} + \beta(\mu_{\cdot} + m_{ij}) + \cdot_{ij} + \delta_{\cdot} \qquad \text{for } j = \,,\ldots, h_i, \text{ and } i = \,,\ldots, k$$

where μ represent fixed trial-level effects for the surrogate outcome variable and m_{ij} represent within-trial treatment-arm-level random effects. The other terms are as in [17]

The generalized model based on treatment differences is given as follows:

$$X_{ij}^{(D)} = m_{ij}^{(D)} + \cdot_{ij}^{(D)} \qquad [20]$$

$$Y_{ij}^{(D)} = \beta\cdot_{ij}^{(D)} + \cdot_{ij}^{(D)} + \delta_{\cdot}^{(D)} \quad \text{for } j = 2,\ldots, h_i, \text{ and } i = \,,\ldots, k$$

Through several applications and simulations, Korn et al [32] propose that model [20] should be used to evaluate surrogate endpoints. Korn et al [32] indicate that a model based on treatment differences is preferable because it is associated with a fewer number of parameters. Further, the generalized model can be applied to the situation in which a relative measure more effectively describes the treatment effect, as compared to an absolute measure for each arm of a trial. The authors recommend the use of the random effects model because the estimator of β from this model is expected to be more efficient than the estimator based on the fixed effects model. The authors also show through some limited simulations that the fixed effects estimator has a larger amount of variability than the random effects estimator and provides some extremely large estimates of β in a very small proportion of the simulation results.

Ordered treatment arms

For the setting in which one of the arms in each trial can be considered the control arm, X_{i1} and Y_{i1} represent the observed surrogate and clinical summary outcome in trial *i*. The subscript $j =$ represents control arm variables. For this setting, Korn et al [32] propose the random effects model based on treatment arm differences in which j=1 is the reference arm:

$$X_{ij}^{(D)} = \mu \quad + \eta_{ij}^{(D)} + \quad _{ij}^{(D)} \qquad [21]$$

$$Y_{ij}^{(D)} = \quad + \beta \quad + \eta_{ij}^{(D)}) + \quad _{ij}^{(D)} + \quad ^{D)} \qquad \text{for } j = ,..., h_i,\ i = ,..., k$$

where, μ represents the average difference between the non-control treatment arms and the control treatment arms. μ represents the average difference between the non-control treatment arms and the control treatment arms that is not explained by differences in the surrogate outcome.

Assumptions

Unordered treatment arms

X_{ij} and Y_{ij} are approximately normally distributed; e_{ij} and δ_{j} are normally distributed with means 0 and variances σ_{j} and v_{ij}^2, respectively; the variances σ_{j} and v_{ij}^2 are assumed to be known and represent the imprecision in the estimated treatment-arm-level effects; g_{ij} is normally distributed with mean 0 and variance $\gamma^{}$; the random effect g_{ij} is independent of e_{ij} and δ_{j}; $e_{ij}^{(D)}$, $\delta_{j}^{D)}$ and $g_{ij}^{(D)}$ are normally distributed with means 0 and variances $\sigma_{j} + \tau$, $v_{ij}^2 + {}_{i1}^2$, and $2\gamma^{}$; the covariance of $e_{ij_1}^{(D)}$ and $e_{ij_2}^{(D)}$ is $\sigma_{}$; the covariance of $\delta_{j_1}^{D)}$ and $\delta_{j_2}^{D)}$ is v_{i1}^2; the covariance of $g_{ij_1}^{(D)}$ and $g_{ij_2}^{(D)}$ is $\gamma^{}$; the covariance of $e_{ij}^{(D)}$ and $\delta_{j}^{D)}$ is $\rho\,\sigma_{j} v_{ij} + \tau\, v_{i1})$; the covariance of $e_{ij_1}^{(D)}$ and $\delta_{j_2}^{D)}$ is $\rho\sigma\, v_{i1}$; $m_{ij}^{(D)}$ is normally distributed with mean 0 and variance 2λ; the covariance of $m_{ij_1}^{(D)}$ and $m_{ij_2}^{(D)}$ is

λ.

Of note, in their applications of model [20] to trials in oncology and HIV, Korn et al [32] assume that $\rho =)$. The authors conducted various simulations and illustrated that in the absence of a large number of trials, or large differences in sample sizes across trials, the assumption that $\rho =)$ results in a more favourable mean squared error (MSE) than estimation of ρ from trial-level data. The authors also conclude that there is little improvement in the performance of the estimator $\hat{\beta}$ when the value of ρ is known through the use of individual patient data. The authors caution that these results were relative to the characteristics of the simulated dataset; however, their dataset is already representative of a 'typical' large number of trials that may not be available in practice.

Ordered treatment arms

The model assumptions are as before for the case of unordered treatment arms and in their application, Korn et al [32] assume that $\rho =)$. In the ordered treatment arms model,

the variance of $m_{ij}^{(D)}$ is $\lambda_c +$ ≀ where, λ and λ represent the variability of the control and non-control arms, respectively. The covariance of $m_{ij_1}^{(D)}$ and $m_{ij_2}^{(D)}$ is λ . The covariance structure for the remaining parameters is as described for model [20].

β as a measure of association- unordered treatment arms

Unordered treatment arms

The β coefficient in model [20] quantifies the linear relationship between the treatment differences on the surrogate outcome variable, which are considered random rather than fixed, and the treatment differences on the clinical outcome variable. The value of β represents the average change in the treatment difference for the clinical summary outcome for a one unit change in the treatment difference for the surrogate summary outcome.

Ordered treatment arms

The β coefficient in model [21] also quantifies the linear relationship between the treatment differences on the surrogate outcome and the treatment differences on the clinical outcome. Unlike model [20], in model [21] the surrogate outcome difference associated with the β coefficient consists of both the random trial-specific treatment difference as well as the average difference between the non-control treatment arms and the control treatment arms. As before, the value of β represents the average change in the treatment difference for the clinical summary outcome for a one unit change in the treatment difference for the surrogate summary outcome.

Δ as a measure of association

Unordered treatment arms

The parameter Δ quantifies "the expected treatment-arm difference in Y associated with an observed treatment-arm difference ξ in X in a $(k + 1)$st trial". Δ applies to the situation in which the surrogate outcome would be used in a clinical trial that did not provide information about the clinical outcome. In other words,

$\Delta \equiv E[\beta m_{k+1,j}^{(D)} \mid X_{k+1,j}^{(D)} = \xi]$.

Due to the assumed normality, the calculation of delta is based on a standard calculation of shrinkage-type estimators:

$$\Delta = \beta (X_{k+1,j} - X_{k+1,1})\left(\frac{2\lambda}{2\lambda + \sigma_{\cdot,j} + \sigma_{\cdot,1}}\right) \qquad [22]$$

where, Δ can be estimated by $\hat{\Delta}$, which is determined by substituting $\hat{\beta}$ for β and $\hat{\lambda}$ for λ. In the interpretation of Δ, its value represents the expected *Y* difference for a trial with a very large sample size, even though this value may be estimated using $X_{k+1,j}^{(D)}$ from a trial with a small sample size. Korn et al [32] note that Δ is not defined as $E[Y_{k+1,j}^{(D)} \mid X_{k+1,j}^{(D)} = \xi]$. This aforementioned quantity represents the expected treatment-arm difference in *Y* for the trial being analyzed, which may have a small sample size. If $\rho = 1$ then Δ and $E[Y_{k+1,j}^{(D)} \mid X_{k+1,j}^{(D)} = \xi]$ are equal. The authors indicate that if the $(k+1)$st trial is very large, then $\sigma_{\cdot,j}$ and $\sigma_{\cdot,1}$ will be close to zero. As such, there will be no shrinkage and $\Delta = \beta\xi$.

Ordered treatment arms

In the setting of ordered treatment arm trials, Δ can also be calculated to predict the expected treatment arm difference in the clinical outcome, given an observed difference in the surrogate outcome. For a $(k+1)$st trial, Δ is determined as follows:

$$\begin{aligned} \Delta &= E[\mu_Y + \beta(u_X + m_{k+1,j}^{(D)}) \mid X_{k+1,j}^{(D)} = \xi] \\ &= \mu_{\cdot} + \beta(X_{k+1,j} - X_{k+1,1})\left(\frac{\lambda_{\cdot} + \lambda}{\lambda_{\cdot} + \lambda + \sigma_{\cdot,j} + \sigma_{\cdot,1}}\right) \\ &\quad + \beta\mu_{\cdot}\left(1 - \frac{\lambda_{\cdot} + \lambda}{\lambda_{\cdot} + \lambda + \sigma_{\cdot,j} + \sigma_{\cdot,1}}\right) \end{aligned} \qquad [23]$$

3.5 Discussion

The review of statistical methods identified 5 procedures that may be used to evaluate the strength of a surrogate outcome from summary statistics for the surrogate and clinical outcome available in published clinical trials. From these procedures, one was a graphical technique. The remaining four procedures were quantitative with measures of association to summarize the relationship between the surrogate and clinical outcome. The five procedures and associated measures of association (where applicable) are listed below:

- **Method 1:** Significance testing for study-wise agreement
  Measure of association: study-wise percent concordance
- **Method 2:** Rank correlation for the correlation coefficient.
  Measure of association: Spearman's rank correlation coefficient, Spearman's ρ.
- **Method 3:** Graphical analysis
  Measure of association: not applicable
- **Method 4:** Regression analysis
  Measure of association: Weighted or unweighted R^2; weighted or unweighted STE.
- **Method 5:** Mixed models
  Measure of association: β for unordered or ordered treatment arm models; Δ for unordered or ordered treatment arm models.

Many applications for the regression analysis have utilized a weighted approach because the variance in the point estimates for the clinical outcome varies across studies. Korn et al [32] have described procedures for both unordered and ordered treatment arm models. The authors recommend the former approach when there is no natural ordering between the intervention and control arm in the set of studies used for the analysis. The ordered treatment arm model is recommended for the case in which a control group (such as placebo or standard therapy) can be identified in the studies used for the analysis.

The identified methods vary in their complexity. The study-wise agreement, rank correlation and graphical analysis are easy to implement. The regression analysis method has a higher complexity in comparison to the previous three methods, and the largest complexity is associated with the Korn et al models. This model may be difficult to

implement by users of surrogate outcomes because of the advanced statistical knowledge required for its implementation. The Korn et al models also require the estimation of a large number of parameters and is likely to require the largest amount of data. Nontheless, all of the identified methods require a sufficient amount of data to produce meaningful results. The rank correlation and graphical analysis are associated with minimal assumptions. For this reason, these methods may be applied to a larger range of data than the other methods.

In evaluating surrogate outcomes in a meta-analysis, it is useful to apply all of the identified methods to determine whether results are consistent across all procedures. Consistency across varying procedures has the advantage of adding credibility to observed findings. This approach is consistent with a recommendation by Green et al[29].

4 SURROGATE OUTCOME EVALUATION FOR CAROTID INTIMA-MEDIA THICKNESS

4.1 Overview

The purpose of this chapter was to evaluate, in a meta-analytic framework, the validity of CIMT as a surrogate outcome for clinical cardiovascular events. First, a systematic review was performed to identify studies that have evaluated the effect of treatment on both CIMT and clinical cardiovascular events. The statistical procedures identified in Chapter 3 were then applied to data obtained from the randomized studies in order to evaluate the strength of CIMT as a surrogate outcome for various cardiovascular events.

4.2 Objective

The objective of the review was:

- To evaluate whether a treatment related change in carotid intima-media thickness is a valid surrogate outcome measure for treatment-related change in the occurrence of clinical cardiovascular events.

4.3 Methods

4.3.1 Criteria for selecting studies for the review

4.3.1.1 Types of studies

The intent is to review CIMT used in any RCT evaluating an intervention for atherosclerosis. More specifically, the review included randomized trials evaluating the effect of an intervention for the treatment or prevention of atherosclerosis. Studies were eligible if they consisted of a minimum of 2 years of follow-up, reported post-treatment CIMT measurements at a 12 month follow-up period and reported clinical cardiovascular events at the trial end. Minimum study duration of 2 years was chosen because restricting eligibility to studies with longer follow-up periods may yield an insufficient number of studies to perform a meaningful analysis. Studies with shorter follow-up durations were not chosen because cardiovascular events develop over a long period of time. It was hypothesized that 12 month follow-up values of CIMT are likely to indicate the rate with

which atherosclerosis will eventually, over time, progress to cardiovascular events are not likely to indicate the risk of cardiovascular events in the immediate short-term. Although it was desirable to collect information on cardiovascular events that occurred at a specific time point during the follow-up period, this was not done because the number of studies for any given follow-up period may be too few to allow for a meaningful analysis. Studies were required to report CIMT at 12 months of follow-up because 12 month values are commonly reported in imaging trials and it was uncertain whether CIMT would have been reported at an earlier time point in a group of studies large enough for the analysis.

4.3.1.2 Types of participants

Men and women aged 18 years or older receiving treatment for the prevention or regression of atherosclerosis. The review was not restricted to subjects of any particular baseline risk level or medical condition. This decision was made a priori because it was of interest to determine the validity of CIMT as a surrogate outcome for all populations in which this outcome may be used to evaluate the effectiveness of treatment for the prevention or regression of atherosclerosis.

4.3.1.3 Types of interventions

The interventions considered included any pharmacologic or non-pharmacologic intervention versus another intervention or placebo. The review was not restricted to any specific drug class, class of vitamin supplements or other intervention (such as exercise). This decision was made because it was of interest to evaluate the use of CIMT as a surrogate outcome across all interventions for which this outcome may be used to evaluate the effectiveness of treatment for the prevention or regression of atherosclerosis.

4.3.1.4 Types of outcome measures

Primary outcome measures

The primary surrogate outcome was the mean in each treatment group, for ultrasonographic measurement of CIMT at 12 months of follow-up. Based on the Mannheim Carotid Intima-Media Thickness Consensus [53], it was preferred that the CIMT

values included in the review were obtained from measurements obtained in the far wall of the right plus left common carotid artery (CCA). For studies in which measurements from the far wall of the right plus left CCA were not available, the measurements that were included in the review were those that were available or were most likely to give results similar to those obtained from the far wall of the right plus left CCA. It was decided that the degree of similarity to the preferred measurement would be determined through discussion with content experts.

The primary clinical outcome was all-cause mortality at the end of the follow-up period.

Secondary outcome measures

Secondary surrogate outcomes were not included in the review. Secondary clinical outcomes were (i) fatal or non-fatal myocardial infarction, (ii) stroke and (iii) cardiovascular mortality.

4.3.2 Search methods for identification of studies

4.3.2.1 Electronic searches

The following electronic databases were searched for prospective controlled clinical trials:

- MEDLINE, through Ovid (1950 to March Week 5 2010)
- EMBASE Classic + EMBASE, through Ovid (1947 to 2010 April 13)
- The Cochrane Central Register of Controlled Trials (search conducted on April 17, 2010)

The search strategy was initially developed in MEDLINE. In the first step of the construction of the search strategy for use in all three databases, several reviews articles that were relevant to the objective of this systematic review were obtained. The bibliographies of these reviews were searched to find reports of randomized trials evaluating the effect of an intervention on CIMT. The titles, abstracts and controlled vocabulary (as appearing in MEDLINE) were then scanned to generate a list of terms that can be used to describe the outcome carotid intima-media thickness. The list of terms

consisted of both text words and controlled vocabulary. In the second step, the list of terms for CIMT was combined with the sensitivity-maximizing Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE [54]. This search strategy was then translated for use in CENTRAL and EMBASE by modifying the syntax and controlled vocabulary as needed. Of note, for use in CENTRAL, the search strategy required syntax changes and the removal of the search filter for identifying randomized trials. For use in EMBASE, the search strategy required modifications to the controlled vocabulary in the list of terms for CIMT. This modified list of terms for CIMT was then combined with two different search filters for identifying reports of randomized trials. One of these search filters has been described in the Cochrane Handbook [54]. The second search filter has been developed by Wong et al and has also been recommended in the Cochrane Handbook [54, 55].

The general approach to incorporating the population, intervention, comparison and outcome (PICO) criteria in the electronic search strategy were not used since a wide search to capture when CIMT was used was the goal. In particular, the search strategy for a review answering a clinical question avoids the inclusion of outcomes because they may not be described in the indexing terms assigned to a relevant study and may not appear in the title and abstract of the report. Because it is accepted that there are situations in which it is legitimate for reviews to limit eligibility criteria to specific outcomes, and given that the review answered a methodological question, a decision was made that it would be reasonable to construct an electronic search strategy that includes a very thorough list of terms that may be used to describe CIMT along with a RCT search filter, where needed. Terms for the clinical outcomes were not included in the search strategy because it was considered that the inclusion of these terms would result in the omission of relevant studies. The full electronic database search strategies for MEDLINE, EMBASE and CENTRAL are available in Appendix B.

4.3.2.2 Searching other resources

The following additional resources were searched:

- Reference lists of eligible studies and 11 relevant reviews were searched to identify additional eligible studies.
- The trials register clinicalTrials.gov was searched to identify ongoing studies.

4.3.3 Data collection and analysis

4.3.3.1 Selection of studies

The title and abstracts of identified reports from the electronic databases were merged and stored in Reference Manager (version 12). Duplicate reports were deleted and the titles and abstracts of the identified reports were screened. Titles and abstracts that suggested fulfillment of the eligibility criteria appeared to be randomized controlled trials and had either reported post-treatment measurements of CIMT or reported that CIMT was measured after treatment. Titles and abstracts were not required to report information about clinical outcomes in order to be selected for full-text review because of the possibility that the clinical outcome data may not be well-described in the titles and abstracts. Similarly, the titles and abstracts were not required to provide information on whether CIMT was measured at 12 months of follow-up in order to be selected for full-text review. Reports that clearly did not meet the inclusion criteria were excluded. The full text of the remaining reports was then obtained and multiple-reports of the same study were linked together. The full-text of these reports was then evaluated to determine eligibility for inclusion in the review. Titles and abstracts were examined by one author (MK) and the full-text articles were examined by two authors (MK, ZK). It was decided that any disagreements between authors about whether a study should be included in the review would be resolved through discussion or consultation with the thesis supervisor. Similarly, any difficulties in interpreting the study information that was required to determine eligibility for inclusion in the review was resolved through discussion (MK, ZK) or consultation with the thesis supervisor. The review authors who examined the reports for inclusion in the review were not masked to information about the articles (e.g. journal name, authors' names and institutions). The eligibility criteria used for examination of the titles and abstracts, as well as the full-text, was pilot tested on approximately 50 articles. These 50 articles included those that were definitely eligible, definitely ineligible and unclear.

4.3.3.2 Data collection

All data extraction, numerical calculations and measurements obtained from graphs were performed by one author (MK) and confirmed by a second author (ZK). A decision was made a priori that the following information would be collected for each of the studies:

Methods:

- Study design; study duration; description of sequence generation, allocation sequence concealment and blinding.

Participants:

- total number; geographic location (such as country and city); setting (such as community practice or hospital); start and end date of study; disease (and duration) or risk factor for which intervention is being administered; age (mean and SD) in each group; percentage of female in each group; percentage of patients with coronary heart disease at baseline;

Interventions:

- Total number of intervention groups; specific interventions; drug class for pharmacological interventions; details of pharmacololgical and non-pharmacological interventions (such as the dosage of drugs, time of drug administration, and components of dietary counselling and methods used to monitor food intake).

Outcomes:

- For CIMT the following information was obtained: the location(s) in the carotid artery where CIMT was measured (such as far or near wall); unit of measurement for CIMT; description for how CIMT was measured (i.e. as the mean of the mean or mean of the maximum); follow-up time points at which CIMT was measured; follow-up time points for which CIMT was reported; citations provided by authors that may provide additional details about how the various CIMT measurements were obtained.
- For each of the clinical outcomes, all-cause mortality, myocardial infarction, stroke and cardiovascular mortality, the following information was obtained: the method of ascertainment; follow-up time points at which the clinical outcome data was collected; follow-up time points for which the clinical outcomes were reported.

Results:

- Number of subjects allocated to each group.
- For CIMT, at each of the follow-up measurements, the following information was collected: the group mean; the within group mean change from baseline; the within-group mean rate of change; the between group difference in means; the between group difference in the change from baseline; the between group difference in the rate of change. As well, for each of these results, the standard deviation, the p-value and 95% CI were collected.
- For each of the clinical outcomes at trial end and any other time points at which they were ascertained, the following information was collected: the number of subjects with events; the between group difference in the number of subjects with events; the standard deviation of the between group difference in the number of subjects with events; the p-value and 95% CI for the significance of the between-group difference in the number of subjects with events; the type of relative effect measure (such as the hazard ratio or risk ratio) used for any between group differences in the number of subjects with events; the value of the relative effect measure; the p-value and 95% CI for the relative effect measure.

Miscellaneous:

- Source of funding; conclusions of the study; important comments made by the study authors; references to studies that do not fit the eligibility criteria but are relevant to the topic area; miscellaneous comments by the review authors.

4.3.3.3 Assessment of risk of bias in included studies

The risk of bias for each study was evaluated according to the domain-based evaluation tool outlined in the *Cochrane Handbook for Systematic Reviews of Interventions*[56]. In this evaluation tool, the risk of bias in each of six domains was determined in a two step process. The domains included sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and other biases. The first step of the evaluation consisted of providing a description about what was reported to have occurred in the study for each of the domains. In the second step, a judgement was made to determine whether there was adequate protection from bias in the specific domain

being considered. The judgement was made according to criteria described in the *Cochrane Handbook for Systematic Reviews of Interventions* [56]. A judgement of 'yes' indicated low risk of bias, a judgement of 'no' indicated high risk of bias, and a judgement of 'unclear' indicated uncertain risk of bias.

The risk of bias assessment for the included trials was summarized in two ways. A 'risk of bias graph' was created to show the proportion of studies that were judged as 'yes', 'no' or 'unclear', for each entry in the tool. A 'risk of bias summary' figure was also created. The purpose of the figure was to depict the judgement for each study through a cross-tabulation of study by entry. The judgement of 'yes' was represented by a + sign, 'no' was represented by a - sign , and 'unclear' was represented as ? sign.

4.3.3.4 Measures of treatment effect

The between group mean difference in CIMT at 12 months was used to evaluate the treatment effect on the surrogate outcome. The log odds ratio was calculated to measure the between group difference in the risk of each of the clinical outcomes. The log odds ratio was calculated for events reported at the end of the follow-up period for each of the included studies.

4.3.3.5 Assessment of heterogeneity

The amount of variation between studies in the estimates of the mean difference in CIMT was calculated from τ^2 in the random effects model for ordered treatment arms described by Korn et al [32]. The amount of variation between studies in the estimates of the log odds ratio was determined in the same way.

4.3.3.6 Data synthesis

The statistical procedures identified in the review of statistical methods in Chapter 3 were used to evaluate the strength of the relationship between CIMT and each of the clinical outcomes. The methods included (i) study wise agreement analysis; (ii) graphical analysis; (iii) R^2 from a weighted regression analysis; (iv) STE; and (v) β from the Korn et al fixed effects model for ordered treatment arms [32].

4.3.3.7 Studies with multiple treatment groups

For three arm studies in which the same group was used as the comparator for two pair-wise comparisons, the sample size of that group was split in half when calculating the mean difference. For calculation of the odds ratios, both the group sample size and number of events was split in half[56].

4.3.3.8 Dealing with missing data

For studies that measured CIMT at 12 months follow-up but only reported the rate of change in mm per year, the 12 month value was determined using the rate and the baseline value. For studies in which standard deviations for the follow-up CIMT were missing for either of the treatment arms, the baseline standard deviations were used.

4.3.3.9 Sensitivity analysis

The regression diagnostics consisting of residual analysis through graphical techniques were checked for weighted linear regression analysis. The assumptions of normality, homoscedasticity, independence and linearity were assessed. The normality assumption was evaluated through a kernel density plot of the distribution of residuals and a normal quantile-quantile plot. Homoscedasticity and independence were evaluated through a plot of the residuals versus fitted values. The linearity assumption was evaluated through a residual versus predictor plot and a LOWESS smooth was added to the plot .Visual analysis was used to determine whether the LOEWSS smooth showed evidence of non-linearity in the average value of the residuals across the values for the 12 month mean difference in CIMT. For models in which there was evidence of departure from linearity, power transformations were applied to the CIMT values. The first choice for the power transformation was based on recommendations from standard statistical texts after assessing the observed curvature in the residual versus predictor plot. Additional transformations were also applied because it was considered that an empirically chosen transformation may offer a greater improvement in linearity then the one originally indicated by the plot. For the diagnostics of the Korn et al fixed effects model, it was assumed that conclusions about the normality and linearity assumptions obtained from

diagnostics performed for the weighted regression analysis could also be applied to the analysis based on the Korn et al fixed effects analysis.

A sensitivity analysis was planned to determine whether various power transformations would remedy violations of (or improve) the linearity assumption for the weighted regression and Korn et al fixed effects model. Of note, the sensitivity analysis was not planned a priori before performing the original analysis.

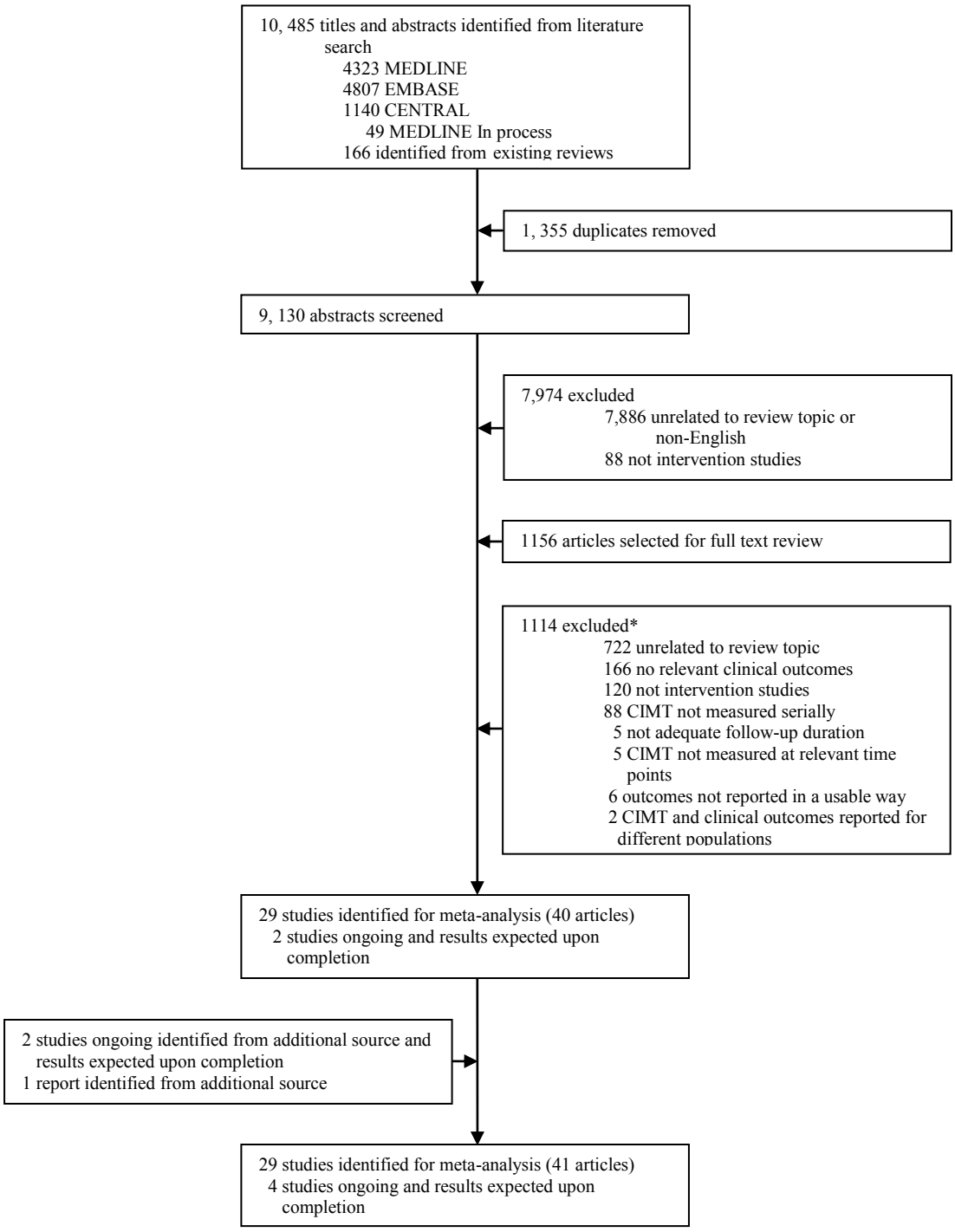

Figure 4.1: Flow diagram for selection of studies of interventions for atherosclerosis. *14 of these excluded articles are well-known and considered relevant to this topic. The references for these studies have been identified in the text and reasons for exclusion are provided.

4.4 Results

4.4.1 Results of the search

Figure 4.1 illustrates how the studies were selected. The search strategy identified a total of 10, 485 articles. From these, 29 studies (corresponding to 41 articles) were eligible for inclusion in the review. One additional report was identified from reference list searching and contained supplementary results for a previously identified study. The search strategy also retrieved reports of 2 studies that were ongoing and for which results still need to be generated upon trial completion. A further 2 were identified from the search of the clinical trials register. Of note, 3 additional reports [57-60] were used to extract some relevant information for 3 included studies; however, these 3 reports did not comprise the 41 eligible articles for the included studies because they did not meet the criteria for inclusion in the review. One of the reports was a trial protocol that did not contain any baseline or other results [59]. The other two reports were publications of the results of the parent trials for ancillary studies that were included in the review [60, 61].

Table 4.1 Summary characteristics of included studies

Reference	Dur-ation (yrs)	Population	Sample size	% CVD	Mean Age (years)	Interventions	All-cause Mortality Measured	Myocardial infarction measured	Stroke Measured	Cardiovascular mortality measured
Hodis (TART)	2	Type 2 diabetes	299	<10	52.5	Glitazone; Placebo		✓	✓	
Hodis (EPAT)	2	Healthy, menopausal women	222	0	61.5	Hormone; Placebo	✓	✓	✓	✓
Smilde (ASAP)	2	Familial hypercholesterolemia	330	23.3	48.5	Statin; Statin	✓			✓
Sawayama (FAST)	2	Asymptomatic Hypercholesterolemia	246	14.2	66.1	Statin; antihypertensive; Placebo	✓	✓		✓
Kastelein (ENHANCE)	2	Familial hypercholesterolemia	720	Not provided	45.9	Statin + cholesterol absorption inhibitor; Statin		✓	✓	✓
Terpstra (ELVERA)	2	Mild to moderate hypertension	166	0	67	CCB; ACE inhibitor	✓			
Bots (RADIANCE2)	2	Mixed dyslipidemia	377	Not provided	57.2	Statin + CETP inhibitor; Statin	✓	✓	✓	
Crouse (METEOR)	2	Mild to moderate subclinical atherosclerosis; low Framingham risk score	984	0	57	Statin; Placebo	✓	✓		
Kastelein (RADIANCE1)	2	Familial hypercholesterolemia	904	Not provided	46	Statin + CETP inhibitor; Statin		✓	✓	✓
Nanayakkara (ATIC)	2	Mild to moderate chronic kidney disease	93	0	53	Statin + vitamin E + vitamin B; Placebo	✓	✓		✓
Meuwese (CAPTIVATE)	2	Familial hypercholesterolemia	892	97	55.1	ACAT inhibitor; Placebo		✓	✓	✓
VanVonderen	2	indication to start combination antiretroviral treatment	37	13.5	41.4	NRTI; cART	✓	✓		✓
Hodis (VEAPS)	3	Elevated cholesterol; no cvd	353	0	56.2	Vitamin E; Placebo	✓	✓	✓	✓
Howard (SANDS)	3	Type 2 diabetes	548	0	56	Aggressive treatment; Standard treatment	✓	✓	✓	✓
Lonn (STARR)	3	Impaired glucose tolerance/ impaired fasting glucose	1425	0	54.4 54.3	Ace inhibitor; Placebo; Glitazone; Placebo	✓	✓	✓	✓
Borhani	3	Hypertension	883	1.5	58.5	CCB;	✓	✓	✓	✓

Reference	Dur-ation (yrs)	Population	Sample size	% CVD	Mean Age (years)	Interventions	All-cause Mortality Measured	Myocardial infarction measured	Stroke Measured	Cardiovascular mortality measured
(MIDAS)						Thiazide diuretic				
Wiklund (ELVA)	3	Hypercholesterolemia	129	4	59.8	Beta blocker; Placebo		✓		
Hedblad (BCAPS)	3	Carotid plaque; no cvd	793	3.75	61.5	Beta blocker; Placebo	✓	✓	✓	
Bots (OPAL)	3	Healthy, postmenopausal women	866	0	58.8	Hormone; Hormone; Placebo	✓			
Hodis (BVAIT)	3	Elevated cholesterol;	506	0	61.4	Vitamin B; Placebo	✓			
Byington (PLAC-II)	3	Postmenopausal; elevated cholesterol (men and women); cvd	151	100	62.6	Statin; Placebo	✓	✓		
Salonen (KAPS)	3	Hypercholesterolemia	447	7.6	57.4	Statin; Placebo	✓	✓	✓	✓
Zanchetti (VHAS)	4	Hypertension	456	4.95	54.1	Verapamil; Chlorthalidone	✓	✓	✓	✓
Magliano (MAVET)	4	Smoking	409	≥6.6	63.5	Vitamin E; Placebo	✓			✓
Zanchetti (ELSA)	4	Hypertension	2334	12.8	56	Lacidipine; atenolol	✓	✓	✓	✓
Elkeles (SENDCAP)	5	Type 2 diabetes	164	0	50.9	Fibrate; Placebo		✓		
Mazzone (CHICAGO)	6	Type 2 diabetes	462	11	59.4	Glitazone; Sulfonylurea	✓	✓	✓	
Salonen (ASAP)	6	Elevated cholesterol	520	13.95	45-69	Vitamin E + vitamin C; Placebo	✓			
Hodis (CLAS)	10	CAD	188	100	54.2	Colestipol + vitamin B + diet therapy; Diet therapy		✓		✓

4.4.1.1 Description of included studies

A summary of the main characteristics of the included studies is presented in Table 4.1. A table of detailed characteristics of the included studies can be found in Appendix C (Table C.1). Of the 29 included studies, 10 had multiple reports involving 12 articles. The multiple reports contained baseline results or results for additional outcomes. For the purposes of this review, the primary reports for studies with multiple publications are as follows: for the PLAC-II trial [62, 63], Byington et al [62] is the primary study; for the ASAP study [64, 65], Smilde et al [65] is the primary study; for another ASAP study [66, 67], Salonen et al [67] is the primary study; for the ELSA trial [68-71], Zanchetti et al [70] is the primary study; for the METEOR trial [72, 73], Crouse et al [73] is the primary study; for the RADIANCE 1 trial [74, 75], Kastelein et al [74] is the primary trial; for the SANDS trial [76, 77], Howard et al [76] is the primary study; for the MIDAS trial [78, 79], Borhani et al [79] is the primary trial; for the CLAS trial [80, 81], Hodis et al [81] is the primary study and for the OPAL trial [82, 83], Bots et al [83] is the primary study.

Design

All studies were randomized controlled trials. The FAST [84] trial consisted of two treatment groups (probucol and pravastatin) and one placebo group. The OPAL trial [83] consisted of two treatment groups (tibolone and conjugated equine oestrogens plus medroxyprogesterone) and one placebo group. The STARR [85] trial followed a 2X2 factorial design in which each of two treatments (ramipril and rosiglitazone) were compared to placebo. The remaining studies were two arm trials.

Sample sizes

The sample sizes ranged from 37 in the study by van Vonderen et al [86] to 2334 in the ELSA trial [70].

Setting

At least some details about the setting were provided for 21 studies. The setting for ELVA [87] and CAPTIVATE [88] were lipid clinics and ELVA [87] was both community and

hospital based. SENDCAP [89] was conducted at diabetes clinics. ATIC [90] was conducted at an outpatient clinic. ENHANCE [91] was conducted in ambulatory centers. METEOR [73] was conducted in primary care centers. The setting for RADIANCE 1 [74], RADIANCE 2 [92], VHAS [93], and CLAS [81] were imaging centers and the CLAS [81] centre was university-based. EPAT [94] and MIDAS [79] were conducted in university based clinics, but no additional details were provided. Similarly, SANDS [76], ELSA [70], and CHICAGO [95] were conducted in a clinical setting, and additional details were not provided. Van Vonderen [86] was conducted in a hospital setting. BCAPS [96], MAVET [97], and ASAP [67] took place in a community setting. KAPS [98] also took place in a community setting, including some rural areas. ELVERA [99] was based in a rural setting.

From the studies, 18 were conducted as multicentre trials. The studies conducted in the United States of America included TART [100], SANDS [76] (Oklahoma, Arizona and South Dakota), MIDAS [79], CHICAGO (Chicago) [95], and CLAS [81]. The ASAP study by Smilde et al [65] and ATIC [90] were conducted in multiple centers in the Netherlands. The study by vanVonderen [86] was also primarily conducted in the Netherlands, although some patients may have been enrolled at sites in Finland, Spain and the United Kingdom. SENDCAP [89] was conducted at multiple sites in England. VHAS [93] was conducted in various centers in Italy. ELSA [70] was conducted in various countries in Europe (France, Germany, Italy, Greece, Spain, Sweden, and United Kingdom). RADIANCE 2 [92] was conducted in North America and Europe. METEOR [73] and OPAL [83] were conducted in the US and Europe. ENHANCE [91] was conducted in North America (Canada, US), Europe (Spain, Denmark, Norway, Sweden, and the Netherlands) and South Africa. RADIANCE 1 [74] was also conducted in these three regions, but information about the specific participating countries was not provided. CAPTIVATE [88] was conducted in North America (Canada, US), Europe, South Africa and Israel. The STARR trial [85] was conducted in North America (Canada, US, Bermuda), South America (Argentina, Brazil), Europe, India and Australia. The remaining studies which were not reported as multicentre trials were conducted in the United States, Japan, the Netherlands, Sweden and Australia.

Participants

A total of 3 diseases were represented in this review. TART [101], SANDS [76], SENDCAP [89] and CHICAGO [95] were conducted in subjects with Type 2 diabetes. The ATIC [90] study included subjects with mild to moderate chronic kidney disease. PLAC-II [62] and CLAS [81] included subjects with coronary artery disease. The remaining studies were conducted in populations with risk factors for coronary artery disease and who may benefit from treatment that aims to prevent or reduce atherosclerosis. ELVERA [99], MIDAS [79], VHAS [93], and ELSA [70] included subjects with hypertension. ASAP [65], ENHANCE [91], RADIANCE 1 [74] and CAPTIVATE [88] were conducted in subjects with familial hypercholesterolemia. FAST [84], ELVA [87], and KAPS [98] were conducted in subjects with hypercholesterolemia. VEAPS [101], BVAIT [102] and ASAP [65] subjects had elevated cholesterol. The population in EPAT [94] and OPAL [83] was healthy, post menopausal women. Risk factors that were specific to each of the remaining studies are listed in Table 4.1. The average age of the participants across the trials was 56.2 years and ranged from 41.4 years in the study by van Vonderen et al [86] to 67.0 years in the ELVERA [99] trial. The ASAP [67] study by Salonen et al was not used in determining these results because the age was reported as a range. From the studies that provided information about baseline CVD, no subjects had CVD at baseline in 10 studies. In the PLAC –II [62] and CLAS [81] trial, 100% of the subjects had CVD at baseline. In the CAPTIVATE [88] trial, the proportion was 97% . In the remaining studies, the proportion of subjects with CVD at baseline varied from 3.75% in the BCAPS trial to 23.3% in the ASAP trial. The study by van Vonderen et al [86], as well as the CLAS [81] and KAPS [98] trials included only males. The EPAT [94] and OPAL [82] trials included only females. For the remaining studies, the proportion of females varied from 13% in the PLAC-II [62] study to 73.5% in the FAST [84] study.

Interventions

The comparison of statin versus placebo was evaluated in 4 studies, FAST [84], METEOR [73], PLAC II [62], and KAPS [98] (statin: 3 pravastatin , rosuvastatin). An additional 5 drug classes were compared in sets of two studies. EPAT [94] and OPAL [82] compared hormone

therapy vs placebo (hormone therapy: micronized 17b-estradiol, tibolone, conjugated equine oestrogens plus medroyxprogesterone acetate (CEE/MPA)). VEAPS [101] and MAVET [97] compared vitamin E vs. placebo (vitamin E: d-α tocopherol, DL- α tocopherol). MIDAS [78] and VHAS [93] compared calcium channel blocker vs. thiazide diuretic (calcium channel blocker: isradipine, verapamil; thiaizide diuretic: hydrochlorothiazide, chlorthalidone). TART [100] and STARR [85] compared glitazone vs placebo (glitazone: troglitazone, rosiglitazone). RADIANCE 1 and 2 [91, 92] compared statin + CETP inhibitor vs statin (statin + CETP inhibitor: atorvastatin + torcetrapib; statin: atorvastatin). The remaining comparisons were evaluated in individual studies (Table 4.1; Table C.1 in Appendix C identifies the specific drugs).

Outcomes

Surrogate outcome

Measurements for the far wall of the right plus left CCA were available in 5 studies, FAST [84], SANDS [76], STARR [85], KAPS [98] and ASAP [67] by Salonen. The far wall of the right CCA was evaluated in 7 studies, TART [100], EPAT [94], VEAPS [101], ELVA [87], BCAPS [96], BVAIT [102] and CLAS [81]. In two studies, ELVERA [99] and SENDCAP [89], the far wall of the CCA was measured, but information was not provided for whether the measurements were performed on the right side, left side or bilaterally. Measurements were available for the right plus left CCA in the study by van Vonderen et al [86] and CHICAGO [95], but information was not provided for whether the measurements were taken on the near and/or far walls. The near and far wall of the right plus left CCA was measured in 7 studies, METEOR [73], RADIANCE 1 [74], RADIANCE 2 [92], OPAL [83], ELSA [70], ASAP [65] by Smilde et al, and PLAC-II [62]. The CCA was measured in ATIC [90] and MAVET [97], but additional details were not provided. The far wall of the right plus left CCA, bulb and internal carotid (6 segments) was measured in 3 studies, ENHANCE [91], CAPTIVATE [88] and VHAS [93]. The near and far walls for the right plus left CCA, carotid bulb and internal carotid (12 segments) were measured in the MIDAS [79] study.

Clinical outcomes

Data for all-cause mortality was available from 23 studies. Data for myocardial infarction, stroke and cardiovascular mortality was available from 26, 16 and 18 studies, respectively.

4.4.1.2 Description of excluded studies

Table 4.2 provides information about 14 studies that were excluded during full-text review. These 14 articles are well-known imaging trials that measured CIMT and are of interest to the topic of using CIMT as a surrogate outcome in clinical trials. The references for these studies are identified in Table 4.2 and reasons for exclusion are provided to explain why these studies were not included in the review.

Table 4.2 Description of excluded studies

Reference	Reason for exclusion
ACAPS (Furberg et al[103])	CIMT and the clinical outcomes were measured on different populations
MARS (Hodis et al[104]; Blankenhorn et al[105])	CIMT and the clinical outcomes were measured on different populations
CAIUS (Mercuri et al[106])	No clinical outcomes were reported[1]
SECURE (Lonn et al[61])	Clinical outcomes not reported in a usable way
HERS (Byington et al[107])	Serial measurements of CIMT not reported
ARBITER (Taylor et al[108])	The study did not have an adequate follow-up duration
EDIC (Nathan et al[109])	No clinical outcomes are reported
ASFAST (Zoungas et al[110])	CIMT not reported in a usable way
ARBITER-6 HALTS (Taylor et al[111])	The study did not have an adequate follow-up duration
LIPID (MacMahon et al[112])	Measurements of CIMT were not reported for relevant time points
PART-2 (MacMahon et al[113])	Measurements of CIMT were not reported for relevant time points
PREVENT (Pitt et al[114])	Serial measurements of CIMT not reported
STOP-NIDDM substudy (Hanefeld et al[115])	Serial measurements of CIMT not reported
REGRESS (de Groot et al[116])	Serial measurements of CIMT not reported

[1]One myocardial infarction was reported, but it occurred during the first 4 months of therapy, which was before the follow-up measurements for CIMT.

4.4.1.3 Description on ongoing studies

Table 4.3 contains brief descriptions of ongoing studies which appear to satisfy the inclusion criteria but for which results still need to be generated when the studies are completed. It is expected that the results will be available upon trial completion.

Table 4.3: Description of ongoing studies

Reference	Description of study
De Lorenzo et al[117]. "Prevention of atherosclerosis in patients living with HIV".	Methods: Randomized, double-blind, placebo-controlled Study duration: 2 years Participants: HIV Interventions: Rosuvastatin Notes: CIMT measured at 6 and 12 months. Adverse events, including those that result in death, will be monitored.
de Fronzo et al[118]. "Actos Now for the prevention of diabetes (ACT NOW) study".	Methods: Randomized, double-blind, placebo-controlled Study duration: 4 years Participants: Impaired glucose tolerance Interventions: Pioglitazone or placebo Notes: CIMT measured at 18 months. Cardiovascular events and mortality will be monitored.
"Targeting Peroxisome Proliferator-activated Receptor-gamma in Peritoneal Dialysis Patients- Will it Reduce Inflammation, Atherosclerosis, Calcification and Improve Survival of Peritoneal Dialysis Patients?"[57] ClinicialTrials.gov identifier: NCT00516880	Methods: Randomized, double-blind, placebo-controlled trial Study duration: 2 years Participants: peritoneal dialysis Interventions: Rosiglitazone or placebo Notes: CIMT will be measured at 6 months and 12 months. Cardiovascular events and all-cause mortality will be monitored.
"Cilostazol Versus Aspirin for Primary Prevention of Atherosclerotic Events."[119] ClinicalTrials.gov identifier: NCT00886574	Methods: Randomized, controlled trial Study duration: 4 years Participants: Type 2 diabetes Interventions: Cilostazol or aspirin Notes: CIMT will be measured at 6 months and 12 months. Events of ischemic heart disease and cerebrovascular disease will be monitored

4.4.2 Risk of bias in included studies

Figures 4.2 and 4.3 summarize the results of the risk of bias assessment.

Figure 4.2 Risk of bias summary figure

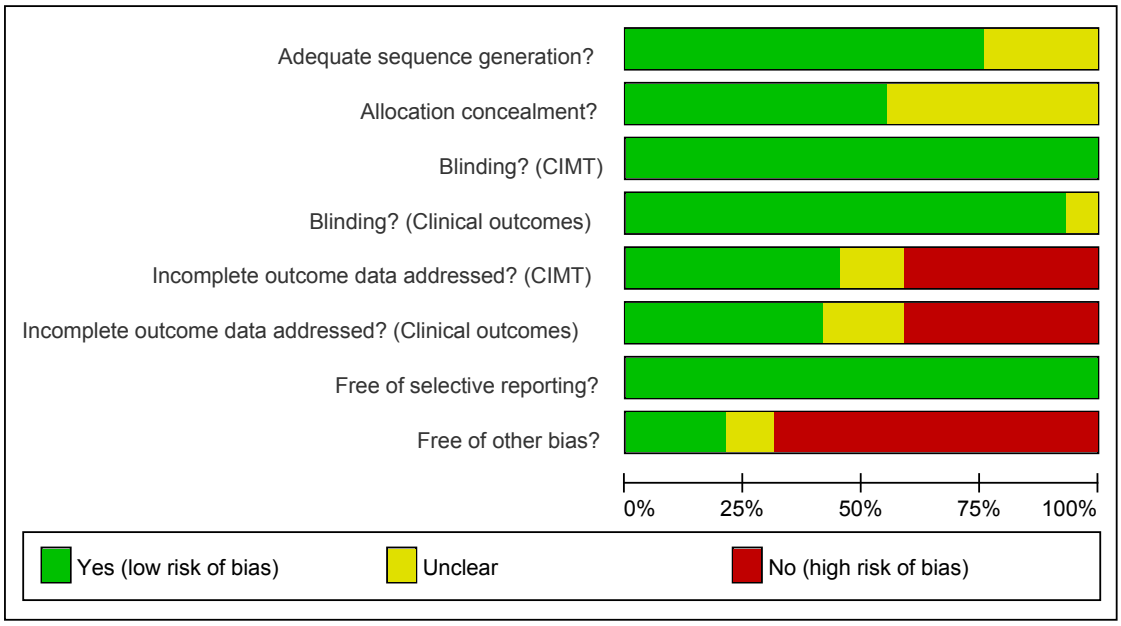

Figure 4.3: Risk of bias graph

4.4.2.1 Sequence generation

In 22 studies (76%), there was a low risk of bias due to inadequate sequence generation. In 6 of these studies, there was insufficient information about the method used for sequence generation, but it was judged to be adequate in other studies by the same authors. These studies included METEOR [73], RADIANCE 1 [74], CAPTIVATE [88], OPAL [82], ASAP [67] by Salonen and VHAS [93]. In another 3 of the studies with low risk of bias, SENDCAP [89], CLAS [79] and KAPS [98], a statistician was responsible for sequence generation and it was assumed that sequence generation was performed properly. In 8 studies, computer generated random numbers were used and in 1 study, STARR [85], a computerized telephone randomization system was used. In 2 studies, FAST [84] and van Vonderen et al [86], the minimization method was used. A random permuted block size of 8 was used in the MAVET [97] study. In the SANDS trial, randomization was performed using the urn method [76].In the other seven studies (24%) insufficient information was provided and the risk of bias was judged to be unclear.

4.2.2.2 Allocation concealment

Risk of bias due to inadequate allocation concealment was judged to be low in 16 (55.17%) of the studies. In 4 of these studies, the methods were not fully described, but the risk of bias was found to be low in other studies by the same authors. These studies

included METEOR [73], RADIANCE 1 [74], CAPTIVATE [88] and OPAL [83]. In the remaining 13 studies, the risk of bias was unclear because there was insufficient information in the reports.

4.4.2.3 Blinding

CIMT

All studies were judged to be at low risk of bias due to inadequate blinding. All of the studies were reported to have blinded outcome assessors. Also, most of the studies provided information which described all study personnel as being blinded or reported that the study was 'double-blind'. Studies in which there was some lack of blinding or inadequate reporting of this information were not judged to introduce bias because CIMT measurements are performed in an objective manner and all outcome assessors were masked.

In 6 studies, patients, study personnel and outcome assessors were blind to treatment assignment. This included TART [101], EPAT [94], VEAPS [101], ELSA [70], BVAIT [102] and MAVET [97]. Another 10 studies were reported as 'double blind' and also reported that the outcome assessors were blinded. Of the studies, 8 were reported as double blind and no additional information was provided. In SANDS [76], CLAS [81] and VHAS [93], outcome assessors were blinded, but some of the other study members were not. In the SANDS [76] trial, the physician was reported as being unmasked, in the CLAS [81] trial, the patients were unmasked and the VHAS [93] trial was open label for part of the study duration. In the FAST [84] trial and the study by van Vonderen et al [86], outcome assessors were blinded and additional information was not provided.

Clinical outcomes

The risk of bias was unclear in the van Vonderen et al [86] study and the FAST [84] trial because no information was provided about blinding. For the remaining studies, there were no concerns about blinding with respect to the clinical outcomes and the studies were judged to be at low risk of bias. This decision was made because for all of these studies, either blinding was described for all study members, the report was stated as

being ‘double blind’ or the outcome assessors were blinded. In similarity to CIMT, because of the objective nature of the clinical outcomes, lack of blinding for some study members or incomplete reporting of blinding was not judged to introduce bias into the studies.

For 5 studies, the reports mentioned that the study was ‘double blind’ and also reported that the outcome assessors were masked to treatment. This included STARR [85], MIDAS [78], BCAPS [96], ELSA [70] and CHICAGO [95]. Another 5 studies reported that the subjects, physicians and clinical staff were blinded. Included in these studies were TART [101], EPAT [94], VEAPS [101], BVAIT [102], and MAVET [97]. For 14 studies, the report stated that the trial was ‘double blind’ and additional information was not provided. In two studies, VHAS [93] and SANDS [76], the outcome assessors were blinded but some other study participants were not. The VHAS [93] study was open label for part of the study duration, and in the SANDS [76] trial the physicians were not blinded. In the CLAS [81] trial, the subjects were not blinded. The report indicated that myocardial infarction was evaluated in a blinded way and it was assumed that the remaining clinical outcomes would also have been obtained in this way.

4.4.2.4 Incomplete outcome data

Studies were judged to be at high risk of bias if overall, <80% of the subjects in the trial had complete outcome data. If completeness of outcome data was provided for the specific groups only, then studies were judged to be at high risk of bias if <80% of the subjects in the experimental group had complete outcome data. Outcome data was considered missing if subjects were excluded from the analysis or if subjects dropped out during the study. For this review, an analysis was defined as intention to treat (ITT) if it evaluated all randomized subjects according to the group to which they were randomized.

CIMT

From the included studies, 7 reported the completeness of the CIMT data for measurements obtained at the 12 month follow-up period. These studies included ASAP [65] by Smilde et al, ELVERA [99], ATIC [90], CAPTIVATE [88], OPAL [83], MAVET [97] and

CHICAGO [95]. In 4 studies, there was no information about the completeness of data and included RADIANCE 2 [92], MIDAS [79], BCAPS [96], and PLAC-II [62]. For these studies (14% of total), the amount of bias was judged to be unclear (Figure 4.2; Figure 4.3). Because most of the remaining studies reported the number of subjects who completed the trial at the end of the follow-up period, this information was used to assess the completeness of data if information for the 12 month follow-up time was unavailable. For the 17 studies that reported the completeness of data at trial end, 15 provided information for each group of randomized subjects and 2 studies, VEAPS [101] and STARR [85], provided data for the total number of subjects randomized. In 1 of the included studies, VHAS [93], all details about the completeness of data were not provided, but less than 80% of subjects in each group were included in the analysis. For this reason, it was judged to be at high risk of bias. Overall, 13 (45%) of studies were judged to be at low risk of bias due to incomplete outcome data (Figure 4.2; Figure 4.3). From the 7 studies that reported completeness of data at the 12 month follow-up period, 5 (72%) were judged to be at low risk of bias. The remaining 12 (41%) studies were judged to be at high risk of bias (Figure 4.2; Figure 4.3).

Of the included studies, 5 studies performed an ITT analysis. Included in these trials were FAST [84], the study by van Vonderen et al [86], SANDS [76], MIDAS [79] and PLAC-II [62]. Of the remaining studies, 19 provided complete information about the analysis strategy that was used. In another 5 studies the analysis strategy was unclear. Overall, 11 studies based the analysis on subjects who had data from an ultrasound scan at baseline and 1 or more times during the follow-up period. For 2 studies in which the analysis strategy was unclear, ENHANCE [91] and BCAPS [96], information in the reports suggested that the analysis followed the same principle. METEOR [73] included subjects who received the medication they were assigned and had both baseline measurements and data from at least 1 follow-up scan. The analysis strategy for STARR [85] and VHAS [93] included subjects who had data from 2 or more and 3 more follow-up scans, respectively, in addition to data from the baseline scan. In the MAVET [97] trial, the results for each of the follow-up measurements were based on subjects who had data for the time point under consideration. For 2 studies in which the analysis strategy was unclear, ATIC [90] and OPAL [83], the information in the reports suggested that the analyses plan was the same as

that followed in MAVET [97]. The ASAP [65] study by Smilde et al excluded subjects who withdrew before receiving treatment. CAPTIVATE [88] excluded subjects who did not have data from scans at least 40 weeks apart. The CLAS [81] trial based the analysis on subjects who completed the trial and had "evaluable coronary endpoints". The interpretation of the aforementioned quote was considered to be unclear. In the ASAP [67] study by Salonen et al, subjects were included in the analysis if they had data from scans taken at the end of the follow-up period. The analysis strategy for the SENDCAP [89] study was unclear; however, information in the report suggested that it was based on subjects who had CIMT data at baseline and for each of the follow-up measurements.

Of the 7 studies that reported the completeness of the CIMT data at the 12 month follow-up time, 3 studies, namely, ELVERA [99], CAPTIVATE [88] and CHICAGO [95], had no withdrawals. In one study, ASAP [65] by Smilde et al, there were withdrawals. In the remaining 3 studies, there were no withdrawals in addition to the subjects who were already excluded from the analysis, since it was limited to subjects who had CIMT data at the 12 month follow-up time. Included in these studies were ATIC [90], OPAL [83], and MAVET [97].

Clinical outcomes

A total of 24 studies provided information on the completeness of data for the clinical outcomes at the end of the follow-up period. Of these, 22 provided information for each of the randomized groups and 2 studies, VEAPS [101] and STARR [85], provided information for the overall sample size. In 5 (17%) studies there was no information about the completeness of clinical outcome data and the risk of bias in these studies was judged to be 'unclear' (Figure 4.2; Figure 4.3). Overall, 10 (34%) studies were judged to be at low risk of bias and 12 (41%) studies were judged to be at high risk of bias (Figure 4.2; Figure 4.3).

Overall, 20 studies performed a true ITT analysis. From the remaining studies, 8 provided information about the analysis strategy that was used and in 1 study, BCAPS [96], no information was provided. ASAP [65] and METEOR [73] based the analysis on subjects who received treatment. In the CAPTIVATE [88] trial, subjects were included if they were not

lost to follow-up and if they had valid lipid values and cardiovascular endpoints. In the ELVA [87] trial, subjects were excluded from the analysis if they were withdrawn during placebo-run. OPAL [83] excluded subjects who dropped out at baseline and CHICAGO [95] excluded subjects who did not receive the treatment as assigned. In the CLAS [81] trial, subjects were included in the analysis if they completed the study and had evaluable carotid endpoints. In the VHAS [93] study, subjects were included in the analysis for the clinical outcomes if they had CIMT data for all carotid segments that were measured in the study.

In total, 21 studies had data missing due to withdrawals. In 2 studies, SANDS [76] and van Vonderen et al [86], there were no withdrawals. In 5 studies, there was no information about withdrawals. These studies included RADIANCE 2 [92], MIDAS [79], PLAC-II [62], VHAS [93] and BCAPS [96]. In the CLAS [81] study, there were no withdrawals in addition to the subjects who were already excluded from the analysis, since it was limited to subjects who completed the trial.

4.4.2.5 Selective reporting

All studies reported CIMT and 25 studies reported clinical outcome data as described in the method's sections of the report(s) of the studies. In 4 trials, there was no information about the clinical events monitoring in the methods section; however, the information presented in the reports suggests that all clinical outcome data would have been reported correctly. All studies were judged to be at low risk of bias.

4.4.2.6 Other potential sources of bias

The source of funding was considered as a potential source of bias because there is evidence to show that industry-funded trials may exaggerate the treatment effect[76, 120]. A total of 24 studies received funding from industry and with the exception of one study, EPAT [94], all were judged to be at high risk of bias. This represented 79% of the studies. The EPAT [94] trial stated that the authors had complete control over the study and for this reason, it was judged to be at low risk of bias. The remaining 5 studies received their funding from government agencies and were also judged to be at low risk of bias.

4.4.3 Effects of interventions

In applying the statistical procedures to the data obtained from the systematic review, it was found that a minimum of 3 studies were required to generate statistics for all procedures. For analyses that were not limited to specific drug classes, the review identified a sufficient number of studies to evaluate the strength of the 12 month difference in CIMT as a surrogate outcome for each of the clinical outcomes at both 2 or more and 4 or more years of follow-up. These results are described and summarized in Table 4.4. For analyses in specific drug classes, there was sufficient data to evaluate CIMT as a surrogate outcome for all-cause mortality and myocardial infarction. For all-cause mortality, there was enough data to consider two drug classes, namely: hormone therapy vs. placebo and statins vs. placebo. For myocardial infarction, data was available for the comparison of statins vs. placebo.

4.4.3.1 Non-drug class specific

Table 4.4: Summary of results for non-drug class specific comparisons

Outcome	Follow-up duration (years)	Surrogate evaluation methods								
		Spearman's ρ	p-value	Regression analysis				Korn et al model		% concord -ance
				R^2	β_1	STE	p-value	β_1	p-value	
All-cause mortality	≥2	0.0650	0.7525	0.0281	1.4614	None	0.4129	1.9977	0.5037	65.4%
	≥4	0.8	0.1041	0.7145	43.3817	Trend	0.0714	43.3818	0.1676	80%
Myocardial infarction	≥2	0.4008	0.0425	0.1399	4.8558	Trend	0.0598	4.6544	Not estimable	84.62%
	≥4	0.1	0.8729	0.4756	24.8234	None	0.1976	24.8233	0.2228	100%
Stroke	≥2	-0.2958	0.2660	0.0948	-22.4059	None	0.2460	-22.4059	0.3272	87.5%
	≥4	-1.0	<0.0001	0.9956	-65.0909	Trend	0.0425	-65.0909	0.3971	100%
Cardio-vascular mortality	≥2	0.5972	0.0054	0.2063	3.5805	Trend	0.0442	4.0907	Not estimable	70%
	≥4	-1	<0.0001	0.5720	-10.1374	None	0.2437	-10.1374	0.7672	75%

All-cause mortality

Studies with ≥ 2 years of follow-up

Table 4.5 contains the study specific 12 month mean differences in CIMT and log odds ratios for all-cause mortality, along with the 95% CIs and standard errors.

To evaluate CIMT as a surrogate endpoint for all-cause mortality at 2 or more years of follow-up, data from 23 studies was used. A total of 26 pairwise comparisons were considered in most of the analyses because 3 studies had multiple treatment arms. For the fixed effects analysis, 25 pairwise comparisons were used to obtain results. Collectively, the results of the analyses indicated that there was no statistically significant association

between CIMT and all-cause mortality. Qualitatively, there was a positive trend between the outcomes. The results of the graphical analysis are shown in Figure 4.4. Each point on the plot corresponds to the estimate of the mean difference in CIMT and log odds ratio for all-cause mortality in a given study. The precision in the estimates for the mean differences and log odds ratios is represented by the length of the horizontal and vertical confidence intervals, respectively. Mean differences in CIMT increasing towards zero from left to right on the graph indicate a successively smaller between group difference in CIMT, with the experimental arm having a lower carotid thickness than the control arm. Further increases in the mean difference in CIMT beyond zero indicate successively large between group differences, with increased carotid thicknesses in the experimental arm relative to the control arm. Likewise, log odds ratios increasing towards zero from bottom to top of the graph indicate successively smaller differences between groups in all-cause mortality, with a relatively lower risk for all-cause mortality in the experimental arm. Further increases in the log odds ratio beyond zero indicate a successively larger difference in all-cause mortality with a larger risk in the experimental arm relative to the control arm. For there to be a strong association between CIMT and all-cause mortality so that CIMT may be considered a very good surrogate for this clinical outcome, the estimates for both outcomes should be precise. Also, there is a positive association between the outcomes. Increasing mean differences in CIMT from left to right on the graph should be associated with consistently increasing log odds ratios for all-cause mortality from bottom to top of the graph. A horizontal scatter of points with precise estimates for both outcomes would suggest that CIMT is not a good surrogate outcome for all-cause mortality. The conclusiveness of any observed pattern (lack of pattern), depends on the precision in the estimates for the outcomes. Compared to patterns in which only one outcome was estimated with imprecision, speculation increases when both outcomes have imprecise estimates.

In the graphical analysis, there was evidence of a positive association between the outcomes, but this finding was inconclusive (Figure 4.4). A pattern of increasing log odds ratios was observed for the mean differences that increased from approximately -0.07 to 0.005; however, the vertical confidence intervals were relatively long and suggested that estimates for the log odds ratios were relatively imprecise. Because of this imprecision in

the estimates for the log odds ratios, it was difficult to ascertain whether any true association exists between CIMT and all-cause mortality. Indeed, for a given mean difference, the plot indicated a range of odds ratios for all-cause mortality. The plot indicated that for most of the studies, the horizontal confidence intervals were short and for this reason, the estimates of the mean differences in CIMT were precise.

The results of the rank correlation analysis indicated that there was no association between the outcomes. Spearman's $\rho = 0.0650$ (p-value = 0.7525). A positive correlation was noted, but it was close to zero and not significant ($p > 0.05$). The results of the weighted regression indicated that there was no association between the outcomes. The results of the weighted regression analysis indicated that $R^2=0.0281$ (p-value = 0.4129). For this regression, the intercept was $\beta_0 = -0.3305$ (p-value = 0.0108; SE = 0.1195) and the slope coefficient was $\beta_1 =1.4614$ (p-value = 0.4129; SE = 1.7539). A positive trend was noted, but only 2.81% of the variation in the log odds ratios was explained by the 12 month mean differences in CIMT and the regression was not significant ($p > 0.05$).

The graph to estimate the STE is shown in Figure 4.5. For CIMT to be a good surrogate outcome for all-cause mortality, a STE should be observed.A STE is observed when there is a negative mean difference in CIMT below which the upper prediction band remains below the x-axis. This value of the mean difference in CIMT is associated with the following characteristics: i) from right-to-left on the graph, the mean differences are successively more negative and the carotid thickness is always lower in the experimental arm; ii) from right-to-left on the graph, values of the log odds ratios along the upper prediction band become successively more negative, and the risk of mortality is always lower in the experimental arm. That is, from right-to-left on the graph, values of the log odds ratios along the upper prediction band are < 0 and the odds ratios are < 1.

In figure 4.5, there was no STE. The lack of a STE indicated that there is no evidence to suggest that CIMT is a good surrogate outcome for all-cause mortality at 2 or more years of follow-up. In the graph, there was no point at which the upper prediction band crossed and remained below the x-axis.

When the fixed effects model as described by Korn et al [32] was applied to the data, $\beta_0 = -0.3128$ (p-value = 0.0760; SE = 0.1680) and $\beta_1 = 1.9977$ (p-value = 0.5037; SE = 2.0208). These results suggested that there was a positive association between the outcomes, but the regression was not significant ($p > 0.05$).

When considering the study wise agreement between the two outcomes, in 17 of the 26 comparisons the effect of treatment on the surrogate outcome agreed with the clinical outcome and the percent concordance was 65.4%. In all of these 17 studies, there was no statistically significant treatment effect on either of the outcomes (all p-values > 0.05). In the remaining 9 comparisons, the treatment had a statistically significant effect on CIMT (p-values for CIMT differences < 0.05) but not on all-cause mortality. In 5 of the 9 comparisons, the treatment associated with a significantly lower carotid thickness was also associated with a non-significant lower risk of all-cause mortality. In 2 comparisons, tibolone vs. placebo and cee/mpa vs. placebo, treatment as compared to placebo was associated with a significantly greater carotid thickness and a non significant lower risk of mortality. For the comparison of tibolone vs. placebo the mean difference in CIMT was 0.021 (95% CI: 0.008, 0.035) and the log odds ratio for all-cause mortality was -1.7541 (-4.9618, 1.4537). For the comparison of cee/mpa vs. placebo, the mean difference was 0.013 (95% CI: 0.001, 0.023) and log odds ratio was -1.7937 (95% CI: -5.0014, 1.4140). In another 2 comparisons, the treatment associated with the significantly lower CIMT had a non-significant, higher risk of all-cause mortality. In the ASAP [67] study by Salonen et al which compared vitamin E plus C versus placebo, the supplementation group had a significantly lower CIMT, and a non-significant higher risk of mortality. The mean difference in CIMT for the supplementation vs. placebo group was -0.004 (95% CI: -0.007, -0.001) and the log odds ratio for all-cause mortality was 0.7738 (95% CI: -0.4604, 2.0080). In the ELVERA [99] trial which compared amlodipine vs. lisinopril, the amlodipine group had a significantly lower CIMT and a non-significant, higher risk of mortality. The mean difference in CIMT for amlodipine vs. lisinopril was -0.083 (95% CI: -0.1483, -0.0177) and the log odds ratio was 3.1863 (95% CI: 0.1279, 79.3588).

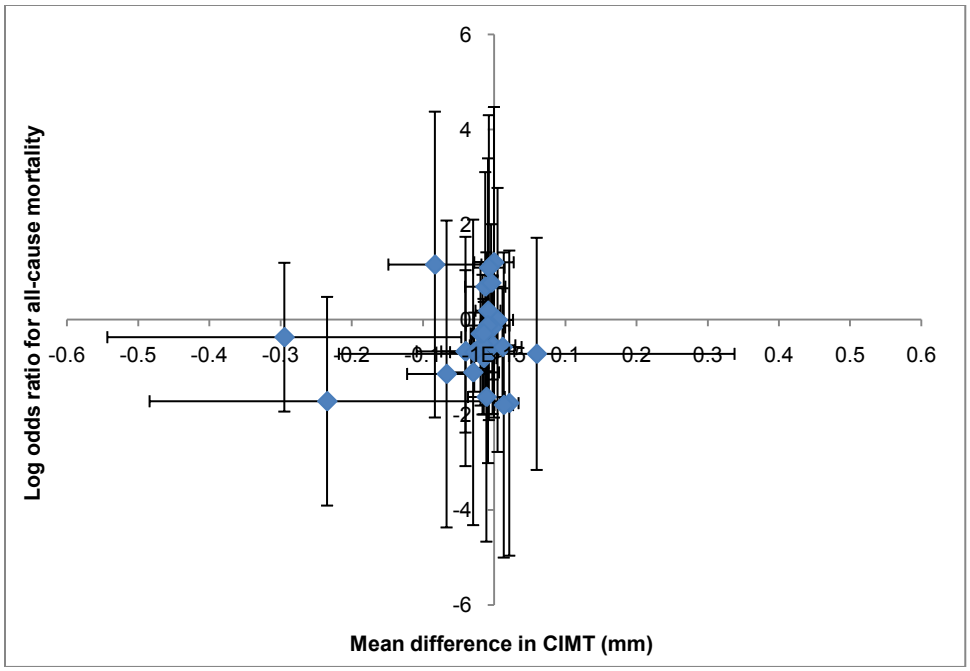

Figure 4.4: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality at 2 or more years of follow-up

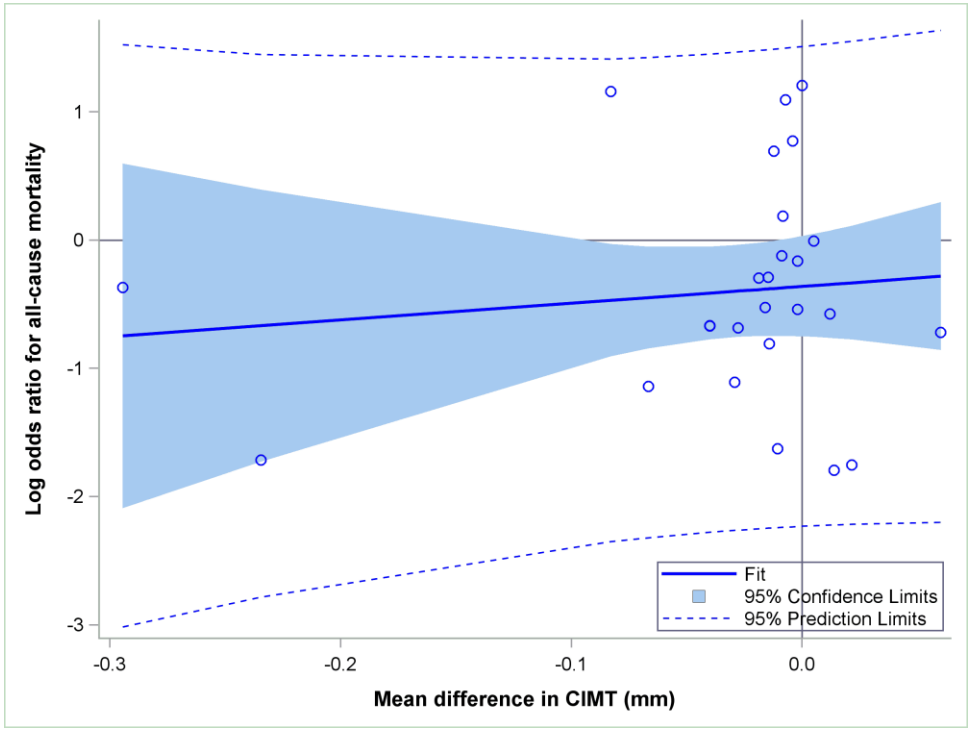

Figure 4.5: STE for 12 month mean difference in CIMT and all-cause mortality at 2 or more years of follow-up.

Studies with ≥ 4 years of follow-up

For the evaluation of CIMT as a surrogate outcome for all-cause mortality measured at 4 or more years post-treatment, data from 5 studies was used. The results indicated that there was no significant relationship between CIMT and all-cause mortality. Qualitatively, a positive trend was observed between the outcomes and the magnitude was greater in comparison to the results for studies with 2 or more years of follow-up. The graphical analysis (Figure 4.6) indicated CIMT may be a good surrogate outcome for all-cause mortality. This observation was somewhat inconclusive. The graph showed that in general, increases in the values for the mean differences in CIMT from left-to-right on the graph were associated with increases in the values for the log odds ratios for all-cause mortality from bottom-to-top of the graph. However, the plot consisted of a small number of points and from this small number, two of the points were associated with imprecision in the estimates for the mean differences and/or log odds ratios. One point (mean difference = -0.04, log odds ratio = -0.67) on the graph was associated with imprecision in the estimates for both outcomes. For another point, (mean difference = -0.007, log odds ratio = 1.094) there was a large amount of imprecision in the estimate of the log odds ratio. Spearman's $\rho = 0.8000$ (p-value = 0.1041). From the weighted regression, $R^2 = 0.7145$ (p-value = 0.0714). In the weighted regression, $\beta_0 = 0.5748$ (p-value = 0.1932; SE = 0.3439) and $\beta_1 = 43.3817$ (p-value = 0.0714; SE = 15.8339). In the graph for the STE, a trend towards a STE was observed (Figure 4.7). From right-to-left on the graph, it was observed that the upper prediction band was approaching the x-axis as the values of the mean differences and log odds ratios were becoming successively more negative. The graph indicated that below a certain threshold value for the CIMT mean difference, carotid thicknesses that are lower in the experimental arm relative to the control arm post treatment, may be able to predict a lower risk of all-cause mortality. Since there was no negative mean difference in CIMT below which the upper prediction band crossed and remained below the x-axis, an actual STE was not observed. From the graph, it appears that to the left of the mean difference value of -0.04, mean differences and log odds ratios would continue to be more negative; however, with a lack of data, the prediction band does not extend to the left of this value and for this reason it is not known whether a STE would be identified. For the Korn [32]fixed effects analysis, $\beta_0 = 0.5748$ (p-value = 0.3496;

SE = 0.5199) and β_1 = 43.3818 (p-value = 0.1676; SE =23.9387). For the study wise agreement, there was concordance between the effect of treatment on each outcome in 4 of 5 comparisons and the percent concordance was 80%. In all 4 studies, there was no significant treatment effect on either of the outcomes (all p-values >0.05). Disagreement between the effect of treatment on the CIMT versus all-cause mortality was observed in the ASAP [67] study as described in the previous section.

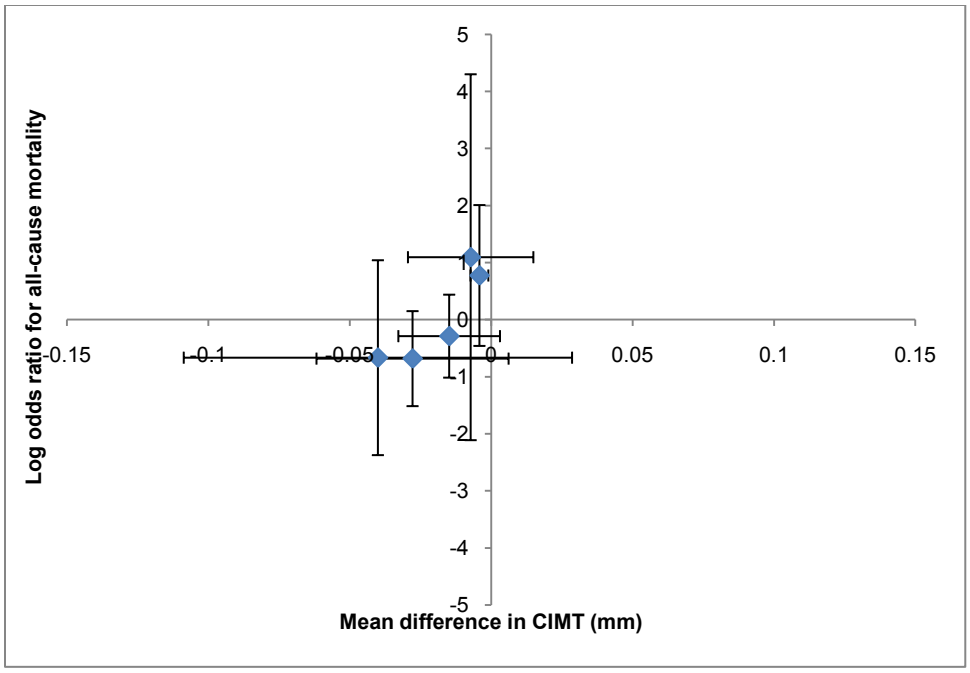

Figure 4.6: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality at 4 or more years of follow-up

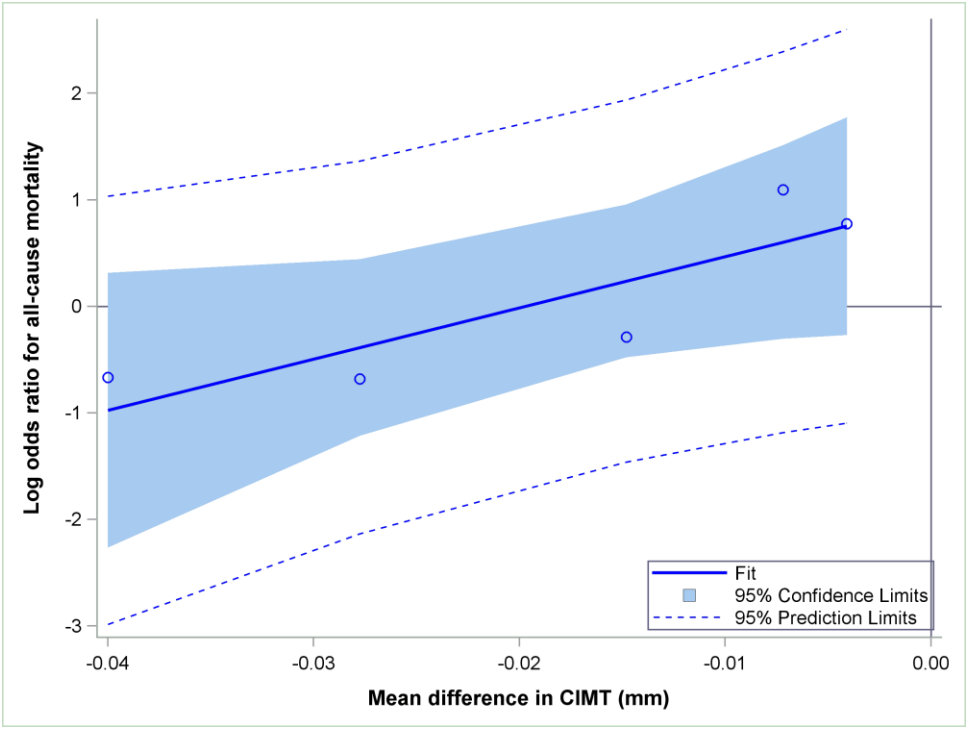

Figure 4.7: STE for 12 month mean difference in CIMT and all-cause mortality at 4 or more years of follow-up

Table 4.5: Estimates of treatment effect for CIMT and all-cause mortality

Reference	Experimental drug class	Control drug class	Study dur-ation	Mean differ-ence (MD) for CIMT	LCL (MD)	UCL (MD)	SE (MD)	Odds ratio (OR) for all-cause mort-ality	LCL (OR)	UCL (OR)	Log odds ratio (LOR) for all-cause mort-ality	LCL (LOR)	UCL (LOR)	SE (LOR)
EPAT (Hodis)	Hormone therapy	Placebo	2	-0.0293	-0.0657	0.0071	0.0186	0.3303	0.0133	8.1976	-1.1076	-4.3191	2.1038	1.6385
ASAP (Smilde)	Statin (atorvstatin)	Statin (simvastatin)	2	-0.0400	-0.0741	-0.0059	0.0174	0.5126	0.0460	5.7094	-0.6683	-3.0787	1.7421	1.2298
FAST (Sawayama)	Antihyertensive	Placebo	2	-0.2345	-0.4838	0.0148	0.1272	0.1800	0.0201	1.6144	-1.7148	-3.9086	0.4790	1.1193
Sawayama (FAST)	Statin	Placebo	2	-0.2945	-0.5431	-0.0459	0.1268	0.6923	0.1449	3.3078	-0.3677	-1.9317	1.1963	0.7980
FAST (Sawayama)	Antihyertensive	Statin	2	0.0600	-0.2179	0.3379	0.1418	0.4875	0.0425	5.5969	-0.7185	-3.1591	1.7222	1.2452
ELVERA (Terpstra)	CCB	ACE inhibitor	2	-0.0830	-0.1483	-0.0177	0.0333	3.1863	0.1279	79.3588	1.1589	-2.0562	4.3740	1.6404
RADIANCE 2 (Bots)	Statin + CETP inhibitor	Statin	2	0.0050	-0.0168	0.0268	0.0111	0.9947	0.0620	15.9621	-0.0053	-2.7809	2.7702	1.4161
METEOR(Crouse)	Statin	placebo	2	-0.0084	-0.0260	0.0092	0.0090	1.2073	0.0490	29.7264	0.1884	-3.0153	3.3920	1.6345
ATIC (Nanayakkara)	Statin + vitamin E + placebo	placebo	2	-0.0666	-0.1220	-0.0112	0.0283	0.3193	0.0127	8.0426	-1.1416	-4.3680	2.0848	1.6461
Van Vonderen	NRTI	cART	2	0.0000	-0.0277	0.0277	0.0141	3.3429	0.1277	87.5240	1.2068	-2.0583	4.4719	1.6659
VEAPS (Hodis)	Vitamin E	Placebo	3	-0.0123	-0.0406	0.0160	0.0144	2.0000	0.1797	22.2596	0.6931	-1.7165	3.1028	1.2294
SANDS (Howard)	Aggressive treatment	Standard treatment	3	-0.0020	-0.0340	0.0300	0.0163	0.5831	0.1378	2.4668	-0.5393	-1.9816	0.9029	0.7359
STARR (Lonn)	ACE inhibitor	Placebo	3	-0.0020	-0.0197	0.0157	0.0090	0.8499	0.2842	2.5415	-0.1626	-1.2581	0.9328	0.5589
STARR (Lonn)	Glitazone	Placebo	3	-0.0143	-0.0320	0.0034	0.0090	0.4457	0.1366	1.4540	-0.8081	-1.9905	0.3743	0.6033
MIDAS (Borhani)	CCB	Thiazide diuretic	3	-0.0089	-0.0397	0.0219	0.0157	0.8848	0.3382	2.3146	-0.1224	-1.0841	0.8393	0.4906
BCAPS (Hedblad)	Beta blocker	Placebo	3	0.0120	-0.0148	0.0388	0.0137	0.5626	0.1634	1.9375	-0.5752	-1.8117	0.6614	0.6309
OPAL (bots)	Hormone therapy (tibolone)	Placebo	3	0.0215	0.0084	0.0345	0.0067	0.1731	0.0070	4.2787	-1.7541	-4.9618	1.4537	1.6366
OPAL (Bots)	Hormone therapy (CEE/MPA)	Placebo	3	0.0138	0.0007	0.0268	0.0067	0.1663	0.0067	4.1124	-1.7937	-5.0014	1.4140	1.6366
BVAIT (Hodis)	Vitamin B	Placebo	3	-0.0107	-0.0365	0.0151	0.0132	0.1969	0.0094	4.1212	-1.6253	-4.6667	1.4161	1.5517
PLAC II (Byington)	Statin	Placebo	3	-0.0161	-0.0320	-0.0002	0.0081	0.5917	0.1363	2.5693	-0.5248	-1.9933	0.9436	0.7492
KAPS (Salonen)	Statin	Placebo	3	-0.0189	-0.0308	-0.0070	0.0061	0.7432	0.1644	3.3596	-0.2968	-1.8054	1.2118	0.7697
VHAS (Zanchetti)	CCB	Thiazide diuretic	4	-0.0400	-0.1087	0.0287	0.0350	0.5135	0.0931	2.8319	-0.6665	-2.3739	1.0410	0.8711
MAVET (Magliano)	Vitamin E	Placebo	4	-0.0277	-0.0617	0.0062	0.0173	0.5051	0.2197	1.1612	-0.6830	-1.5154	0.1494	0.4247
ELSA (Zanchetti)	CCB	Beta blocker	4	-0.0148	-0.0328	0.0032	0.0092	0.7489	0.3621	1.5491	-0.2891	-1.0158	0.4376	0.3708
CHICAGO (Mazzone)	Glitazone	Sulfonylurea	6	-0.0072	-0.0294	0.0150	0.0113	2.9869	0.1210	73.7108	1.0942	-2.1117	4.3001	1.6357
ASAP (Salonen)	Vitamin E + vitamin C	Placebo	6	-0.0041	-0.0073	-0.0009	0.0016	2.1680	0.6310	7.4487	0.7738	-0.4604	2.0080	0.6297

Myocardial infarction

Studies with ≥ 2 years of follow-up

Table 4.6 contains the study specific 12 month mean differences in CIMT and log odds ratios for myocardial infarction, along with the 95% CIs and standard errors.

For the evaluation of CIMT as a surrogate outcome for myocardial infarction at 2 or more years of follow-up, data from 23 studies was used. A total of 26 pair-wise comparisons were used in most of the analyses because two trials had multiple treatment arms. The trials with multiple arms were FAST [84] and STARR [85]. For the application of the fixed effect model, 25 pair-wise comparisons were used to obtain results. Overall, a significant, positive relationship was observed between CIMT and myocardial infarction. From the graphical analysis, there was evidence of a positive association between the outcomes (Figure 4.8). Spearman's ρ=0.4008 (p-value = 0.0425), indicating that the correlation was significant. The weighted R^2 = 0.1399 (p=0.0598) and the regression was marginally significant. In the weighted regression, β_0 = -0.1676 (p-value = 0.3227; SE = 0.1660) and β_1 = 4.8558 (p-value = 0.0598; SE = 2.4580). A trend towards a STE was observed (Figure 4.9). For the Korn fixed effects analysis [32], β_0 = -0.1588 (p-value = 0.3775; SE = 0.1764) and β_1 = 4.6544 (p-value = not estimable; SE = 2.0818).

The analysis for study wise percent agreement indicated that the effect of treatment on CIMT was concordant with its effect on the clinical outcome in 22 of the 26 comparisons and the percent concordance was 84.62%. In 20 of the 22 comparisons with concordant results, there was no significant treatment effect on either of the outcomes (all p-values > 0.05). In the 2 remaining comparisons, there was a statistically significant treatment effect in favour of the experimental arm for each of the outcomes. In the PLAC-II [62] trial which compared pravastatin vs. placebo, the mean difference for CIMT was -0.0161 (95% CI: -0.0320, -0.0002) and the log odds ratio for myocardial infarction was -1.7102 (95% CI: -3.2645, -0.1560). In the CLAS [81] trial which compared colestipol plus niacin plus diet therapy vs. placebo plus diet therapy, the mean difference was -0.045 (95% CI: -0.0536, -0.0364) and the log odds ratio was -1.1209 (95% CI: -1.9793, -0.2625). For the 4 studies in which there was disagreement in the effect of treatment on CIMT versus the

effect on myocardial infarction, there was a significant difference between groups in CIMT (p-values for CIMT differences < 0.05) but there was no significant treatment effect on the risk of myocardial infarction. In 3 of these 4 studies- FAST [84], KAPS [98] and CAPTIVATE [88]- the effect of treatment was in the same direction for both outcomes. In FAST [84] and KAPS [98], the experimental group had a significantly lower CIMT and a non-significant, lower risk of myocardial infarction. For the comparison of pravastatin vs. placebo in the FAST [84] trial, the mean difference in CIMT was -0.2945 (95% CI: -0.54311, -0.04589) and the log odds ratio for myocardial infarction was -1.0498 (95% CI: -2.7513, 0.6516). In the KAPS [98] trial, the mean difference in CIMT for pravastatin vs. placebo was -0.0189 (95% CI: -0.0308, -0.0070) and the log odds ratio for myocardial infarction was -1.0084 (95% CI: -2.3485, 0.3318). In CAPTIVATE [88], which compared pactimibe vs. placebo, the experimental group had a significantly higher carotid thickness and a non-significant, higher risk of myocardial infarction. The mean difference in CIMT for pactimibe vs. placebo was 0.023 (95% CI: 0.0001, 0.0459) and the log odds ratio for mortality was 13.0297 (95% CI: 0.7318, 232.0084). In the remaining study with discordant results, the effect of treatment was in the opposite direction for each of the outcomes. In the ATIC [90] study which compared pravastatin plus vitamin E plus vitamin B vs. placebo, the experimental group had a significantly lower CIMT and a non-significant, greater risk of myocardial infarction as compared to the placebo group. The mean difference in CIMT between the pravastatin plus vitamin E plus vitamin C vs. placebo group was -0.0666 (95% CI = -0.1222, -0.0112) and the log odds ratio was 0.6931 (95% CI: -1.7227, 3.1290).

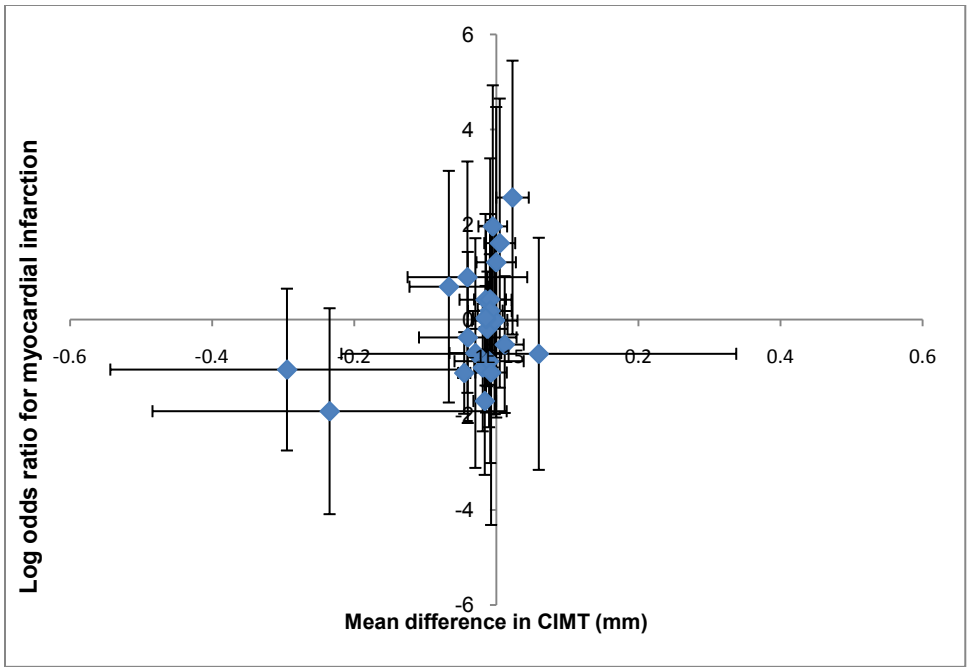

Figure 4.8: Graphical analysis for 12 month mean difference in CIMT and myocardial infarction at 2 or more years of follow-up

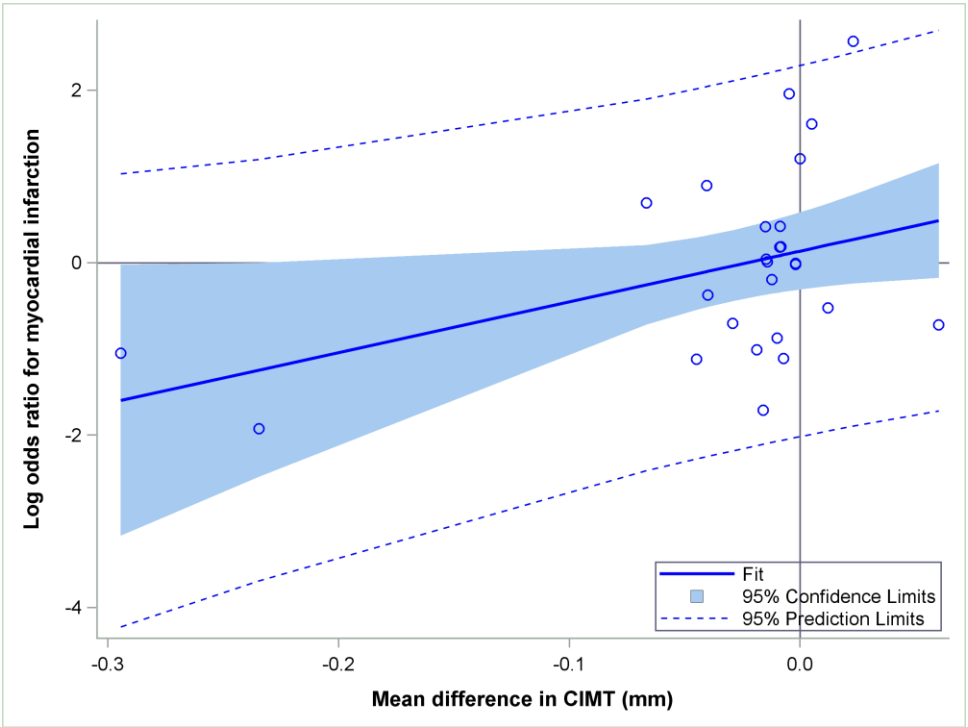

Figure 4.9: STE for 12 month mean difference in CIMT and myocardial infarction at 2 or more years of follow-up.

Studies with ≥ 4 years of follow-up

To evaluate CIMT as a surrogate endpoint for myocardial infarction at 4 or more years of follow-up, data was used from 5 studies. Overall, the results indicated that there was no association between the two outcomes. The results showed a positive, non-significant association between CIMT and myocardial infarction. From the graphical analysis, no association was observed between the outcomes (Figure 4.10). The Spearman's ρ = 0.1000 (p=0.8729) and was no longer significant. The weighted R^2 = 0.4756 (p=0.1976) and the regression was not significant. In the weighted regression, β_0 = 0.1516 (p-value = 0.7508; SE = 0.4356) and β_1 = 24.8234 (p-value = 0.1976; SE = 15.0489). No STE was observed (Figure 4.11). For the Korn fixed effects analysis [32], β_0 = 0.1516 (p-value = 0.7676; SE = 0.4687) and β_1 = 24.8233 (p-value = 0.2228; SE = 16.1906). There was agreement between all 5 comparisons included in the analysis of study wise agreement, resulting in a percent concordance of 100%. In 4 of the 5 comparisons, there was no significant treatment effect for either of the outcomes (all p-values > 0.05). In the CLAS [81] study, a statistically significant benefit was observed in the experimental group for both outcomes, as described in the previous section.

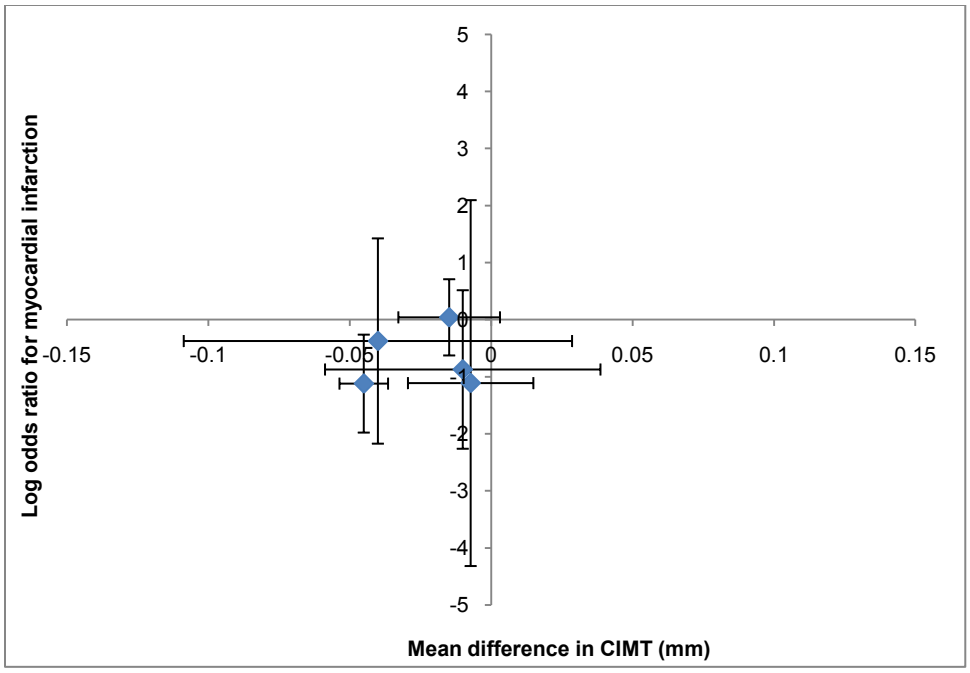

Figure 4.10: Graphical analysis for 12 month mean difference in CIMT and myocardial infarction at 4 or more years of follow-up

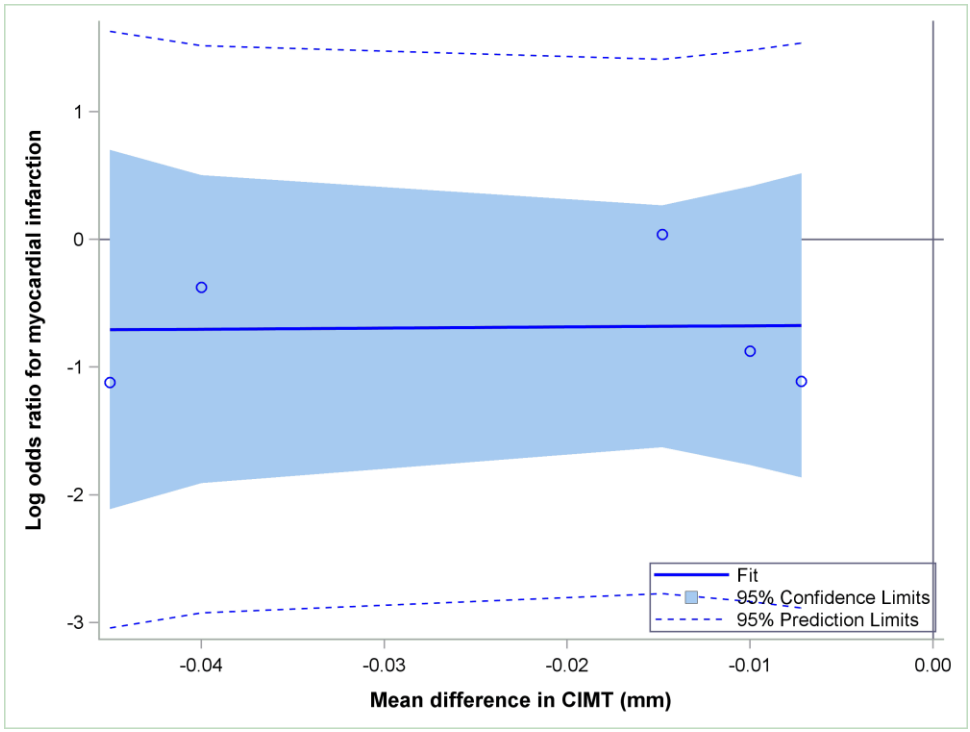

Figure 4.11: STE for 12 month mean difference in CIMT and myocardial infarction at 4 or more years of follow-up.

Table 4.6: estimates of treatment effect for CIMT and myocardial infarction

Reference	Experimental drug class	Control drug class	Study dur-ation	Mean differ-ence (MD) for CIMT	LCL (MD)	UCL (MD)	SE (MD)	Odds ratio for all-cause mort-ality	LCL (OR)	UCL (OR)	Log odds ratio	LCL (LOR)	UCL (LOR)	SE (LOR)
TART (Hodis)	Glitazone	Placebo	2	-0.0150	-0.0516	0.0216	0.0187	1.5205	0.2504	9.2336	0.4191	-1.3847	2.2228	0.9203
EPAT (Hodis)	Hormone therapy	Placebo	2	-0.0293	-0.0657	0.0071	0.0186	0.4955	0.0443	5.5444	-0.7023	-3.1174	1.7128	1.2322
FAST (Sawayama)	Antihypertensive	Placebo	2	-0.2345	-0.4838	0.0148	0.1272	0.1458	0.0167	1.2710	-1.9253	-4.0904	0.2398	1.1046
FAST (Sawayama)	Statin	Placebo	2	-0.2945	-0.5431	-0.0459	0.1268	0.3500	0.0638	1.9187	-1.0498	-2.7513	0.6516	0.8681
FAST (Sawayama)	Antihypertensive	Statin	2	0.0600	-0.2179	0.3379	0.1418	0.4875	0.0425	5.5969	-0.7185	-3.1591	1.7222	1.2452
ENHANCE (Kastelein)	Statin + cholesterol absorption inhibitor	Statin	2	-0.0087	-0.0311	0.0137	0.0115	1.5297	0.2541	9.2098	0.4250	-1.3702	2.2203	0.9159
RADIANCE 2 (Bots)	Statin + CETP inhibitor	Statin	2	0.0050	-0.0168	0.0268	0.0111	5.0000	0.2392	104.5039	1.6094	-1.4303	4.6492	1.5509
METEOR (Crouse)	Statin	Placebo	2	-0.0084	-0.0260	0.0092	0.0090	1.2073	0.0490	29.7264	0.1884	-3.0153	3.3920	1.6345
RADIANCE 1 (Kastelein)	Statin + CETP inhibitor	Statin	2	-0.0048	-0.0250	0.0154	0.0103	7.1095	0.3662	138.0404	1.9614	-1.0047	4.9275	1.5133
ATIC (Nanayakkara)	Statin + vitamin E + vitamin B	Placebo	2	-0.0666	-0.1220	-0.0112	0.0283	2.0000	0.1751	22.8500	0.6931	-1.7427	3.1290	1.2428
CAPTIVATE (Meuwese)	ACAT inhibitor	Placebo	2	0.0230	0.0001	0.0459	0.0117	13.0297	0.7318	232.0084	2.5672	-0.3123	5.4468	1.4692
Van Vonderen	NRTI	cART	2	0.0000	-0.0277	0.0277	0.0141	3.3429	0.1277	87.5240	1.2068	-2.0583	4.4719	1.6659
VEAPS (Hodis)	Vitamin E	Placebo	3	-0.0123	-0.0406	0.0160	0.0144	0.8236	0.2467	2.7500	-0.1940	-1.3996	1.0116	0.6151
SANDS (Howard)	Aggressive treatment	Standard treatment	3	-0.0020	-0.0340	0.0300	0.0163	0.9800	0.1370	7.0125	-0.0202	-1.9881	1.9477	1.0040
STARR (Lonn)	ACE inhibitor	Placebo	3	-0.0020	-0.0197	0.0157	0.0090	0.9930	0.1395	7.0690	-0.0070	-1.9698	1.9557	1.0014
STARR (Lonn)	Glitazone	Placebo	3	-0.0143	-0.0320	0.0034	0.0090	1.0099	0.1419	7.1894	0.0099	-1.9529	1.9726	1.0014
MIDAS (Borhani)	CCB	Thiazide diuretic	3	-0.0089	-0.0397	0.0219	0.0157	1.2000	0.3635	3.9613	0.1823	-1.0119	1.3766	0.6093
ELVA (Wiklund)	Beta blocker	Placebo	3	-0.0405	-0.1247	0.0437	0.0430	2.4444	0.2146	27.8389	0.8938	-1.5388	3.3264	1.2411
BCAPS (Hedblad)	Beta blocker	Placebo	3	0.0120	-0.0148	0.0388	0.0137	0.5923	0.1406	2.4957	-0.5237	-1.9620	0.9146	0.7338
PLAC II (Byington)	Statin	Placebo	3	-0.0161	-0.0320	-0.0002	0.0081	0.1808	0.0382	0.8556	-1.7102	-3.2645	-0.1560	0.7930
KAPS (Salonen)	Statin	Placebo	3	-0.0189	-0.0308	-0.0070	0.0061	0.3648	0.0955	1.3935	-1.0084	-2.3485	0.3318	0.6837
VHAS (Zanchetti)	CCB	Thiazide diuretic	4	-0.0400	-0.1087	0.0287	0.0350	0.6877	0.1138	4.1550	-0.3744	-2.1731	1.4243	0.9177
ELSA (Zanchetti)	CCB	Beta blocker	4	-0.0148	-0.0328	0.0032	0.0092	1.0415	0.5341	2.0310	0.0406	-0.6273	0.7085	0.3408
SENDAP (Elkeles)	Fibrate	Placebo	5	-0.0100	-0.0587	0.0387	0.0248	0.4176	0.1041	1.6747	-0.8733	-2.2622	0.5157	0.7086
CHICAGO (Mazzone)	Glitazone	Sulfonylurea	6	-0.0072	-0.0294	0.0150	0.0113	0.3290	0.0133	8.1189	-1.1117	-4.3176	2.0942	1.6357
CLAS (Hodis)	Colestipol + niacin + diet therapy	Placebo + diet therapy	10	-0.0450	-0.0536	-0.0364	0.0044	0.3260	0.1382	0.7691	-1.1209	-1.9793	-0.2625	0.4379

Stroke

Studies with ≥ 2 years of follow-up

Table 4.7 contains the study specific 12 month mean differences in CIMT and log odds ratios for stroke, along with the 95% CIs and standard errors.

For the evaluation of CIMT as a surrogate outcome for stroke in studies with 2 or more years of follow-up, data was obtained from 15 studies. A total of 16 pairwise comparisons were used in the analysis because the STARR[85] trial had multiple treatment arms. Overall, the results showed that there was no significant association between CIMT and stroke. Qualitatively, a negative relationship was observed between the outcomes. From the graphical analysis (Figure 4.12), there was some indication of a negative relationship between the outcomes, but there was a large amount of imprecision in this observation. The Spearman's ρ= -0.2958 (p=0.2660) and the weighted R^2 = 0.0948. From the weighted regression, β_0 = -0.5526 (p-value = 0.0846; SE = 0.2977) and β_1 = -22.4059 (p-value = 0.2460; SE = 18.5044). There was no STE (Figure 4.13). For the Korn fixed effects analysis [32], β_0 = -0.5526 (p-value = 0.1419; SE = 0.3551) and β_1 = -22.4059 (p-value = 0.3272; SE = 22.0701). The study wise agreement analysis indicated that in 14 of the 16 comparisons, treatment was associated with the same effect on both outcomes and the percent concordance was 87.5%. In all of these studies, the treatment had no significant effect on either of the outcomes (all p-values > 0.05). In each of the 2 studies with discordant results, the difference between groups in the surrogate outcome was statistically significant (p-values for CIMT differences < 0.05), and the risk of stroke was not statistically significant. Qualitatively, the direction of treatment effect for each outcome was the same in these studies. In CAPTIVATE [88], subjects who received pactimibe had a significantly higher CIMT than placebo subjects and a non-significant, higher risk of stroke. The mean difference for the pactimibe vs placebo group in CIMT was 0.023 (95% CI: 0.00001, 0.0459) and the log odds ratio was 1.0895 (-2.1139, 4.2929). In the KAPS [98] study which compared pravastatin vs. placebo, subjects who receive the statin had a significantly lower CIMT and a non-significant, lower risk of stroke. The mean difference in CIMT for the pravastatin vs. placebo group was -0.0189

(95% CI: -0.0308, -0.0070) and the log odds ratio was -0.7068 (95% CI: -2.4144, 1.0009).

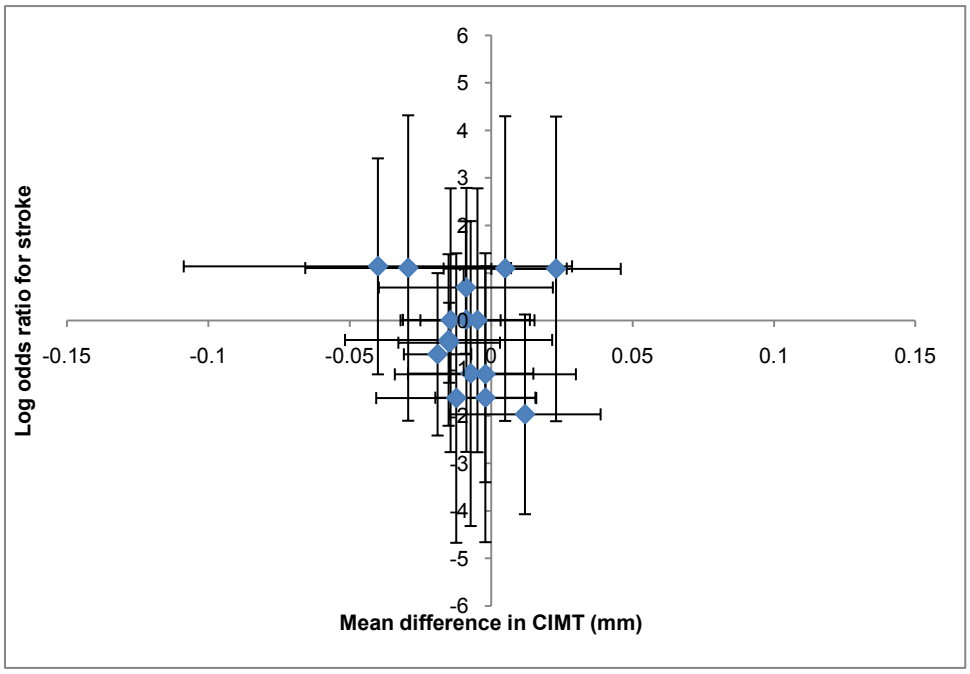

Figure 4.12: Graphical analysis for 12 month mean difference in CIMT and stroke at 2 or more years of follow-up

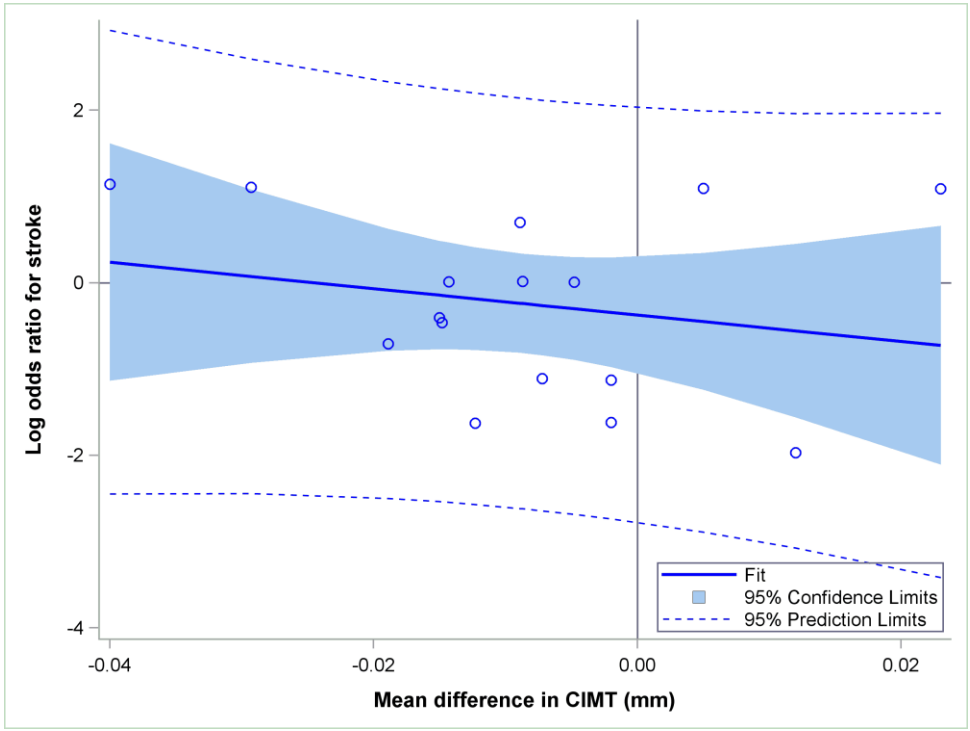

Figure 4.13: STE for 12 month mean difference in CIMT and stroke at 2 or more years of follow-up.

Studies with ≥ 4 years of follow-up

The evaluation of CIMT as a surrogate outcome for stroke at 4 or more years of follow-up was performed using data from 3 studies. Collectively, the results indicated that there was a statistically significant, negative association between the outcomes. There was evidence of a negative association in the graphical analysis (Figure 4.14). This observation was supported by the rank analysis in which Spearman's ρ= -1.0000 (p-value <0.0001). The weighted R^2 = 0.9956 (p-value = 0.0425) and the regression was significant. From the weighted regression analysis, β_0 = -1.4395 (p-value = 0.0368; SE = 0.0833) and β_1 = -65.0909 (p-value = 0.0425; SE = 4.3500). There was a trend towards a STE (Figure 4.15). Since the upper prediction band appeared to be approaching the positive x-axis, the graph for the STE indicated that above a certain threshold value for the CIMT mean difference, carotid thicknesses that are higher in the experimental arm relative to the control arm post treatment, may be able to predict a lower risk of stroke. This finding is in agreement with the negative association observed in prior analyses; however, it is inconclusive due to the small number of studies in the analysis. For the Korn fixed effects analysis [32], β_0 = -1.4394 (p-value = 0.3547; SE = 0.8968) and β_1 = -65.0909 (p-value = 0.3971; SE = 46.8326). The analysis for study-wise agreement indicated that in all studies, the result of the between group comparison on CIMT was the same as the result based on stroke. The percent concordance was 100%. In all 3 studies, no significant treatment effect was observed on either of the outcomes (all p-values >0.05).

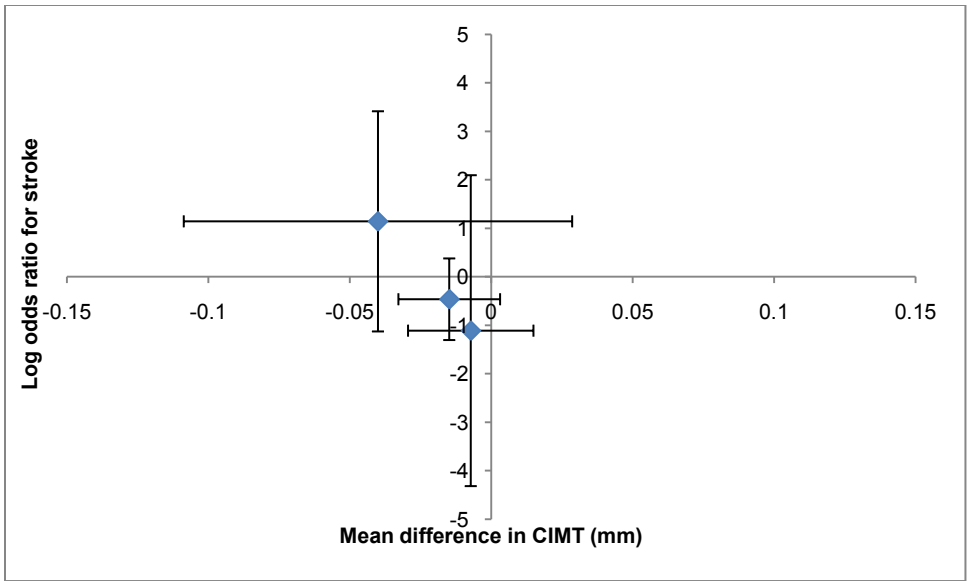

Figure 4.14: Graphical analysis for 12 month mean difference in CIMT and stroke at 4 or more years of follow-up

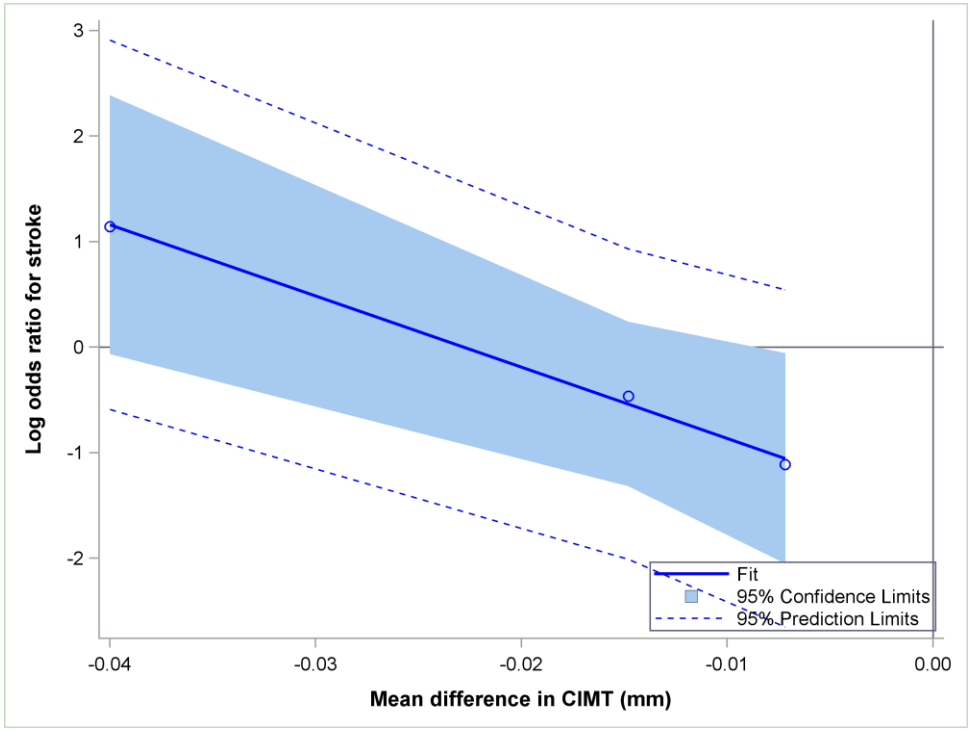

Figure 4.15: STE for 12 month mean difference in CIMT and stroke at 4 or more years of follow-up.

Table 4.7: Estimates of treatment effect for CIMT and stroke

Reference	Experimental drug class	Control drug class	Study dur-ation	Mean differ-ence (MD) for CIMT	LCL (MD)	UCL (MD)	SE (MD)	Odds ratio for all-cause mort-ality	LCL (OR)	UCL (OR)	Log odds ratio	LCL (LOR)	UCL (LOR)	SE (LOR)
TART (Hodis)	Glitazone	Placebo	2	-0.0150	-0.0516	0.0216	0.0187	0.6667	0.1098	4.0483	-0.4055	-2.2092	1.3983	0.9203
EPAT (Hodis)	Hormone therapy	Placebo	2	-0.0293	-0.0657	0.0071	0.0186	3.0271	0.1220	75.1197	1.1076	-2.1038	4.3191	1.6385
ENHANCE (Kastelein)	Statin + cholesterol absorption inhibitor	Statin	2	-0.0087	-0.0311	0.0137	0.0115	1.0169	0.0634	16.3206	0.0167	-2.7590	2.7924	1.4162
RADIANCE 2 (Bots)	Statin + CETP inhibitor	Statin	2	0.0050	-0.0168	0.0268	0.0111	2.9920	0.1215	73.6858	1.0960	-2.1079	4.2998	1.6346
RADIANCE 1 (Kastelein)	Statin + CETP inhibitor	Statin	2	-0.0048	-0.0250	0.0154	0.0103	1.0089	0.0629	16.1804	0.0089	-2.7661	2.7838	1.4158
CAPTIVATE (Meuwese)	ACAT inhibitor	Placebo	2	0.0230	0.0001	0.0459	0.0117	2.9729	0.1208	73.1800	1.0895	-2.1139	4.2929	1.6344
VEAPS (Hodis)	Vitamin E	Placebo	3	-0.0123	-0.0406	0.0160	0.0144	0.1966	0.0094	4.1252	-1.6265	-4.6701	1.4171	1.5529
SANDS (Howard)	Aggressive treatment	Standard treatment	3	-0.0020	-0.0340	0.0300	0.0163	0.3240	0.0335	3.1367	-1.1269	-3.3970	1.1432	1.1582
STARR (Lonn)	ACE inhibitor	Placebo	3	-0.0020	-0.0197	0.0157	0.0090	0.1980	0.0095	4.1327	-1.6193	-4.6575	1.4189	1.5501
STARR (Lonn)	Glitazone	Placebo	3	-0.0143	-0.0320	0.0034	0.0090	1.0099	0.0630	16.1779	0.0098	-2.7640	2.7836	1.4152
MIDAS (Borhani)	CCB	Thiazide diuretic	3	-0.0089	-0.0397	0.0219	0.0157	2.0092	0.4993	8.0848	0.6977	-0.6945	2.0900	0.7103
BCAPS (Hedblad)	Beta blocker	Placebo	3	0.0120	-0.0148	0.0388	0.0137	0.1396	0.0171	1.1399	-1.9691	-4.0692	0.1309	1.0715
KAPS (Salonen)	Statin	Placebo	3	-0.0189	-0.0308	-0.0070	0.0061	0.4932	0.0894	2.7207	-0.7068	-2.4144	1.0009	0.8712
VHAS (Zanchetti)	CCB	Thiazide diuretic	4	-0.0400	-0.1087	0.0287	0.0350	3.1357	0.3237	30.3739	1.1429	-1.1278	3.4136	1.1585
ELSA (Zanchetti)	CCB	Beta blocker	4	-0.0148	-0.0328	0.0032	0.0092	0.6291	0.2712	1.4592	-0.4635	-1.3048	0.3779	0.4293
CHICAGO (Mazzone)	Glitazone	Sulfonylurea	6	-0.0072	-0.0294	0.0150	0.0113	0.3290	0.0133	8.1189	-1.1117	-4.3176	2.0942	1.6357

Cardiovascular mortality

Studies with ≥ 2 years of follow-up

Table 4.8 contains the study specific 12 month mean differences in CIMT and log odds ratios for cardiovascular mortality, along with the 95% CIs and standard errors.

For the evaluation of CIMT as a surrogate outcome for cardiovascular mortality at 2+ years of follow-up, data was obtained from 17 studies. A total of 20 pair-wise comparisons were used for most of the analyses because the FAST and STARR trials had multiple treatment arms. For the Korn et al [32] fixed effects analysis, 19 pair-wise comparisons were used. Overall, there was a statistically significant, positive association between CIMT and cardiovascular mortality. From the graphical analysis (Figure 4.16), there was evidence of a positive relationship between the outcomes. Spearman's ρ=0.5972 (p= 0.0054) and the correlation was significant. The weighted R^2 = 0.2063 (p=0.0442) and the regression was significant. In the weighted analysis, β_0 = -0.1692 (p-value = 0.2985; SE = 0.1580) and β_1 = 3.5805 (p-value = 0.0442; SE = 1.6554). There was a trend towards a STE (Figure 4.17). For the Korn fixed effects analysis [32], β_0 = -0.1631(p-value = 0.5628; SE = 0.2764) and β_1 = 4.0907 (p-value not estimable; SE = 2.4812). The results of the study-wise agreement analysis indicated that in 14 of the 20 studies, the results obtained for the effect of treatment on CIMT were in agreement with the results based on the clinical outcome. The percent concordance was 70%. In each of these studies, there was no significant treatment effect on either of the outcomes. For the 6 studies with discordant results, there was a statistically significant difference between groups in CIMT (p-values for CIMT differences < 0.05) but there was no significant difference in the risk for cardiovascular mortality. In 5 of these studies the same direction of treatment effect was observed for each outcome and in most of these 5 studies, as compared to placebo, the experimental arm was associated with a significantly lower CIMT and non-significant, lower risk of cardiovascular mortality. The exception was the CAPTIVATE [88] trial in which the placebo group had better outcomes than the pactimibe group. The mean difference in CIMT for pactimibe vs. placebo was 0.023 (95% CI: 0.0001, 0.0459) and the log odds ratio for cardiovascular mortality was 2.9800 (95% CI: 0.3087, 28.7562). The remaining one study with discordant results was the ASAP [65] trial

in which subjects receiving atorvastatin had a significantly lower CIMT than subjects receiving simvastatin and a non-significant, higher risk of cardiovascular mortality. The mean difference in CIMT for atorvastatin versus simvastatin was -0.04 (95% CI: -0.0741, -0.0059) and the log odds ratio was 0.0310 (95% CI: -2.7495, 2.8114).

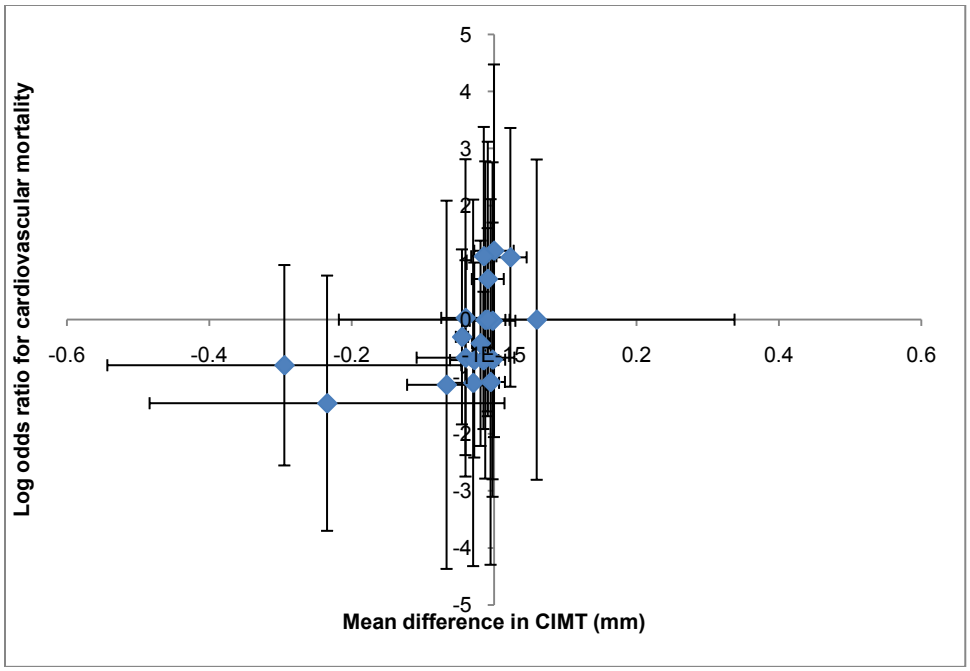

Figure 4.16: Graphical analysis for 12 month mean difference in CIMT and cardiovascular mortality at 2 or more years of follow-up.

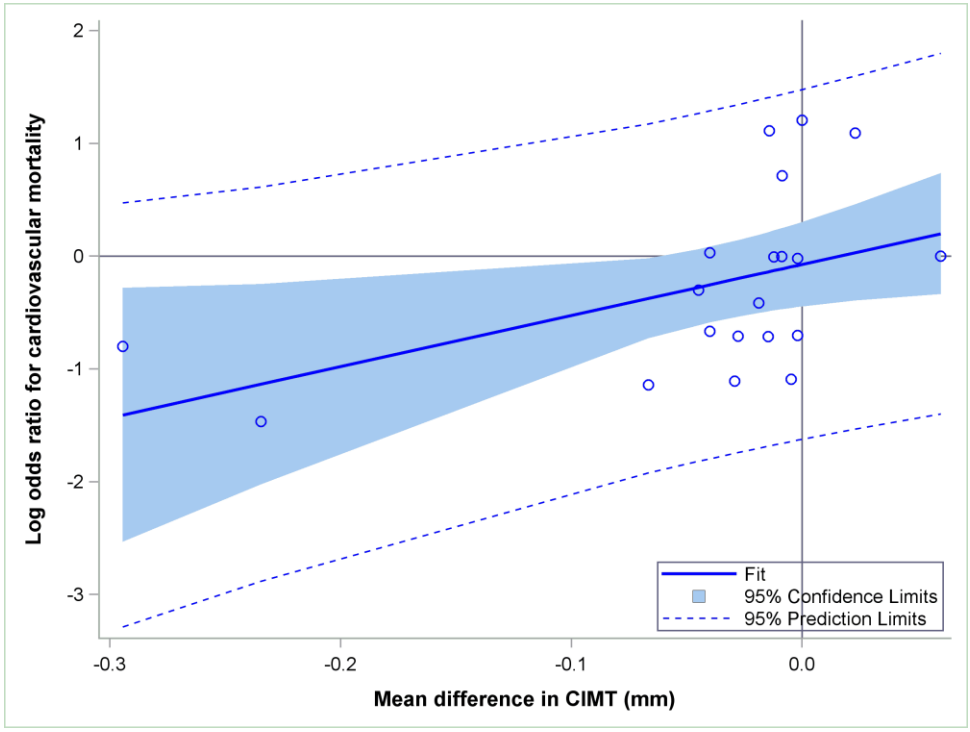

Figure 4.17: STE for 12 month mean difference in CIMT and cardiovascular mortality at 2 or more years of follow-up.

Studies with ≥ 4 years of follow-up

In the evaluation of CIMT as a surrogate outcome for cardiovascular mortality at 4 or more years of follow-up, data was obtained from 4 studies. Overall, there was a negative association between the outcomes but the significance of this observation was inconsistent across results. In the graphical analysis (Figure 4.18) a negative association was observed. between CIMT and cardiovascular mortality. In the rank correlation analysis, there was a statistically significant, negative correlation between the outcomes. The Spearman's $\rho = -1$ ($p < 0.0001$). A negative, non-significant association was observed in the weighted analysis. The weighted $R^2 = 0.5720$ (p-value = 0.2437, $\beta_0 = -0.9024$ (p-value = 0.0447; SE = 0.1974) and $\beta_1 = -10.1374$ (p-value = 0.2437; SE = 6.2011). There was no STE. (Figure 4.19). The results of the fixed effects analysis were not significant. For the Korn fixed effects analysis[32], $\beta_0 = -0.9024$ (p-value = 0.4437; SE = 0.9532) and $\beta_1 = -10.1374$ (p-value = 0.7672; SE = 29.9512). The study wise agreement analysis indicated that the results for both outcomes were in agreement for 3 of the 4 studies. In these studies, no significant treatment effect was observed for either outcome (all p-values > 0.05). Discordant results were observed for the CLAS [81] study but qualitatively, the direction of treatment effect was the same for both outcomes. In the trial, the intervention group had a significantly lower CIMT (p-value for CIMT difference < 0.05) and a non-significant, lower risk of cardiovascular mortality. The mean difference in CIMT for the experimental vs. control group was -0.045 (95% CI: -0.0536, -0.0364) and the log odds ratio was -0.3021 (95% CI: -1.8355, 1.2314).

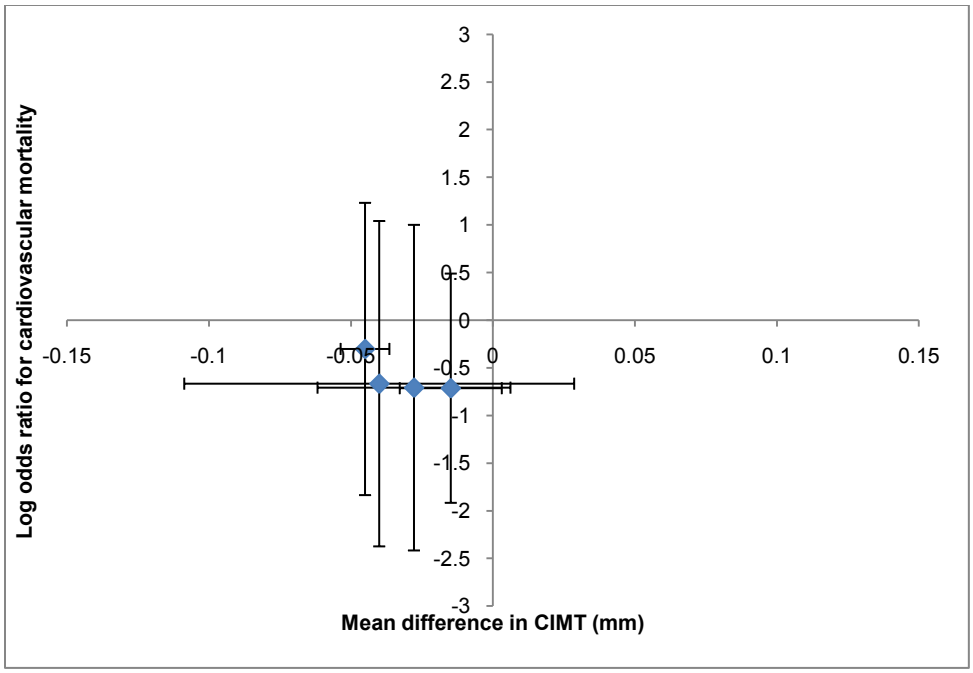

Figure 4.18: Graphical analysis for 12 month mean difference in CIMT and cardiovascular mortality at 4 or more years of follow-up.

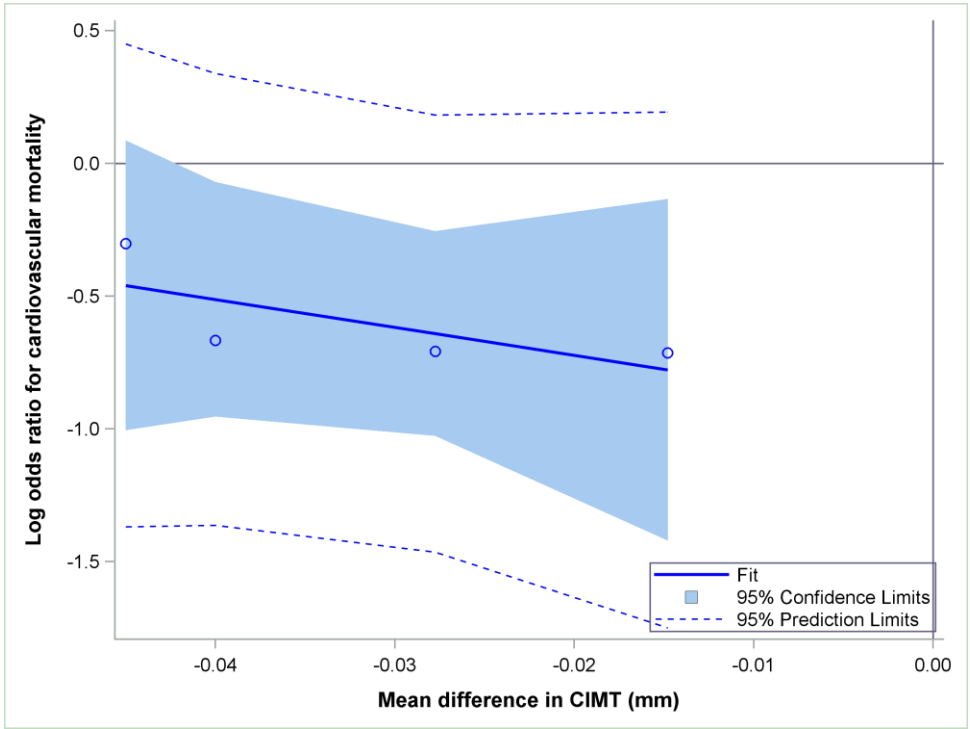

Figure 4.19: STE for 12 month mean difference in CIMT and cardiovascular mortality at 4 or more years of follow-up.

Table 4.8: Estimates of treatment effect for CIMT and cardiovascular mortality

Reference	Experimental drug class	Control drug class	Study dur-ation	Mean differ-ence (MD) for CIMT	LCL (MD)	UCL (MD)	SE (MD)	Odds ratio for all-cause mort-ality	LCL (OR)	UCL (OR)	Log odds ratio	LCL (LOR)	UCL (LOR)	SE (LOR)
EPAT (Hodis)	Hormone therapy	Placebo	2	-0.0293	-0.0657	0.0071	0.0186	0.3303	0.0133	8.1976	-1.1076	-4.3191	2.1038	1.6385
ASAP (Smilde)	Statin (atorvastatin)	Statin (simvastatin)	2	-0.0400	-0.0741	-0.0059	0.0174	1.0314	0.0640	16.6330	0.0310	-2.7495	2.8114	1.4186
FAST (Sawayama)	Antihypertensive	Placebo	2	-0.2345	-0.4838	0.0148	0.1272	0.2313	0.0247	2.1646	-1.4643	-3.7007	0.7722	1.1411
FAST (Sawayama)	Statin	Placebo	2	-0.2945	-0.5431	-0.0459	0.1268	0.4500	0.0777	2.6055	-0.7985	-2.5546	0.9576	0.8960
FAST (Sawayama)	Antihypertensive	Statin	2	0.0600	-0.2179	0.3379	0.1418	1.0000	0.0604	16.5485	0.0000	-2.8063	2.8063	1.4318
ENHANCE (Kastelein)	Statin + cholesterol absorption inhibitor	Statin	2	-0.0087	-0.0311	0.0137	0.0115	2.0394	0.1841	22.5929	0.7127	-1.6923	3.1176	1.2270
RADIANCE 1 (Kastelein)	Statin + CETP inhibitor	Statin	2	-0.0048	-0.0250	0.0154	0.0103	0.3356	0.0136	8.2593	-1.0920	-4.2953	2.1113	1.6343
ATIC (Nanayakkara)	Statin + vitamin E + vitamin B	Placebo	2	-0.0666	-0.1220	-0.0112	0.0283	0.3193	0.0127	8.0426	-1.1416	-4.3680	2.0848	1.6461
CAPTIVATE (Meuwese)	ACAT inhibitor	Placebo	2	0.0230	0.0001	0.0459	0.0117	2.9795	0.3087	28.7561	1.0918	-1.1753	3.3589	1.1567
Van Vonderen	NRTI	cART	2	0.0000	-0.0277	0.0277	0.0141	3.3429	0.1277	87.5240	1.2068	-2.0583	4.4719	1.6659
VEAPS (Hodis)	Vitamin E	Placebo	3	-0.0123	-0.0406	0.0160	0.0144	0.9943	0.0617	16.0233	-0.0057	-2.7854	2.7740	1.4182
SANDS (Howard)	Aggressive treatment	Standard treatment	3	-0.0020	-0.0340	0.0300	0.0163	0.9801	0.0610	15.7574	-0.0201	-2.7976	2.7573	1.4171
STARR (Lonn)	Ace inhibitor	Placebo	3	-0.0020	-0.0197	0.0157	0.0090	0.4958	0.0449	5.4803	-0.7016	-3.1043	1.7012	1.2259
STARR (Lonn)	Glitazone	Placebo	3	-0.0143	-0.0320	0.0034	0.0090	3.0382	0.3153	29.2793	1.1113	-1.1543	3.3769	1.1559
MIDAS (Borhani)	CCB	Thiazide diuretic	3	-0.0089	-0.0397	0.0219	0.0157	0.9977	0.2003	4.9705	-0.0023	-1.6081	1.6035	0.8193
KAPS (Byington)	Statin	Placebo	3	-0.0189	-0.0308	-0.0070	0.0061	0.6607	0.1093	3.9924	-0.4145	-2.2134	1.3844	0.9178
VHAS (Zanchetti)	CCB	Thiazide diuretic	4	-0.0400	-0.1087	0.0287	0.0350	0.5135	0.0931	2.8319	-0.6665	-2.3739	1.0410	0.8711
MAVET (Magliano)	Vitamin E	Placebo	4	-0.0277	-0.0617	0.0062	0.0173	0.4926	0.0892	2.7198	-0.7080	-2.4166	1.0006	0.8717
ELSA (Zanchetti)	CCB	Beta blocker	4	-0.0148	-0.0328	0.0032	0.0092	0.4898	0.1471	1.6310	-0.7138	-1.9168	0.4892	0.6138
CLAS (Hodis)	Colestipol + niacin + diet therapy	Placebo + diet	10	-0.0450	-0.0536	-0.0364	0.0044	0.7393	0.1595	3.4260	-0.3021	-1.8355	1.2314	0.7824

4.4.3.2 Drug class specific comparisons

All-cause mortality

Hormone therapy versus placebo

To evaluate CIMT as a surrogate outcome for all-cause mortality in trials of hormone therapy vs. placebo, data was used from 2 studies. Because one study, OPAL [83], consisted of 3 treatment groups, 3 pairwise comparisons were used in the analysis. Overall, the results indicated that there was no statistically significant relationship between CIMT and all-cause mortality. Qualitatively, there was a negative trend between the outcomes. The graphical analysis (Figure 4.20) showed a negative relationship between CIMT and all-cause mortality. This observation remains speculative based on the large imprecision in the estimates of the log odds ratios. Spearman's $\rho = -0.5000$ (p-value = 0.6667) and the weighted R^2 was 0.9633 (p-value = 0.1228). From the regression, $\beta_0 = -1.5246$ (p-value = 0.0252; SE = 0.0605) and $\beta_1 = -13.8225$ (p-value = 0.1228; SE = 2.6991). A STE=0.007 was observed (Figure 4.21). At the mean difference= 0.007, the upper prediction band dipped below the x-axis. At this mean difference value, from left-to-right on the graph mean differences were successively more positive and log odds ratios along the upper prediction band were successively more negative before reaching a constant value. The STE indicated that carotid thicknesses that are at least 0.007 mm higher in the experimental arm relative to the control arm post-therapy predict a lower risk of all-cause mortality. This finding is consistent with the negative trend observed in prior analyses, but remains inconclusive due to the small number of studies in the analysis. For the Korn fixed effects analysis [32], $\beta_0 = -1.4118$ (p-value = 0.3810; SE = 0.9626) and $\beta_1 = -9.9726$ (p-value = not estimable; SE = 42.5430). The study wise agreement analysis indicated that the results based on each outcome were in agreement in one study, and the percent concordance was 33.33%. In this study, the treatment had no effect on either of the outcomes (all p-values > 0.05). Discordant results were obtained for both comparisons of the OPAL [83] study. In each comparison, the treatment group was associated a significantly higher CIMT (p-values for CIMT differences < 0.05) and a non-significant, lower risk of mortality. The mean difference in CIMT for tibolone vs placebo was 0.0215

(95% CI: 0.0084, 0.0345) and the log odds ratio for all-cause mortality was -1.7541 (95% CI: -4.9618, 1.4537). For cee/mpa vs. placebo, the mean difference in CIMT was 0.0138 (95% CI: 0.0007, 0.0268) and the log odds ratio was -1.7937 (95% CI: -5.0014, 1.4140).

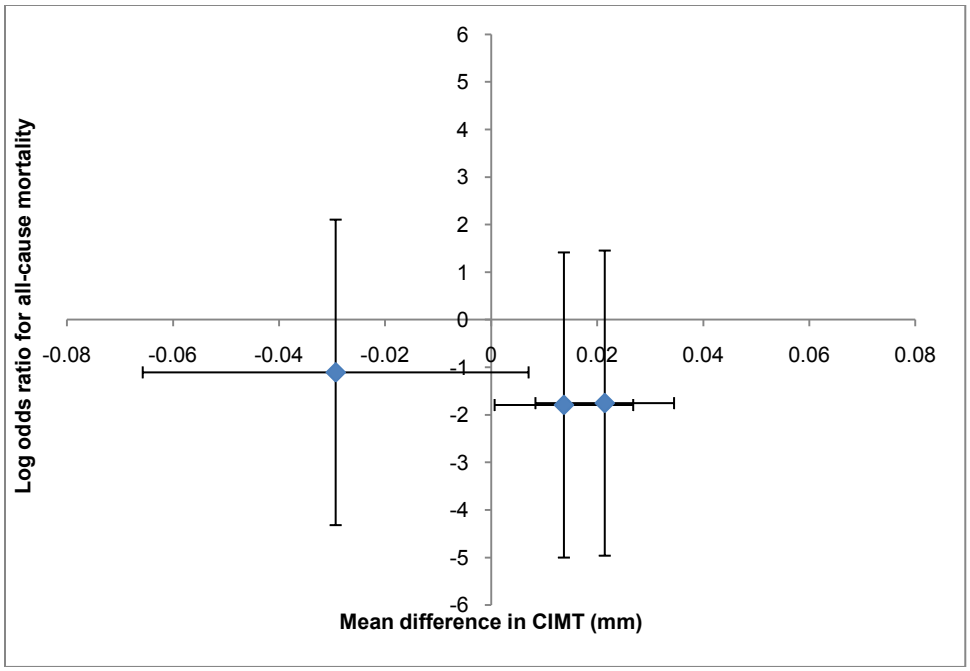

Figure 4.20: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality for hormone therapy versus placebo.

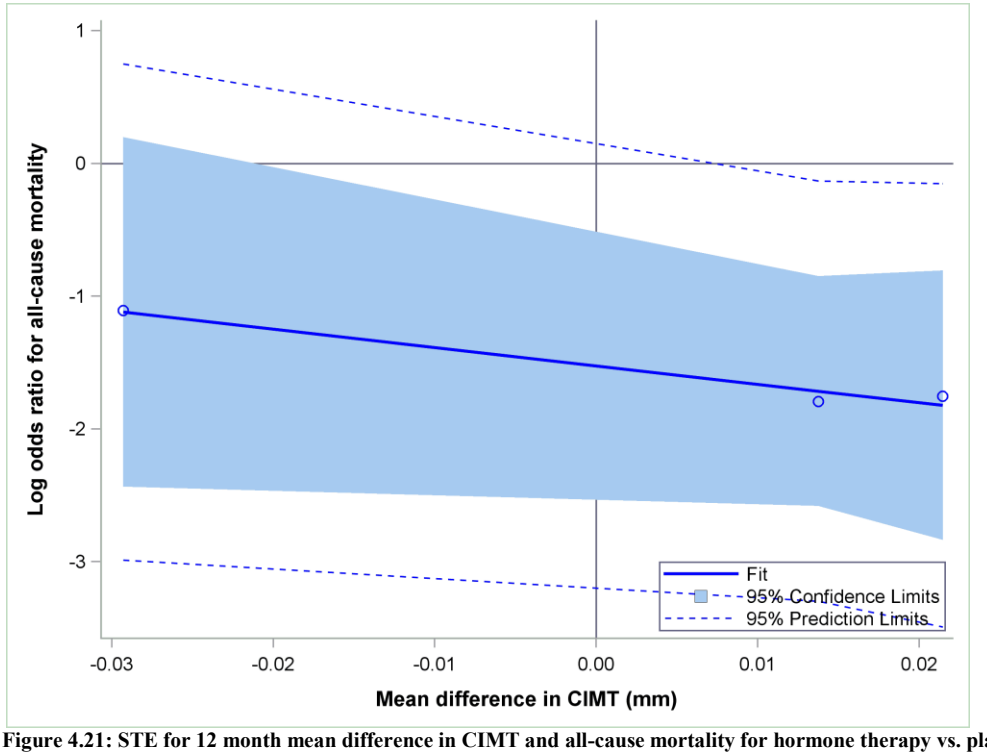

Figure 4.21: STE for 12 month mean difference in CIMT and all-cause mortality for hormone therapy vs. placebo.

Statin vs. placebo

To evaluate CIMT as a surrogate outcome for all-cause mortality in trials of statins vs. placebo, data was used from 4 studies. Overall, there was no statistically significant association between CIMT and all-cause mortality. Qualitatively, the results indicated that there was a positive relationship between the outcomes. No association was observed between the outcomes in the graphical analysis (Figure 4.22). Spearman's $\rho = 0.4000$ (p-value = 0.6000) and from the weighted regression analysis, $R^2 = 0.0018$ (p-value = 0.9570). The $\beta_0 = -0.3530$ (p-value = 0.1535; SE = 0.1570) and $\beta_1 = 0.0599$ (p-value = 0.9570; SE = 0.9859). No STE was observed (Figure 4.23). For the Korn fixed effects analysis [32], $\beta_0 = -0.3352$ (p-value = 0.5990; SE = 0.5416) and $\beta_1 = 1.1345$ (p-value = 0.7229; SE = 2.7820). There was agreement in 1 study for trial results based on the surrogate outcome versus the clinical outcome and the percent concordance was 25%. In the study, there was no significant treatment effect observed for either of the outcomes. In the remaining three studies, a statistically significant treatment effect was observed for the surrogate outcome (p-values for CIMT differences < 0.05), and the log odds ratios for all-cause mortality was not significant. In each of these studies, the same direction of treatment effect was observed for both outcomes, and was in favour of the statin group.

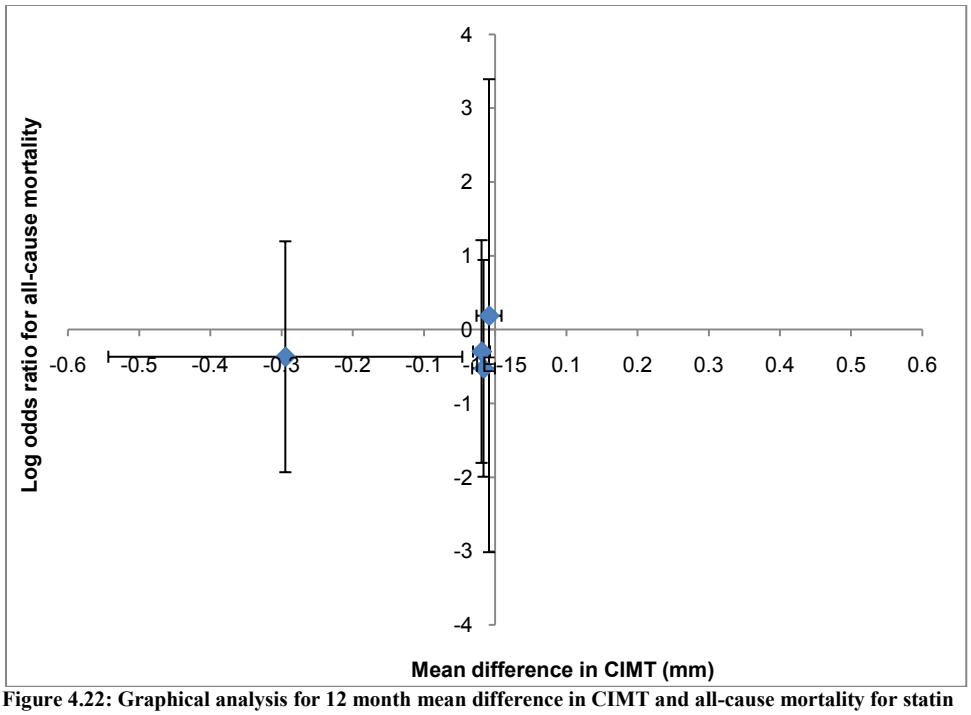

Figure 4.22: Graphical analysis for 12 month mean difference in CIMT and all-cause mortality for statin versus placebo.

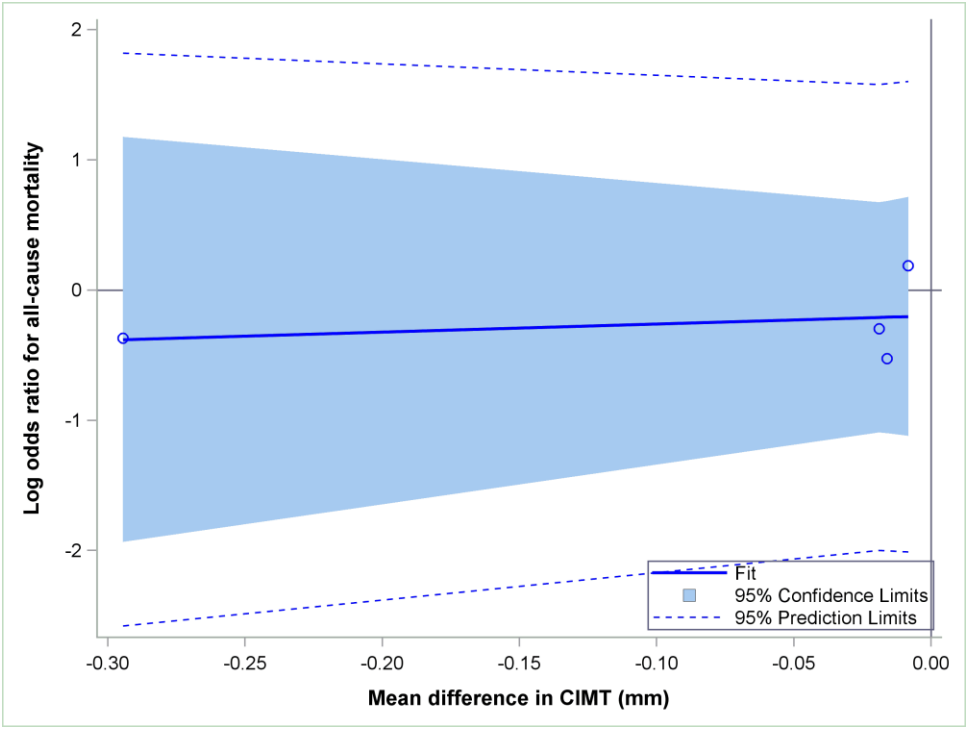

Figure 4.23: STE for 12 month mean difference in CIMT and all-cause mortality for statin vs. placebo.

Myocardial infarction

Statin versus placebo

To evaluate CIMT as a surrogate outcome for myocardial infarction in trials comparing statins versus placebo, data was available from 4 studies. No statistically significant association was observed between the outcomes. Qualitatively, the direction of the association between CIMT and myocardial infarction was inconsistent across the results. From the graphical analysis (Figure 4.24), no pattern could be identified to describe the relationship between the outcomes. The rank correlation analysis indicated that there was a non-significant, positive correlation. The Spearman's $\rho = 0.4000$ (p-value = 0.6000). The results of the weighted regression showed a non-significant, negative association between the outcomes. The weighted $R^2 = 0.0100$ (p-value = 0.8998). The $\beta_0 = -1.1753$ (p-value = 0.1031; SE = 0.4098) and $\beta_1 = -0.3989$ (p-value = 0.8998; SE = 2.8006). No STE was observed (Figure 4.25). A negative, non-significant association was also observed in the results for the fixed effects analysis. For the Korn fixed effects analysis [32], $\beta_0 = -1.1706$ (p-value = 0.1555; SE = 0.5249) and $\beta_1 = -0.1158$ (p-value = 0.9709; SE = 2.8162). For the study wise agreement analysis, two studies showed concordant results between outcomes and the percent concordance was 50%. In one of these studies, PLAC-II [62], a statistically significant treatment effect was observed for both outcomes and favoured the statin group. The mean difference in CIMT for the statin vs. placebo group was -0.0161 (95% CI: -0.0320, -0.0002) and the log odds ratio was -1.7102 (95% CI: -3.2645, -0.1560). In the other study, METEOR [73], no significant treatment effect was observed for either of the outcomes. Discordant results were observed in FAST [84] and KAPS [98]. In both of these studies, a significant treatment effect was observed on the surrogate outcome (p-values for CIMT differences < 0.05) but not on the clinical outcome. The direction of treatment effect was the same for both outcomes and in favour of the statin group. Collectively, in 3 of the 4 studies, the treatment effect was in the same direction for both outcomes and in favor of the statin group.

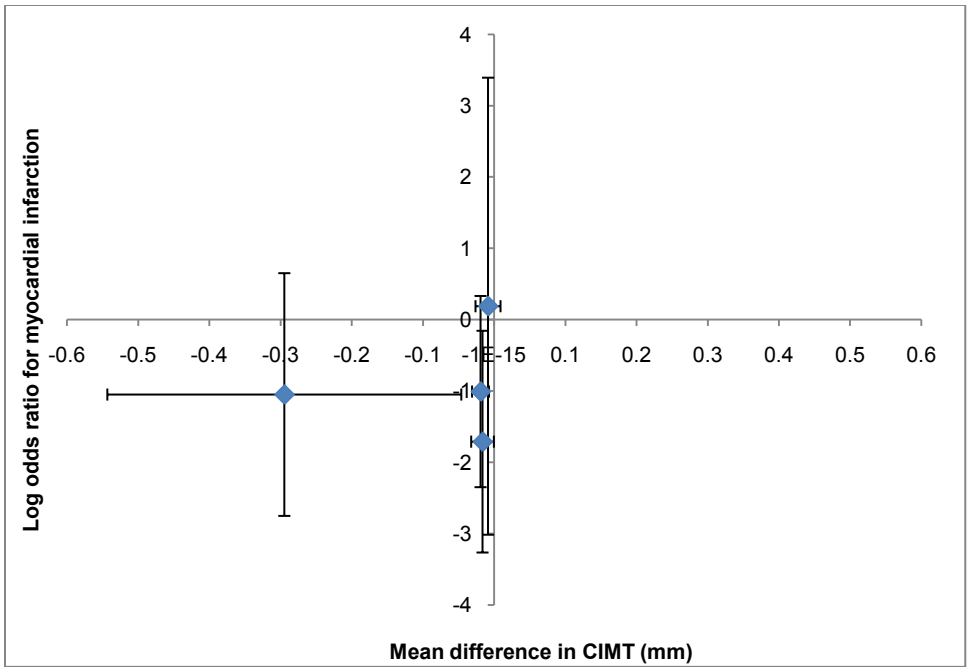

Figure 4.24: Graphical analysis for 12 month mean difference in CIMT and myocardial infarction for statin versus placebo.

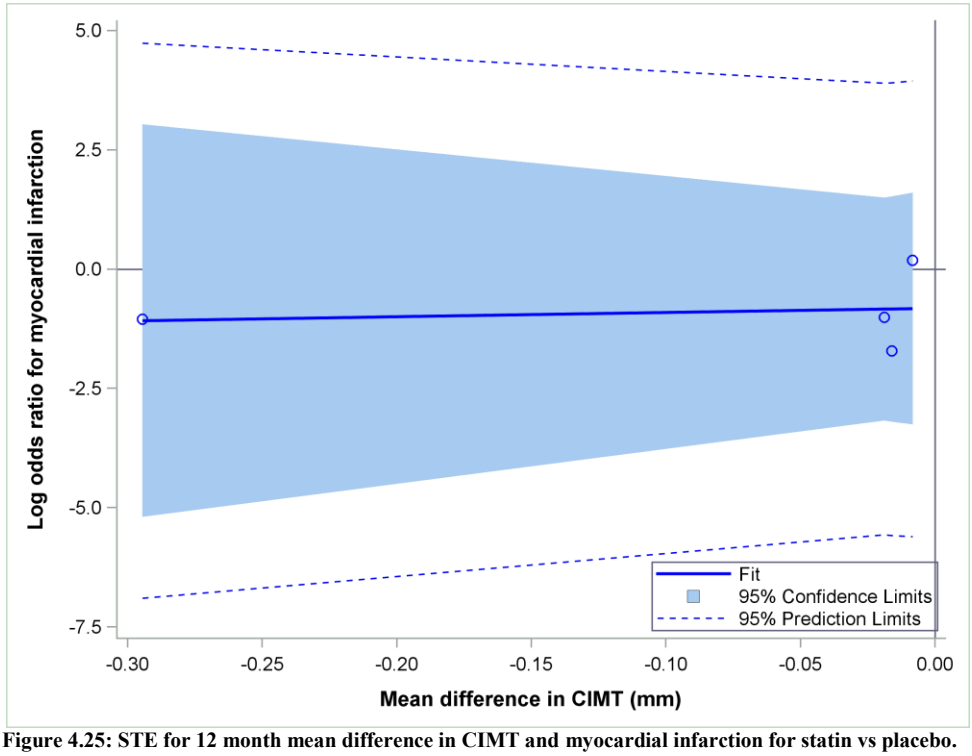

Figure 4.25: STE for 12 month mean difference in CIMT and myocardial infarction for statin vs placebo.

4.4.3.2 Sensitivity analysis

Diagnostics for the weighted regression were only performed for the models that included clinical outcome data at 2 or more years of follow-up. Diagnostics were not performed for the models that were based on clinical outcome data at 4 or more years of follow-up because these analyses were based on a small number of studies. It was decided that a meaningful evaluation of the regression assumptions could not be performed with the small amount of data available in these analyses.

With the exception of linearity, there were no serious violations of the regression assumptions. The kernel dentisty plots showed an approximately normal distribution. There was some evidence of departures from normality in the tail regions of the normal quantile-quantile plots; because regression analysis is robust to violations of the normality assumption, these deviations were not considered serious. The plots of the residuals versus fitted values showed an approximately equal vertical range of residuals across the fitted values which suggested that there was no evidence of violation for the assumption of homogeneity and independence.

The LOWESS smooth applied to the residual versus predictor plot showed departures from linearity for each of the models evaluated. Four transformations were applied to the 12 month mean differences in CIMT, including the log, square, square root, and exponential transformation. For each of these transformations, there was no significant improvement in the linearity and the overall results did not change. Table 4.9 summarizes the results of the weighted regression analyses based on applying transformations. Because the overall results did not change, the transformations were not applied to the Korn et al models [32].

Table 4.9: Results of applying transformations for the weighted regression for studies with 2 or more years of follow-up

Transform-ation	All-cause mortality			Myocardial infarction			Stroke			Cardiovascular mortality		
	R^2	β_1	p-value	R^2	β_1	p-value	R^2	β_1	p-value	R^2	β_1	p-value
None	0.0281	1.4614	0.4129	0.1399	4.8558	0.0598	0.0948	-22.4059	0.2460	0.2063	3.5805	0.0442
Log (CIMT)	0.0273	1.6583	0.4201	0.1347	5.5480	0.0651	0.0969	-23.5727	0.2405	0.1967	4.0561	0.0501
$(CIMT)^{1/2}$	0.0277	3.1174	0.4163	0.1373	10.3911	0.0624	0.0959	-45.9670	0.2432	0.2015	7.6285	0.0471
$(CIMT)^2$	0.0288	0.6373	0.4076	0.1447	2.1075	0.0552	0.0927	-10.6424	0.2516	0.2159	1.5690	0.0390
e^{CIMT}	0.0288	0.4617	0.4076	0.1455	1.5285	0.0545	0.0926	-7.8194	0.2519	0.2177	1.1404	0.0381

4.5 Discussion

Below is a summary of the main results obtained from the application of the various statistical procedures to data obtained from the systematic review of imaging trials:

All-cause mortality

- At 2 or more years of follow-up, there was no statistically significant association between CIMT and all-cause mortality. Qualitatively, a positive trend was observed between the outcomes.
- At 4 or more years of follow-up, there was no statistically significant association between CIMT and all-cause mortality. Qualitatively, a positive trend was observed between the outcomes and the magnitude of the relationship between CIMT and all-cause mortality was higher as compared to the results obtained for 2 or more years of follow-up.
- For the comparison of hormone therapy vs. placebo, there was no statistically significant association between CIMT and all-cause mortality. Qualitatively, a negative trend was observed between the outcomes.
- For the comparison of statins vs. placebo, there was no statistically significant association between CIMT and all-cause mortality. Qualitatively, there was a positive association between the outcomes.

Myocardial infarction

- At two or more years of follow-up, there was a statistically significant, positive association between CIMT and myocardial infarction.
- At 4 or more years of follow-up, there was no statistically significant association between CIMT and myocardial infarction. Qualitatively, there was a positive relationship between the outcomes.
- For the comparison of statin vs. placebo, there was no statistically significant association between the CIMT and myocardial infarction. Qualitatively, the direction of the association between the outcomes was inconsistent across the results from the varying analysis methods.

Stroke

- At 2 or more years of follow-up, there was no statistically significant association between CIMT and stroke. Qualitatively, a negative trend was observed between the outcomes.
- At 4 or more years of follow-up, there was a statistically significant, negative association between CIMT and stroke.

Cardiovascular mortality

- At 2 or more years of follow-up, there was a statistically significant, positive association between CIMT and cardiovascular mortality.
- At 4 or more years of follow-up, the significance of the relationship between CIMT and cardiovascular mortality was inconsistent across the various analysis methods. Qualitatively, a negative trend was observed between the outcomes.

Overall, the results indicated that CIMT is a good surrogate outcome for myocardial infarction and cardiovascular mortality at 2 or more years of follow-up. For the relationship between CIMT and each of these clinical outcomes, the results showed a statistically significant association which may be clinically relevant. There was no evidence to suggest that CIMT was a good surrogate outcome for cardiovascular mortality. The results suggested that CIMT was a poor surrogate outcome for stroke. The negative association between CIMT and stroke was statistically significant in studies with 4 or more years of follow-up, and a negative, non significant trend was also noted for the analyses for 2 or more years of follow-up. Because the analysis in studies with 4 or more years of follow-up consisted of a small sample size, there is a degree of speculation in the findings for the relationship between CIMT and stroke. For the same reason, there was insufficient information to suggest that CIMT was a good surrogate outcome for all-cause mortality, myocardial infarction and cardiovascular mortality occurring at 4 or more years of follow-up.

It was noted that the percent concordance was generally high and did not reflect the results obtained from the remaining analysis methods. For this reason, the percent

concordance values were not considered in determining the consistency of results across the various methods. Also, in this review, the high percent concordance was based primarily on studies in which there was no significant treatment effect for either outcome. In settings such as this, where the majority of studies show null effects on both outcomes, the usefulness of the percent concordance is significantly reduced.

The graphical analysis method was a useful exploratory tool to develop an idea about the type of trend that may exist between the outcomes[29,40]. The inclusion of the vertical and horizontal confidence intervals was a useful strategy to develop an immediate sense of the precision in the estimates for each outcome. Despite these advantages, the graphical analysis approach did not provide specific quantitative criteria, such as p-values, to judge the strength of the CIMT[29,40]. For these reasons, the observed trends remained speculative in each of the graphs. Application of this method indicates that its greatest usefulness is in the setting where both estimates are measured very precisely and a consistent positive association is observed. This scenario is unlikely to happen in reality, and some of the other models that were applied did show a significant association between the outcomes.

Although Spearman's rank correlation analysis is usually used as a supportive procedure alongside other surrogate validation methods, its application to this dataset indicated that it may deserve greater consideration. Because the procedure does not rest on assumptions about the type of relationship (e.g. linear) that should exist between the outcomes, it has the potential to detect a clinically relevant and significant, positive association between the outcomes even when the relationship is not linear.

The primary advantage of determining the STE was that a STE was likely to be identified when the R^2 was very high. This finding is in agreement with a simulation study conducted by Lassere et al[33].

Application of the Korn et al model suggested that its usefulness was limited in evaluationg CIMT as a surrogate outcome. Application of the the Korn et al model required the estimation of a large number of parameters, given the size of the dataset. The results from the Korn et al model were not significicant and in some cases, the p-values were not estimable.

In general, the results obtained across different statistical procedures were in agreement with each other for any given evaluation of CIMT as a surrogate outcome. A few exceptions were noted, and were likely due to the small number of studies included in these analyses. One exception was found in the results obtained for the relationship between CIMT and cardiovascular mortality at 4 or more years of follow-up. The Spearman rank correlation analysis indicated that there was a statistically significant and negative correlation between the outcomes. The results of the weighted regression and Korn et al model were not statistically significant. Another exception was the results obtained for the relationship between CIMT and myocardial infarction in statins versus placebo. No statistically significant association was observed across the different procedures; however, qualitatively the trend (positive or negative) was inconsistent. The Spearman rank analysis showed a positive correlation between the outcomes while the weighted regression and Korn et al model showed a negative association between outcomes.

The results for the analyses not limited to specific drug classes indicated that generally, the trend (positive or negative) in the relationship between CIMT and the clinical outcome under consideration was the same at 2 or more versus 4 or more years of follow-up. These trends were not replicated in the analyses performed for specific drug class comparisons. Considered together, the results indicated that the trends observed in non-drug class specific analyses were likely only true for some drug classes and that drug class specific analyses would need to be performed in order to determine the applicability of the findings for specific drug classes.

The lack of a significant association between CIMT and all-cause mortality at 2 or more years of follow-up was somewhat surprising in the context of significant findings for the relationship between CIMT and cardiovascular mortality. It was expected that the results for the relationship between CIMT and all-cause mortality would reflect the observed increase in the log odds ratio for cardiovascular mortality with increasing mean differences in CIMT.

Inspection of the data indicated that for 2 of the 3 studies which were only included in the analysis for all-cause mortality and which also reported reasons for death, the groups with a lower CIMT had the only deaths which occurred in the studies and all were of non-cardiovascular causes. These results likely had the effect of reducing the correlation between CIMT and all-cause mortality, which would decrease the likelihood of finding a statistically significant finding. In the METEOR [73] trial, the rosuvastatin group had a lower CIMT than the placebo group and the only death occurred in the rosuvastatin group. The reason for the death was Creutzfeldt-Jakob disease. In the CHICAGO [95] trial, the pioglitazone group had a lower CIMT than the glimepiride group and the only death in the study, which was reported as being due to pancreatic cancer. RADIANCE 2 [92] was the only other study which was used in the all-cause mortality analysis and which also reported reasons for death. In this study, the statin plus torcetrapib group had a larger follow-up CIMT than the statin only group and there was one death in each group. One subject in the experimental group died of pneumonia and one subject in the control group died in a car accident. It is likely that the results of this study further contributed to a reduction in the correlation between CIMT and all-cause mortality, and a subsequent reduction in the likelihood of finding a significant association between the outcomes.

There were 7 additional studies (8 comparisons) that were included in the analysis for all-cause mortality but which did not report reasons for death. Interpretation of the results for these studies remains uncertain without information about the reasons for mortality. It is plausible that the interpretation given to the results for METEOR [73] and CHICAGO [95] may also apply to those studies in which one death was reported and occurred in the group with a lower CIMT. In the ELVERA [99] trial, there was one death overall and it occurred in the amlodipine group; this group also had a lower CIMT than the lisinopril group.

Another possibility is that there is no significant association between CIMT and both of all-cause mortality and cardiovascular mortality in certain drug classes that were included in the analysis for all-cause mortality, but under represented in the analysis for cardiovascular mortality. The inclusion or greater representation of these drug classes in the analysis for CIMT and cardiovascular mortality would have had the effect of reducing

the observed correlation and possibly the significance of the findings. Evidence to support this interpretation may be provided by the comparison of hormone therapy versus placebo. In the OPAL [83] study, which was included in the analysis for all-cause mortality, two subjects died in the placebo group and this group had a lower CIMT than each of the two hormone therapy groups. No subjects died in the hormone therapy groups. The Women's Health Initiative (WHI) [121] was a large randomized outcome trial comparing hormone therapy versus placebo. The results indicated that hormone therapy was not significantly associated with cardiovascular mortality or all-cause mortality. In that trial, the 3 year follow-up comparison of combined continuous hormone therapy versus placebo indicated that the hazard ratio for all-cause mortality was 1.22 (95% CI: 0.70, 1.37). For cardiovascular mortality, the hazard ratio was 1.18 (95% CI: 0.47, 2.98). Similarly, a systematic review of hormone therapy versus placebo suggests that hormone therapy is not significantly associated with either all-cause mortality or cardiovascular mortality [122]. The review compared estrogen only, combined continuous or combined sequential hormone therapy versus placebo in separate analyses for women with or without cardiovascular disease at a variety of follow-up durations and the same result was obtained in all analyses.

Consideration to the lack of significant findings in the comparison between CIMT and all-cause mortality suggests that the observed results may be due to a combination of reasons- some drug class comparisons may have shown a spurious negative relationship between the outcomes as a result of a higher number of non-cardiovascular disease deaths in the group with a lower CIMT. In other drug classes that were under represented in the analysis for cardiovascular mortality, there may have been no association between CIMT and both of all-cause and cardiovascular mortality.

The significant negative association observed between CIMT and stroke at 4 or more years of follow-up was unexpected. When the results of the 3 studies used in this analysis were considered individually, it was noted that in two of the studies, there was a negative mean difference in CIMT and a negative log odds ratio for stroke. The exception was the VHAS [93] study, which showed a negative mean difference in CIMT and a positive log odds ratio for stroke. The results seemed to suggest that as the carotid thickness in the

experimental arm becomes lower compared to the control arm, either through regression or slower progression of atherosclerosis in the experimental arm, the protective effect of the treatment for the risk of stroke decreases until a certain threshold value is reached where further differences in carotid thickness are associated with increased risk for stroke.

This observation cannot be readily explained especially because the experimental and control arms were not the same in each study. One explanation is that the observed results are a chance finding. In order to better understand the importance of the negative relationship between mean difference in CIMT and the log odds ratio for stroke, it is informative to understand the results which indicated an increased risk of stroke, albeit non-significant, in drug classes that showed a lower CIMT in the randomized studies included in this review. To this end, the results of the VHAS [93] study interpreted in the context of other studies that have compared calcium channel blockers to thiazide diuretics indicate that it is likely a chance finding that calcium channel blockers are associated with a lower CIMT and a greater risk of stroke. Although the VHAS [93] study and MIDAS [79] showed a lower CIMT and an increased risk of stroke in the CCB group compared to diuretic, a recent meta-analysis did not conclude that CCBs were associated with a greater risk for stroke as compared to diuretics. Chen et al [123] performed a meta-analysis of 5 studies which compared a CCB vs. diuretic and found that CCB was associated with a non-significant 6% lower risk of stroke as compared to diuretics. The results of the pooled analysis indicated that the risk ratio was 0.94 (95% CI: 0.84, 1.05). Further to this, in the parent trial for the VHAS [93] study, there was a lower number of strokes in the CCB group as compared to the thiazide diuretic group.

Despite these explanations, it is of clinical relevance to consider that treatments which may have a relatively higher effectiveness in reducing the progression or inducing regression of CIMT may be associated with lower protection from stroke or even an increased risk of stroke. There is some evidence to suggest that this association may be real. One source of supporting evidence is an exploratory analysis in which a negative association between CIMT and myocardial infarction was not observed in the subgroup of studies that were used for the analysis involving stroke. From these results, it seems

plausible that while CIMT may be a good surrogate outcome for a purely cardiovascular outcome such as MI, it may be a poor surrogate outcome for a cerebrovascular outcome such as stroke. Treatments may have unintended harmful effects on those disease pathways that lead to stroke and which are not present in the context of cardiovascular outcomes. It is unclear whether any harmful effect of these treatments on risk of stroke is related to fibrofatty by-products of the atherosclerosis regression entering into the circulation. This is because in VHAS [93], which as mentioned earlier was the only study that showed a negative mean difference in CIMT and a positive log odds ratio for stroke, both treatment groups were associated with progression in CIMT. Further to this, because most of the imaging trials have shown a reduced net progression in CIMT rather than regression, changes that occur to the carotid wall during regression are unclear.

From a review of the literature it was not possible to discern whether the treatments are associated with any degree of atherosclerosis regression, even if there is net progression due to atherosclerosis from other sources, and whether any regression would result in a tendency for atherosclerotic plaque and thrombi to dislodge from the carotid wall and increase the risk of ischemic stroke. Although the mechanisms remain elusive, it seems plausible the increase in stroke may be due to pro-coagulant properties of the treatments leading to ischemic stroke rather than anti-coagulant effects which would result in hemorrhagic strokes. Support for this may be found in the interpretation of the results for the EPAT [94] study in light of additional data. The EPAT [94] study showed a lower CIMT and a higher risk of stroke for the hormone therapy group compared to placebo, and the large Women's Health Initiative (WHI) trial[121] found an increased risk of ischemic stroke attributable to procoagulant properties of hormone therapy. For the comparison of estrogen vs. placebo, the hazard ratio for ischemic stroke was 1.55 (95% CI: 1.19, 2.01). Although there were a lower number of hemorrhagic strokes, the study did not find an increased risk of hemorrhagic stroke in the hormone therapy group. These interpretations for evaluation of CIMT as a surrogate outcome for stroke are largely speculative and may or may not be true for all drug classes. It is likely that the observed results are a combination of chance and significant findings, depending on the drug class.

5 DISCUSSION

5.1 Summary

In the first part, a comprehensive review of the literature to identify and describe statistical methods that may be used to evaluate the strength of a surrogate outcome was conducted. The underlying purpose of the review was to identify methods that may be implemented using summary level data for the surrogate and clinical outcome, which can be easily obtained from reports of randomized trials. In the second phase of this thesis, a systematic review of the literature was performed to identify randomized imaging trials for atherosclerosis which have evaluated the effect of treatment on changes CIMT. The statistical methods identified in the first phase were then applied to the data obtained from this systematic review in order to evaluate the strength of CIMT as a surrogate outcome for 4 clinical outcomes: all-cause mortality, myocardial infarction, stroke and cardiovascular mortality.

The review of statistical methods identified 5 procedures. The first procedure was significance testing for study-wise agreement. In this procedure, the measure of association to describe the strength of the surrogate outcome was study-wise percent concordance. The second procedure was rank correlation analysis and its measure of association was Spearman's ρ. A third procedure was a simple graphical analysis. The fourth procedure was regression analysis and this was associated with two measures of association: R^2 and STE. The fifth procedure was a mixed effects model by Korn et al, also associated with two measures of association: β and delta.

In the second part, a systematic review of randomized trials for atherosclerosis was performed. The review identified trials that measured the effectiveness of interventions on CIMT. The statistical procedures were applied to determine whether CIMT is a good surrogate outcome for all-cause mortality, myocardial infarction, stroke and cardiovascular mortality. The results indicated that CIMT may be a good surrogate outcome for myocardial infarction and cardiovascular mortality.

5.2 The role of the findings in relation to current literature

Based on background knowledge and literature searches about this topic, the work in this thesis is the first attempt towards the evaluation of CIMT as a surrogate outcome in each of the drug classes for which this outcome has been used to measure the effectiveness of treatments for atherosclerosis. The results of the systematic review indicated that there was insufficient data to perform an evaluation of CIMT for all of the drug class comparisons. This lack of data indicates that currently, there is not enough evidence to suggest that CIMT is a valid surrogate outcome in any drug class. In addition to the results of this work, one other meta-analytic evaluation of CIMT as a surrogate outcome has been performed for trials of statins versus placebo [124]. Similarities and differences were noted between the results of the meta-analytic evaluation by Espeland et al [124] and the results of the current work. Overall, Espeland et al's [124] findings were in agreement with the findings of this thesis. Neither source was able to document a statistically significant association between CIMT and the clinical outcomes that were considered. The thesis work furthered the methodology used by Espeland et al [124]. Because it is possible that the strength of CIMT as a surrogate outcome varies for different clinical outcomes, the current evaluation considered CIMT separately for each of all-cause mortality, myocardial infarction, stroke and cardiovascular mortality. In the evaluation by Espeland et al [124], the clinical outcome was a composite measure which included any of cardiovascular disease death, myocardial infarction, stroke, clinical coronary events, or clinical events. By indicating that the relationship of CIMT with stroke may be different from its relationship with purely cardiovascular outcomes, the results indicate that composite measures including stroke should be avoided when evaluating CIMT as a surrogate outcome. The current analysis was not based some of the studies that were used in the analysis performed by Espeland et al [124]. In the current work, the results of ACAPS [125], CAIUS [106], REGRESS [116] and the statin arm of the BCAPS [96] trial were not used to evaluate CIMT. The ACAPS [125] study was excluded because CIMT and the clinical outcomes were reported for different populations. CAIUS [106] was excluded because there was one myocardial infarction and it occurred during the first 4 months of treatment. In this way, the study did not satisfy the underlying assumption in surrogate outcome evaluation that change in the surrogate outcome needs to be predictive of *future* clinical

outcomes. The REGRESS [116] study was excluded because the authors did not measure CIMT at 12 months. The authors of this study provided a yearly rate of change in CIMT based on a simple rate calculation that involves baseline CIMT, follow-up CIMT and study duration. Data from the statin arm of the BCAPS [96] study was not used because the authors reported a composite measure for the clinical outcomes and it was not possible to determine the number of events for each of the relevant clinical outcomes.

Application of the models proposed by Korn et al [32] to the CIMT data provided an evaluation of the methods proposed by these authors. Korn et al [32] have recommended use of the random effects models and both the applications presented in the Korn et al [32] methodology paper were based on the random effects models. In applying the random effects model for the surrogate outcome evaluation of CIMT, it was noted that this model can only be fitted to data in which the τ^2 value is sufficiently different from zero. Because τ^2 is required to calculate Δ, this parameter cannot be determined and used to evaluate the strength of a surrogate outcome in situations where only a fixed effects model could be applied to the data.

The percent concordance values were uninformative and this finding is in agreement with the results of previous research for surrogate outcome evaluation in oncology. Johnson et al [30] evaluated response rate and time to progression as surrogate outcomes for mortality in metastatic colorectal cancer and non-small-cell lung cancer. The authors performed study wise agreement analysis and concluded that the percent concordance values were not useful in evaluating the surrogate outcomes. The current results strengthen the conclusions of Johnson et al [30] and suggest that study wise agreement should not be used as a method to evaluate the strength of surrogate outcomes.

5.3 Recommendations for future work

The regression models for surrogate outcome evaluation that were identified in this work have not considered covariate adjustment. Because the observed relationship between a predictor and response may change with the inclusion of covariates, future developments may consider the inclusion of covariates when evaluating the strength of surrogate outcomes. For the methods proposed by Korn et al [32], future work may focus on defining

Δ or a similar parameter for situations in which τ^2 is approximately zero and only fixed effects models can be fitted to the available data.

The evaluation of CIMT in specific drug classes indicated that even for the simple statistical methods that were identified for surrogate outcome evaluation, there is insufficient data within specific drug classes to adequately determine the usefulness of CIMT as a surrogate outcome in clinical trials. To this end, it is important for investigators to perform a larger number of clinical trials that measure response to treatment on both CIMT and clinical outcomes. Because there are a large number of industry funded trials, it is plausible that this may be achieved if regulatory bodies required data on both types of outcomes in the setting where data from industry funded trials is used towards the drug-approval process. Where funding is received from granting agencies, this may be achieved if granting agencies were to suggest the measurement of relevant surrogate outcomes to those research groups who receive funding for large outcome trials. This type of initiative would require communication between granting agencies, regulatory bodies and researchers with an interest in surrogate outcome evaluation. These researchers may be able to provide information about which surrogate outcome are hypothesized to be the best substitutes for clinical outcomes. This information can subsequently be used to suggest which surrogate outcomes should be measured when clinical trialists receive funding for large outcomes trials. Alternatively, this may inform investigators with industry funded grants about the outcomes for which data is required when submitting results to regulatory authorities.

Until there is sufficient data from clinical trials that have measured both clinical and surrogate outcomes, it may be possible to evaluate surrogate outcomes by combining data from clinical outcomes trials with data from surrogate outcomes trials. Duivenvoorden et al [126] have explored this method to evaluate CIMT in trials for lipid modifying drugs. The investigators identified a set of randomized trials which evaluated the effect of lipid modifying therapy on the annual rate of change in CIMT. For each these trials, the authors also identified outcomes trials that evaluated the same drug class comparison in patients with similar characteristics. The authors then performed a regression analysis in which the relative risk of all-cause mortality was regressed on annual rate of change in

CIMT. In this way, each observation in the regression analysis consisted of data from a pair of studies. Because it may be challenging to determine whether a given pair of studies is similar both clinically and methodologically, this type of analysis would require close collaboration between methodologists and content experts.

The results of this work indicated that CIMT may be a good surrogate outcome for myocardial infarction and cardiovascular mortality. There is insufficient evidence to suggest that CIMT is a good surrogate outcome for all-cause mortality. The results indicate that CIMT may be a poor surrogate outcome for stroke. Application of the statistical methods indicated that the best procedures to evaluate surrogate outcomes in a meta-analytic setting are the STE, the weighted R^2 and Spearman's ρ. The graphical analysis method is a useful method to do a preliminary exploration of the strength of a surrogate outcome, but observations are likely to be too subjective to form conclusions[29,40]. The Korn et al model is useful in theory, but requires the estimation of a large number of parameters for which data from a large number of studies is required. Also, the Korn et al random effects models may fail to produce results. Conclusions based on the percent concordance parameter may be inaccurate and it has limited usefulness when there are a large number of studies with null effects for both outcomes. In this setting, it may overestimate the strength of a surrogate outcome.

References

1. De Gruttola VG, Clax P, DeMets DL et al. Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. 2001.

2. Heinonen T, Waters DD, Libby P, Tardif JC. A winter's tale: report from the First Annual Canadian Biomarkers and Surrogate Endpoints Symposium. [Review] [46 refs]. Canadian Journal of Cardiology 2009;25(9):527-532.

3. Kastelein JJ, de GE. Ultrasound imaging techniques for the evaluation of cardiovascular therapies. Eur Heart J 2008;29(7):849-858.

4. Duivenvoorden R, de GE, Stroes ES, Kastelein JJ. Surrogate markers in clinical trials--challenges and opportunities. [Review] [72 refs]. Atherosclerosis 2009;206(1):8-16.

5. Espeland MA, O'leary DH, Terry JG, Morgan T, Evans G, Mudra H. Carotid intimal-media thickness as a surrogate for cardiovascular disease events in trials of HMG-CoA reductase inhibitors. Curr Control Trials Cardiovasc Med 2005;6(1):3.

6. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Therapy and applying the results surrogate outcomes. In: Guyatt G, Rennie D, editors. Users' guides to the medical literature a manual for evidence-based clinical practice. Chicago. AMA press; 2002: 393.

7. Burzykowski T, Buyse M, Molenberghs G. The evaluation of surrogate endpoints. New York: Springer; 2005.

8. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5(3):173-186.

9. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996;125(7):605-613.

10. Lassere MN, Johnson KR, Boers M et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol 2007;34(3):607-615.

11. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 2008;17(3):303-340.

12. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8(4):431-440.

13. Li Z, Chines AA, Meredith MP. Statistical validation of surrogate endpoints: is bone density a valid surrogate for fracture? J Musculoskelet Neuronal Interact 2004;4(1):64-74.

14. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med 1992;11(2):167-178.

15. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 1998;54(3):1014-1029.

16. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 2000;1(1):49-67.

17. Psaty BM, Weiss NS, Furberg CD et al. Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. JAMA 1999;282(8):786-790.

18. Gordon MB, Libby P. Atherosclerosis. In: Lilly LS, editor. Pathophysiology of heart disease: a collaborative project of medical students and faculty. 3 ed. Philadelphia: Lippincott Williams and Wilkins; 2003:111.

19. Schoen FJ. Blood Vessels. In: Kumar V, Abbas AK, Fausto N, editors. Robbins and Cotran pathologic basis of disease. 7 ed. Philadelphia: Elsevier/Saunders; 2005:511.

20. Grobbee DE, Bots ML. Atherosclerotic disease regression with statins: studies using vascular markers. Int J Cardiol 2004;96(3):447-459.

21. Popma J. Coronary Arteriography and Intravascular Imaging. In: Libby P, Bonow RO, Mann DL, editors. Braunwald's heart disease : a textbook of cardiovascular medicine. 8 ed. Philadelphia: Saunders/Elsevier; 2008: 423.

22. Revkin JH, Shear CL, Pouleur HG, Ryder SW, Orloff DG. Biomarkers in the prevention and treatment of atherosclerosis: need, validation, and future. Pharmacol Rev 2007;59(1):40-53.

23. Glenny AM, Altman DG, Song F et al. Indirect comparisons of competing interventions. Health Technol Assess 2005;9(26):1-iv.

24. Begg CB, Leung DHY. On the use of surrogate end points in randomized trials. Journal of the Royal Statistical Society Series A: Statistics in Society 2000;163(1):15-28.

25. Freedman L. Quantitative science methods for biomarker validation in chemoprevention trials. [Review] [21 refs]. Cancer Biomarkers: Section A of Disease Markers 2007;3(3):135-140.

26. Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Statistical methods in medical research 2008;17(5):477-486.

27. Hotta K, Kiura K, Fujiwara Y et al. Association between incremental gains in the objective response rate and survival improvement in phase III trials of first-line chemotherapy for extensive disease small-cell lung cancer. Annals of Oncology 2009;20(5):829-834.

28. Hotta K, Fujiwara Y, Matsuo K et al. Time to progression as a surrogate marker for overall survival in patients with advanced non-small cell lung cancer. [Review] [31 refs]. Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer 2009;4(3):311-317.

29. Hughes MD, DeGruttola V, Welles SL. Evaluating surrogate markers. Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association 1995;10 Suppl 2.

30. Johnson KR, Ringland C, Stokes BJ et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncology 2006;7(9):741-746.

31. Johnson KR, Freemantle N, Anthony DM, Lassere MN. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE). [Review] [72 refs]. Journal of Clinical Epidemiology 2009;62(3):328-336.

32. Korn EL, Albert PS, McShane LM. Assessing surrogates as trial endpoints using mixed models. Statistics in medicine 2005;24(2):163-182.

33. Lassere M, Johnson K, Hughes M et al. Simulation studies of surrogate endpoint validation using single trial and multitrial statistical approaches. Journal of Rheumatology 2007;34(3):616-619.

34. Macchia A, Marchioli R, Marfisi R et al. A meta-analysis of trials of pulmonary hypertension: a clinical condition looking for drugs and research methodology. [Review] [44 refs]. American Heart Journal 2007;153(6):1037-1047.

35. Miksad RA, Zietemann V, Gothe R et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. International Journal of Technology Assessment in Health Care 2008;24(4):371-383.

36. Mills EJ, Kelly S, Bradley M, Mollon P, Cooper C, Nachega J. Antiretroviral effects on HIV-1 RNA, CD4 cell count and progression to AIDS or death: a meta-regression analysis. [Review] [29 refs]. HIV Medicine 2008;9(10):849-857.

37. Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. [Review] [105 refs]. International Journal of Clinical Oncology 2009;14(2):102-111.

38. Sormani MP, Bonzano L, Roccatagliata L, Cutter GR, Mancardi GL, Bruzzi P. Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. Annals of Neurology 2009;65(3):268-275.

39. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. [Review] [63 refs]. Journal of Clinical Oncology 2007;25(29):4562-4568.

40. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Statistics in medicine 2006;25(2):183-203.

41. Freedman L. Comment on: Assessing surrogates as trial endpoints using mixed models. Stat Med 2005;24(2):183-185.

42. Montagnani F, Migali C, Fiorentini G. Progression-free survival in bevacizumab-based first-line treatment for patients with metastatic colorectal cancer: is it a really good end point? J Clin Oncol 2009;27(28):e132-e133.

43. Perrone F. Don't forget survival, please.. Lancet Oncol 2006;7(9):703-704.

44. Rich S. The value of approved therapies for pulmonary arterial hypertension. Am Heart J 2007;153(6):889-890.

45. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5(3):173-186.

46. Fleming TR. Surrogate markers in AIDS and cancer trials. Stat Med 1994;13(13-14):1423-1435.

47. Guyatt GH, Cranney A, Griffith L et al. Summary of meta-analyses of therapies for postmenopausal osteoporosis and the relationship between bone density and fractures. Endocrinol Metab Clin North Am 2002;31(3):659-79, xii.

48. Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. Br J Cancer 2005;93(11):1215-1221.

49. Hotta K, Fujiwara Y, Kiura K et al. Relationship between response and survival in more than 50,000 patients with advanced non-small cell lung cancer treated with systemic chemotherapy in 143 phase III trials. J Thorac Oncol 2007;2(5):402-407.

50. Torri V, Simon R, Russek-Cohen E, Midthune D, Friedman M. Statistical model to determine the relationship of response and survival in patients with advanced ovarian cancer treated with chemotherapy. J Natl Cancer Inst 1992;84(6):407-414.

51. Macchia A, Marchioli R, Tognoni G et al. Systematic review of trials using vasodilators in pulmonary arterial hypertension: why a new approach is needed. Am Heart J 2010;159(2):245-257.

52. Sherrill B, Amonkar M, Wu Y et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. [Review] [15 refs].

53. Touboul PJ, Hennerici MG, Meairs S et al. Mannheim carotid intima-media thickness consensus (2004-2006). An update on behalf of the Advisory Board of the 3rd and 4th Watching the Risk Symposium, 13th and 15th European Stroke Conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006. Cerebrovasc Dis 2007;23(1):75-80.

54. Lefebvre C, Manheimer E, Glanville J. Searching for studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for systematic reviews of interventions. 5 ed. West Sussex: Cochrane Collaboration and John Wiley & Sons; 2009:95-150.

55. Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc 2006;94(1):41-47.

56. Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions. 5 ed. West Sussex: Cochrane Collaboration and John Wiley & Sons; 2009.

57. Targeting Peroxisome Proliferator-activated Receptor-gamma in Peritoneal Dialysis Patients - Will it Reduce Inflammation, Atherosclerosis, Calcification and Improve Survival of Peritoneal Dialysis Patients? ClinicalTrials gov 2010.

58. Dagenais GR, Gerstein HC, Holman R et al. Effects of ramipril and rosiglitazone on cardiovascular and renal outcomes in people with impaired glucose tolerance or impaired fasting glucose: results of the Diabetes REduction Assessment with ramipril and rosiglitazone Medication (DREAM) trial. Diabetes Care 2008;31(5):1007-1014.

59. Kastelein JJ, Sager PT, de GE, Veltri E. Comparison of ezetimibe plus simvastatin versus simvastatin monotherapy on atherosclerosis progression in familial hypercholesterolemia. Design and rationale of the Ezetimibe and Simvastatin in Hypercholesterolemia Enhances Atherosclerosis Regression (ENHANCE) trial. American Heart Journal 2005;149(2):234-239.

60. Rosei EA, Dal PC, Leonetti G, Magnani B, Pessina A, Zanchetti A. Clinical results of the Verapamil inHypertension and Atherosclerosis Study. VHAS Investigators. J Hypertens 1997;15(11):1337-1344.

61. Lonn E, Yusuf S, Dzavik V et al. Effects of ramipril and vitamin E on atherosclerosis: the study to evaluate carotid ultrasound changes in patients treated with ramipril and vitamin E (SECURE). Circulation 2001;103(7):919-925.

62. Byington RP, Furberg CD, Crouse JR, III, Espeland MA, Bond MG. Pravastatin, Lipids, and Atherosclerosis in the Carotid Arteries (PLAC-II). American Journal of Cardiology 1995;76(9):54C-59C.

63. Crouse JR, Byington RP, Bond MG et al. Pravastatin, lipids, and atherosclerosis in the carotid arteries: design features of a clinical trial with carotid atherosclerosis outcome. Controlled Clinical Trials 1992;13(6):495-506.

64. Smilde TJ, Trip MD, Wollersheim H, van WS, Kastelein JJP, Stalenhoef AFH. Rationale, design and baseline characteristics of a clinical trial comparing the effects of robust vs conventional cholesterol lowering and intima media thickness in patients with familial hypercholesterolaemia: The atorvastatin versus simvastatin on atherosclerosis progression (ASAP) study. Clinical Drug Investigation 2000;20(2):67-79.

65. Smilde TJ, van WS, Wollersheim H, Trip MD, Kastelein JJ, Stalenhoef AF. Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. Lancet 2001;357(9256):577-581.

66. Salonen JT, Nyyssonen K, Salonen R et al. Antioxidant Supplementation in Atherosclerosis Prevention (ASAP) study: a randomized trial of the effect of vitamins E and C on 3-year progression of carotid atherosclerosis. Journal of Internal Medicine 2000;248(5):377-386.

67. Salonen RM, Nyyssonen K, Kaikkonen J et al. Six-year effect of combined vitamin C and E supplementation on atherosclerotic progression: the Antioxidant Supplementation in Atherosclerosis Prevention (ASAP) Study. Circulation 2003;107(7):947-953.

68. Zanchetti A. Prevalence of carotid atherosclerosis in hypertension: preliminary baseline data from the European Lacidipine Study on Atheroscelerosis (ELSA). Blood pressure Supplement 1996;4:30-35.

69. Zanchetti A, Bond MG, Hennig M et al. Risk factors associated with alterations in carotid intima-media thickness in hypertension: baseline data from the European Lacidipine Study on Atherosclerosis. Journal of Hypertension 1998;16(7):949-961.

70. Zanchetti A, Bond MG, Hennig M et al. Calcium antagonist lacidipine slows down progression of asymptomatic carotid atherosclerosis: principal results of the European Lacidipine Study on Atherosclerosis (ELSA), a randomized, double-blind, long-term trial. Circulation 2002;106(19):2422-2427.

71. Zanchetti A, Bond MG, Hennig M et al. Absolute and relative changes in carotid intima-media thickness and atherosclerotic plaques during long-term antihypertensive treatment: further results of the European Lacidipine Study on Atherosclerosis (ELSA). Journal of Hypertension 2004;22(6):1201-1212.

72. Bots ML, Palmer MK, Dogan S et al. Intensive lipid lowering may reduce progression of carotid atherosclerosis within 12 months of treatment: the METEOR study. Journal of Internal Medicine 2009;265(6):698-707.

73. Crouse JR, III, Raichlen JS, Riley WA et al. Effect of rosuvastatin on progression of carotid intima-media thickness in low-risk individuals with subclinical atherosclerosis: the METEOR Trial. JAMA 2007;297(12):1344-1353.

74. Kastelein JJ, van Leuven SI, Burgess L et al. Effect of torcetrapib on carotid atherosclerosis in familial hypercholesterolemia. New England Journal of Medicine 2007;356(16):1620-1630.

75. Kastelein JJ, van Leuven SI, Evans GW et al. Designs of RADIANCE 1 and 2: carotid ultrasound studies comparing the effects of torcetrapib/atorvastatin with atorvastatin alone on atherosclerosis. Current Medical Research & Opinion 2007;23(4):885-894.

76. Howard BV, Roman MJ, Devereux RB et al. Effect of lower targets for blood pressure and LDL cholesterol on atherosclerosis in diabetes: The SANDS randomized trial. JAMA - Journal of the American Medical Association 2008;299(14):1678-1689.

77. Russell M, Fleg JL, Galloway WJ et al. Examination of lower targets for low-density lipoprotein cholesterol and blood pressure in diabetes--the Stop Atherosclerosis in Native Diabetics Study (SANDS). Am Heart J 2006;152(5):867-875.

78. Borhani NO, Bond MG, Sowers JR et al. The Multicenter Isradipine/Diuretic Atherosclerosis Study: a study of the antiatherogenic properties of isradipine in hypertensive patients. MIDAS Research Group. Journal of Cardiovascular Pharmacology 1991;18 Suppl 3:S15-S19.

79. Borhani NO, Mercuri M, Borhani PA et al. Final outcome results of the Multicenter Isradipine Diuretic Atherosclerosis Study (MIDAS). A randomized controlled trial. JAMA : the journal of the American Medical Association 1996;276:785-791.

80. Blankenhorn DH, Johnson RL, Nessim SA, Azen SP, Sanmarco ME, Selzer RH. The Cholesterol Lowering Atherosclerosis Study (CLAS): design, methods, and baseline results. Control Clin Trials 1987;8(4):356-387.

81. Hodis HN, Mack WJ, LaBree L et al. The role of carotid arterial intima-media thickness in predicting clinical coronary events. Annals of Internal Medicine 1998;128(4):262-269.

82. Bots ML, Evans GW, Riley W et al. The Osteoporosis Prevention and Arterial effects of tiboLone (OPAL) study: design and baseline characteristics. Controlled Clinical Trials 2003;24(6):752-775.

83. Bots ML, Evans GW, Riley W et al. The effect of tibolone and continuous combined conjugated equine oestrogens plus medroxyprogesterone acetate on progression of carotid intima-media thickness: the Osteoporosis Prevention and Arterial effects of tiboLone (OPAL) study. European Heart Journal 2006;27(6):746-755.

84. Sawayama Y, Shimizu C, Maeda N et al. Effects of probucol and pravastatin on common carotid atherosclerosis in patients with asymptomatic hypercholesterolemia. Fukuoka Atherosclerosis Trial (FAST). Journal of the American College of Cardiology 2002;39(4):610-616.

85. Lonn EM, Gerstein HC, Sheridan P et al. Effect of ramipril and of rosiglitazone on carotid intima-media thickness in people with impaired glucose tolerance or impaired fasting glucose: STARR (STudy of Atherosclerosis with Ramipril and Rosiglitazone). Journal of the American College of Cardiology 2009;53(22):2028-2035.

86. van Vonderen MG, Hassink EA, van Agtmael MA et al. Increase in carotid artery intima-media thickness and arterial stiffness but improvement in several markers of endothelial function after initiation of antiretroviral therapy. Journal of Infectious Diseases 2009;199(8):1186-1194.

87. Wiklund O, Hulthe J, Wikstrand J, Schmidt C, Olofsson SO, Bondjers G. Effect of controlled release/extended release metoprolol on carotid intima-media thickness in patients with hypercholesterolemia: a 3-year randomized study. Stroke; a journal of cerebral circulation 2002;33:572-577.

88. Meuwese MC, de GE, Duivenvoorden R et al. ACAT inhibition and progression of carotid atherosclerosis in patients with familial hypercholesterolemia: the CAPTIVATE randomized trial. JAMA 2009;301(11):1131-1139.

89. Elkeles RS, Diamond JR, Poulter C et al. Cardiovascular outcomes in type 2 diabetes. A double-blind placebo-controlled study of bezafibrate: the St. Mary's, Ealing, Northwick Park Diabetes Cardiovascular Disease Prevention (SENDCAP) Study. Diabetes Care 1998;21(4):641-648.

90. Nanayakkara PW, van GC, ter Wee PM et al. Effect of a treatment strategy consisting of pravastatin, vitamin E, and homocysteine lowering on carotid intima-media thickness,

endothelial function, and renal function in patients with mild to moderate chronic kidney disease: results from the Anti-Oxidant Therapy in Chronic Renal Insufficiency (ATIC) Study. Archives of Internal Medicine 2007;167(12):1262-1270.

91. Kastelein JJ, Akdim F, Stroes ES et al. Simvastatin with or without ezetimibe in familial hypercholesterolemia.[Erratum appears in N Engl J Med. 2008 May 1;358(18):1977]. New England Journal of Medicine 2008;358(14):1431-1443.

92. Bots ML, Visseren FL, Evans GW et al. Torcetrapib and carotid intima-media thickness in mixed dyslipidaemia (RADIANCE 2 study): a randomised, double-blind trial. Lancet 2007;370(9582):153-160.

93. Zanchetti A, Rosei EA, Dal PC, Leonetti G, Magnani B, Pessina A. The Verapamil in Hypertension and Atherosclerosis Study (VHAS): results of long-term randomized treatment with either verapamil or chlorthalidone on carotid intima-media thickness. Journal of Hypertension 1998;16(11):1667-1676.

94. Hodis HN, Mack WJ, Lobo RA et al. Estrogen in the prevention of atherosclerosis. A randomized, double-blind, placebo-controlled trial. Annals of Internal Medicine 2001;135(11):939-953.

95. Mazzone T, Meyer PM, Feinstein SB et al. Effect of pioglitazone compared with glimepiride on carotid intima-media thickness in type 2 diabetes: a randomized trial. JAMA 2006;296(21):2572-2581.

96. Hedblad B, Wikstrand J, Janzon L, Wedel H, Berglund G. Low-dose metoprolol CR/XL and fluvastatin slow progression of carotid intima-media thickness: Main results from the Beta-Blocker Cholesterol-Lowering Asymptomatic Plaque Study (BCAPS). Circulation 2001;103(13):1721-1726.

97. Magliano D, McNeil J, Branley P et al. The Melbourne Atherosclerosis Vitamin E Trial (MAVET): a study of high dose vitamin E in smokers. European Journal of Cardiovascular Prevention & Rehabilitation 2006;13(3):341-347.

98. Salonen R, Nyssönen K, Porkkala SE, Salonen JT. The Kuopio Atherosclerosis Prevention Study (KAPS): effect of pravastatin treatment on lipids, oxidation resistance of lipoproteins, and atherosclerotic progression. The American journal of cardiology 1995;76:34C-39C.

99. Terpstra WF, May JF, Smit AJ, Graeff PA, Meyboom-de JB, Crijns HJ. Effects of amlodipine and lisinopril on intima-media thickness in previously untreated, elderly hypertensive patients (the ELVERA trial). Journal of Hypertension 2004;22(7):1309-1316.

100. Hodis HN, Mack WJ, Zheng L et al. Effect of peroxisome proliferator-activated receptor gamma agonist treatment on subclinical atherosclerosis in patients with insulin-requiring type 2 diabetes. Diabetes Care 2006;29(7):1545-1553.

101. Hodis HN, Mack WJ, LaBree L et al. Alpha-tocopherol supplementation in healthy individuals reduces low-density lipoprotein oxidation but not atherosclerosis: the Vitamin E Atherosclerosis Prevention Study (VEAPS). Circulation 2002;106(12):1453-1459.

102. Hodis HN, Mack WJ, Dustin L et al. High-dose B vitamin supplementation and progression of subclinical atherosclerosis: a randomized controlled trial. Stroke 2009;40(3):730-736.

103. Furberg CD, Adams J, Applegate WB et al. Effect of lovastatin on early carotid atherosclerosis and cardiovascular events. Circulation 1994;90(4 I):1679-1687.

104. Hodis HN, Mack WJ, LaBree L et al. Reduction in carotid arterial wall thickness using lovastatin and dietary therapy: a randomized controlled clinical trial. Annals of Internal Medicine 1996;124(6):548-556.

105. Blankenhorn DH, Azen SP, Kramsch DM et al. Coronary angiographic changes with lovastatin therapy: The monitored atherosclerosis regression study (MARS). Annals of Internal Medicine 1993;119(10):969-976.

106. Mercuri M, Bond MG, Sirtori CR et al. Pravastatin reduces carotid intima-media thickness progression in an asymptomatic hypercholesterolemic mediterranean population: the Carotid Atherosclerosis Italian Ultrasound Study. American Journal of Medicine 1996;101(6):627-634.

107. Byington RP, Furberg CD, Herrington DM et al. Effect of estrogen plus progestin on progression of carotid atherosclerosis in postmenopausal women with heart disease: HERS B-mode substudy. Arteriosclerosis, Thrombosis & Vascular Biology 2002;22(10):1692-1697.

108. Taylor AJ, Kent SM, Flaherty PJ, Coyle LC, Markwood TT, Vernalis MN. ARBITER: Arterial Biology for the Investigation of the Treatment Effects of Reducing Cholesterol: a randomized trial comparing the effects of atorvastatin and pravastatin on carotid intima medial thickness. Circulation 2002;106(16):2055-2060.

109. Nathan DM, Lachin J, Cleary P et al. Intensive diabetes therapy and carotid intima-media thickness in type 1 diabetes mellitus. New England Journal of Medicine 2003;348(23):2294-2303.

110. Zoungas S, McGrath BP, Branley P et al. Cardiovascular morbidity and mortality in the Atherosclerosis and Folic Acid Supplementation Trial (ASFAST) in chronic renal failure: a multicenter, randomized, controlled trial. Journal of the American College of Cardiology 2006;47(6):1108-1116.

111. Taylor AJ, Villines TC, Stanek EJ et al. Extended-release niacin or ezetimibe and carotid intima-media thickness. New England Journal of Medicine 2009;361(22):2113-2122.

112. MacMahon S, Sharpe N, Gamble G et al. Effects of lowering average of below-average cholesterol levels on the progression of carotid atherosclerosis: results of the LIPID Atherosclerosis Substudy. LIPID Trial Research Group. Circulation 1998;97:1784-1790.

113. MacMahon S, Sharpe N, Gamble G et al. Randomized, placebo-controlled trial of the angiotensin-converting enzyme inhibitor, ramipril, in patients with coronary or other occlusive arterial disease. PART-2 Collaborative Research Group. Prevention of Atherosclerosis with Ramipril. Journal of the American College of Cardiology 2000;36(2):438-443.

114. Pitt B, Byington RP, Furberg CD et al. Effect of amlodipine on the progression of atherosclerosis and the occurrence of clinical events. Circulation 2000;102(13):1503-1510.

115. Hanefeld M, Chiasson JL, Koehler C, Henkel E, Schaper F, Temelkova KT. Acarbose slows progression of intima-media thickness of the carotid arteries in subjects with impaired glucose tolerance. Stroke; a journal of cerebral circulation 2004;35:1073-1078.

116. de GE, Jukema JW, van Boven AJ et al. Effect of pravastatin on progression and regression of coronary atherosclerosis and vessel wall changes in carotid and femoral arteries: a report from the Regression Growth Evaluation Statin Study. American Journal of Cardiology 1995;76(9):40C-46C.

117. De LF, Boffito M, Collot-Teixeira S et al. Prevention of atherosclerosis in patients living with HIV. Vascular Health & Risk Management 2009;5(1):287-300.

118. DeFronzo RA, Banerji MA, Bray GA et al. Actos Now for the prevention of diabetes (ACT NOW) study. BMC Endocrine Disorders 2009;9.

119. Cilostazol Versus Aspirin for Primary Prevention of Atherosclerotic Events (CAPPA). ClinicalTrials gov 2010.

120. Bhandari M, Busse JW, Jackowski D et al. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. CMAJ 2004;170(4):477-480.

121. Hendrix SL, Wassertheil-Smoller S, Johnson KC et al. Effects of conjugated equine estrogen on stroke in the Women's Health Initiative. Circulation 2006;113(20):2425-2434.

122. Farquhar C, Marjoribanks J, Lethaby A, Suckling JA, Lamberts Q. Long term hormone therapy for perimenopausal and postmenopausal women. Cochrane Database Syst Rev 2009;(2):CD004143.

123. Chen N, Zhou M, Yang M et al. Calcium channel blockers versus other classes of drugs for hypertension. Cochrane Database Syst Rev 2010;(8):CD003654.

124. Espeland MA, O'Leary DH, Terry JG, Morgan T, Evans G, Mudra H. Carotid intimal-media thickness as a surrogate for cardiovascular disease events in trials of HMG-CoA reductase inhibitors. Curr Control Trials Cardiovasc Med 2005;6(1):3.

125. Furberg CD, Adams HP, Jr., Applegate WB et al. Effect of lovastatin on early carotid atherosclerosis and cardiovascular events. Asymptomatic Carotid Artery Progression Study (ACAPS) Research Group. Circulation 1994;90(4):1679-1687.

126. Duivenvoorden R, Nederveen AJ, de GE, Kastelein JJ. Atherosclerosis imaging as a benchmark in the development of novel cardiovasular drugs. Curr Opin Lipidol 2007;18(6):613-621.

Appendix A: search strategy for trial-level methods for surrogate outcome evaluation

1. *Biological Markers/
2. surrogate$.mp.
3. intermediate$ outcome$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
4. intermediate$ endpoint$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
5. intermediate$ marker$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
6. intermediate$ indicator$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
7. intermediate end point$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
8. multiple outcome$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
9. multiple endpoint$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
10. multiple end point$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
11. meta-analysis.pt.
12. exp Meta-Analysis as Topic/
13. meta-anal$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
14. summary data.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
15. trial-level data.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
16. trial-level summar$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
17. (randomi#ed adj4 trial#).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
18. 2 or 3 or 4 or 5 or 6 or 7
19. 1 and 18
20. limit 19 to yr="2007 -Current"
21. 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
22. 11 or 12 or 13 or 14 or 15 or 16 or 17
23. 21 and 22
24. limit 23 to yr="2007 -Current"
25. 20 or 24

Appendix B: search strategy for randomized trials evaluating intervention effects on CIMT and clinical outcomes

Database: Ovid MEDLINE(R) <1950 to April Week 1 2010>
Search Strategy:
---

1 Carotid Artery, Common/ (3919)
2 Carotid Arteries/ (24039)
3 Tunica Intima/ (8201)
4 Tunica Media/ (3181)
5 (intima$ adj3 media$ adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (5448)
6 (carotid adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (5412)
7 (carotid adj10 wall$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2171)
8 (intima?media$ adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (43)
9 (carotid adj30 IMT).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2503)
10 (arter$ wall$ adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (1304)
11 randomized controlled trial.pt. (288341)
12 controlled clinical trial.pt. (81073)
13 randomized.ab. (196238)
14 placebo.ab. (118006)
15 drug therapy.fs. (1369966)
16 randomly.ab. (142606)
17 trial.ab. (202938)
18 groups.ab. (960101)
19 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 (2519248)
20 (animals not (humans and animals)).sh. (3376936)
21 19 not 20 (2135968)
22 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 (37009)
23 21 and 22 (4323)
24 from 23 keep 1-4323 (4323)

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <April 16, 2010>
Search Strategy:
---

1 random$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (26052)
2 placebo$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (3223)
3 trial.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (9094)
4 (doubl$ adj blind$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (1937)
5 (singl$ adj blind$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (243)
6 assign$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (8664)

7 allocat$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2088)
8 volunteer$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2749)
9 factorial$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (664)
10 crossover$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2487)
11 cross over$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (328)
12 cross-over$.mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (328)
13 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 (43486)
14 (intima$ adj3 media$ adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (213)
15 (carotid adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (198)
16 (carotid adj10 wall$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (50)
17 (intima?media$ adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (2)
18 (carotid adj30 IMT).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (90)
19 (arter$ wall$ adj30 thick$).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (28)
20 14 or 15 or 16 or 17 or 18 or 19 (289)
21 13 and 20 (49)
22 from 21 keep 1-49 (49)
23 from 22 keep 1-49 (49)

CENTRAL:
Search Name: CIMT v4
Comments: April 17
Save Date: 2010-04-17 18:26:01

ID	Search
#1	intima* NEAR/3 media* NEAR/30 thick* in Clinical Trials
#2	carotid NEAR/30 thick* in Clinical Trials
#3	carotid NEAR/10 wall* in Clinical Trials
#4	intima?media* NEAR/30 thick* in Clinical Trials
#5	intimamedia* NEAR/30 thick* in Clinical Trials
#6	intima*media* NEAR/30 thick* in Clinical Trials
#7	carotid NEAR/30 IMT in Clinical Trials
#8	arter* wall* NEAR/30 thick* in Clinical Trials
#9	MeSH descriptor Carotid Arteries, this term only
#10	MeSH descriptor Tunica Intima, this term only
#11	MeSH descriptor Tunica Media, this term only
#12	MeSH descriptor Carotid Artery, Common, this term only
#13	(#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12)

Database: EMBASE Classic+EMBASE <1947 to 2010 April 16>
Search Strategy:
---
1 random$.mp. (502755)

2 factorial$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (19342)
3 crossover$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (41604)
4 cross over$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (14540)
5 cross-over$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (14540)
6 placebo$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (212397)
7 (doubl$ adj blind$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (123770)
8 (singl$ adj blind$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (14022)
9 assign$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (127052)
10 allocat$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (46312)
11 volunteer$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (123418)
12 crossover procedure/ (23101)
13 double blind procedure/ (81951)
14 randomized controlled trial/ (188076)
15 single blind procedure/ (9322)
16 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 (868104)
17 random:.tw. (457735)
18 clinical trial:.mp. (674592)
19 random$.mp. (502755)
20 factorial$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (19342)
21 crossover$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (41604)
22 cross over$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (14540)
23 cross-over$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (14540)
24 placebo$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (212397)
25 (doubl$ adj blind$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (123770)
26 (singl$ adj blind$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (14022)
27 assign$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (127052)
28 allocat$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (46312)
29 volunteer$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (123418)
30 crossover procedure/ (23101)
31 double blind procedure/ (81951)
32 randomized controlled trial/ (188076)
33 single blind procedure/ (9322)
34 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 (868104)
35 random:.tw. (457735)
36 clinical trial:.mp. (674592)
37 exp health care quality/ (902771)

38 35 or 36 or 37 (1628205)
39 artery intima/ (2824)
40 artery intima proliferation/ (8910)
41 carotid artery/ (23555)
42 common carotid artery/ (5795)
43 tunica media/ (1808)
44 artery diameter/ (9349)
45 artery media/ (2432)
46 artery wall/ (8240)
47 intima/ (6821)
48 internal carotid artery/ (8798)
49 arterial wall thickness/ (1264)
50 (intima$ adj3 media$ adj30 thick$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (5603)
51 (carotid adj30 thick$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (5644)
52 (carotid adj10 wall$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (2830)
53 (intima?media$ adj30 thick$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (108)
54 (carotid adj30 IMT).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (2524)
55 (arter$ wall$ adj30 thick$).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (2768)
56 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 (66681)
57 34 and 56 (4807)
58 from 57 keep 1-4807 (4807)
59 from 57 keep 1-100 (100)

Appendix C: characteristics of included studies

C.1: Detailed tables for characteristics of included studies

Borhani 1996 (MIDAS)		
Methods	Randomized, double-blind trial. Multicentre. 3 year duration.	
Participants	Hypertension. Isradipine N = 442, age (mean, SD) = 58.2 ± 8.3, % female = 20.1, duration of diabetes in years (mean, SD) (median) = 9.9 ± 8.9 (7), % CVD = 0.7 Hydrochlorothiazide N = 441, age (mean, SD) = 58.7 ± 8.7, % female = 24.3, duration of diabetes in years (mean, SD) (median) = 10.2 ± 8.9 (8), % CVD = 2.3	
Interventions	Isradipine 2.5 – 5 mg twice per day or Hydrochlorothiazide 12.5 – 25 mg twice per day	
Outcomes	Surrogate: common carotid artery, carotid bifurcation, internal carotid, near and far wall, right and left. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from Sandoz Research Institute (SRI), Sandoz Pharmaceuticals. This study has also been referenced by Borhani 1991	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: no information provided
Allocation concealment?	UC	Comment: insufficient information provided.
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "scans were evaluated and IMTs measured centrally at the Ultrasound Reading Centre (URC), without knowledge of patients' randomization assignment". Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind" Quote: "all reported clinical events were reviewed, adjudicated, and classified by the MIDAS Investigators' Morbidity and Mortality Committee, consisting of 6 clinicians…all blinded to the randomization assignments". Comment: it is likely that the key study personnel and participants were blinded
Incomplete outcome data addressed? Surrogate	UC	Exclusions: None, because all randomized subjects (isradipine = 442; hydrochlorothiazide = 441) were included in the analysis. Attrition: no information is provided. Overall completeness of data: no information

outcome		is provided
Incomplete outcome data addressed? Clinical outcome	UC	Exclusions: None, because all randomized subjects (isradipine = 442; hydrochlorothiazide = 441) were included in the analysis. Attrition: no information is provided. Overall completeness of data: no information is provided
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: The study was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Bots 2006 (OPAL)		
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 3 year duration.	
Participants	Healthy, postmenopausal women Tibolone N = 290, age (mean, SD) = 52.4 ± 9.4, % female = 100, % CVD = 0 CEE plus MPA N = 288, age (mean, SD) = 52.6 ± 8.7, % female = 100, % CVD = 0 Placebo N = 288, age (mean, SD) = 52.6 ± 8.7, % female = 100, % CVD = 0	
Interventions	Tibolone 2.5 mg/d or CEE plus MPA 0.625 plus 2.5 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, near and far wall, right and left. Clinical: all-cause mortality	
Notes	Funding received from NV Organon. This study has also been referenced by Bots 2003.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "subjects were randomized". Comment: Probably done because another report including some of the investigators has described adequately generated allocation sequence (Bots 2007)
Allocation concealment?	YES	Comment: No information provided but probably done because another report including some of the same

		authors has described adequately concealed allocation (Bots 2007).
Blinding? Surrogate outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	YES	Of the subjects randomized (tibolone = 290; cee/mpa = 288; placebo = 288), at twelve months, 175 (tibolone = 64; cee/mpa = 57; placebo = 54) subjects were excluded from the analysis. Although not explicit, it seems that the authors have considered only those subjects with data available at the relevant time points. Attritions: while the authors have reported the number of subjects who did not complete the trial, this information is not related to the risk of bias because the analysis is based on the subjects for whom data was available. Number of subjects in analysis: At twelve months, 226 tibolone subjects, 231 cee/mpa subjects and 234 placebo subjects were included in the analysis. Overall completeness of data: At twelve months, 226 (78%) tibolone subjects, 231 (80%) cee/mpa subjects and 234 (81%) placebo subjects completed follow-up. Overall, 691 (approximately 80%) subjects completed the trial. Overall completion rate was ~ 80%.
Incomplete outcome data addressed? Clinical outcome	NO	Of the subjects randomized (tibolone = 290; cee/mpa = 288; placebo = 288), 9 (tibolone = 4; cee/mpa = 4; placebo = 1) were excluded because they were baseline drop-outs. Attritions: 254 (tibolone = 88; cee/mpa = 80; placebo = 86) did not complete the study but were included in the analysis. Number of subjects in analysis: 286 (99%) tibolone subjects, 284 (99%) cee/mpa subjects; 287 (~100%) placebo subjects were included in the analysis. Overall completeness of data: 198 (68%) tibolone subjects, 204 (71%) cee/mpa subjects and 201 (70%) placebo subjects completed the trial. Overall 603 (70%) subjects completed the trial. Overall completion rate was < 80%.
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported. Clinical events monitoring was listed in the study protocol and reported in the results.
Free of other	NO	Comment: The study was funded by a drug company.

bias?		Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Bots 2007 (RADIANCE 2)		
Methods	Randomized, double-blind, trial. Multicentre. 2 year duration.	
Participants	Mixed dyslipidemia. Atorvastatin plus torcetrapib N = 377, age (mean, SD) = 57.9 ± 8.1, % female = 37.1 Atorvastatin N = 375, age (mean, SD) = 56.5 ± 8.2, % female = 34.7,	
Interventions	Atorvastatin dose was established during the atorvastatin-only run in phase. The dose was titrated to achieve LDL concentrations in accordance with NCEP guidelines based on individual cardiovascular risk Torcetrapib 60 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, near and far, right and left Clinical: all-cause mortality, myocardial infarction, stroke	
Notes	Funding received from Pfizer	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "computer-generated permuted block design" was used for randomization. Comment: probably done.
Allocation concealment?	YES	Quote: "the placebo tablets were identical in appearance to the torcetrapib tablets." Quote: "patients were randomized by use of a central scheme". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	UC	Exclusion: Of the subjects who were randomized (atorvastatin + torcetrapib = 377; atorvastatin = 375), 69 (atorvastatin + torcetrapib = 38; atorvastatin = 31) were excluded because they did not meet the analysis strategy: "subjects with at least one measurement of carotid intima-media thickness after baseline". Attritions: Authors do not report the number of subjects who met the analysis criteria but did not complete the study. Number of subjects in efficacy analysis: atorvastatin +

		torcetrapib = 339 (90%) and atorvastatin = 344 (92%). Overall completion rate: the authors do not report the overall completion rates and the completeness of data at 12 months is not known.
Incomplete outcome data addressed? Clinical outcome	UC	Exclusion: None, because all randomized subjects (atorvastatin + torcetrapib = 377; atorvastatin = 375) were included in the analysis strategy which was ITT. Attritions: The overall number of attritions is not provided. Overall completion rate: the authors do not report the overall completion rates and the completeness of data.
Free of selective reporting?	YES	Comment: Prespecified CIMT measurement as described in the methods section was reported in the results. Monitoring of clinical events listed in methods and reported in the results.
Free of other bias?	NO	Comment: The study was funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Byington 1995 (PLAC-II)		
Methods	Randomized, double-blind, placebo controlled trial. Single centre. 3 year duration.	
Participants	Coronary artery disease . Pravastatin N = 75, age (mean) = 62.7, % female = 16, % CVD = 100 Placebo N = 76, age (mean) = 62.4, % female = 13, % CVD = 100	
Interventions	Pravastatin 20 mg/d 3-4 hours post evening meal. If the treatment dose did not achieve LDL-c <100mg/dl, then the dose was increased to 40 mg/d. If the treatment achieved LDL-c <90 mg/dL, then the dose was decreased to 10 mg/dL or Placebo	
Outcomes	Surrogate: common carotid artery, near and far wall, right and left Clinical: all-cause mortality, myocardial infarction,	
Notes	Funding received from Bristol-Myers Squibb and by a Preventive Cardiology Academic Award from the National Heart, Lung and Blood Institute. This study has also been referenced by Crouse 1992.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: insufficient information provided.

Allocation concealment?	YES	Quote: "Placebo participants received matched “dose” capsules to maintain the blind" Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-blind". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: It is likely that key study personnel and participants were blinded
Incomplete outcome data addressed? Surrogate outcome	UC	Exclusions: of the subjects who were randomized (pravastatin = 75; placebo = 76), none were excluded because they all met the analysis strategy which was intention-to-treat. Attritions: the authors do not report the number of subjects who met the analysis strategy but did not have follow-up measurements at the relevant time points. Number of subjects in analysis: 75 pravastatin subjects and 76 placebo subjects were included in the analysis. Overall completeness of data: authors do not provide information on overall completeness of data.
Incomplete outcome data addressed? Clinical outcome	UC	Exclusions: of the subjects who were randomized (pravastatin = 75; placebo = 76), none were excluded because they all met the analysis strategy which was intention-to-treat. Attritions: the authors do not report the number of subjects (if any) who met the analysis strategy but did not have follow-up outcome data. Number of subjects in analysis: 75 pravastatin subjects and 76 placebo subjects were included in the analysis. Overall completeness of data: authors do not provide information on overall completeness of data.
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: Funding was partially received from a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Crouse 2007 (METEOR)	
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 2 year duration.

Participants	Mild to moderate subclinical atherosclerosis and a low Framingham risk score. Rosuvastatin N = 702, age (mean, SD) = 57.0 ± 6.2, % female = 40.0, % CVD = 0 Placebo N = 282, age (mean, SD) = 57.0 ± 6.0, % female = 40.8, % CVD = 0	
Interventions	Rosuvastatin 40 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, near and far wall, right and left. Clinical: all-cause mortality, myocardial infarction	
Notes	Funding received from Astra-Zeneca This study was also referenced by Bots 2009.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "eligible individuals were randomized to either placebo or rosuvastatin in blocks of 7 (5 to the rosuvastatin group and 2 to the placebo group)". Comment: Probably done because another report including some of the investigators has described adequately generated allocation sequence (Bots 2007).
Allocation concealment?	YES	Quote: "blinded study medication was supplied in individual numbered bottles prepared prior to the clinic visits and eligible individuals were allocated study medication sequentially". Comment: probably done because another report including some the authors has described adequately concealed allocation (Bots 2007).
Blinding? Surrogate outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double blind". Comment: It is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions and attritions reported. Exclusions: Of the subjects who were randomized (rosuvastatin = 702; placebo = 282), 108 (rosuvastatin = 78; placebo = 30) were excluded because they did not meet the analysis strategy: subjects who received study medication and a minimum of one CIMT measurement after the baseline reading. Attritions: 138 (rosuvastatin = 94; placebo = 44) subjects did not complete the study but did meet the analysis strategy and were included in the analysis. Number of subjects in analysis: 624 (89%) rosuvastatin subjects and 252 (89%) placebo subjects were included

		in the analysis. Overall completeness of data: the trial was completed by rosuvastatin = 530 (75%) and placebo = 208 (74%) subjects. Completion rate is <80 % overall < 80% in the treatment group.
Incomplete outcome data addressed? Clinical outcome	NO	Exclusions and attritions reported. Exclusions: Of the subjects who were randomized (rosuvastatin = 702; placebo = 282), 3 (rosuvastatin = 2; placebo = 1) were excluded because they did not meet the analysis strategy: subjects "who received at least 1 dose of study medication during the 104-week treatment period." Attritions: 243 (rosuvastatin = 170; placebo = 73) subjects did not complete the study but did meet the analysis strategy and were included in the analysis. Number of subjects in analysis: 700 (~100%) rosuvastatin subjects and 281 (~100%) placebo subjects completed the study. Overall completeness of data: the trial was completed by rosuvastatin = 530 (75%) and placebo = 208 (74%) subjects. Overall completion rate is <80 % < 80% in the treatment group.
Free of selective reporting?	YES	Comment: Pre-specified CIMT measurement as described in the methods section was reported in the results. Monitoring of clinical events listed in methods and reported in the results.
Free of other bias?	NO	Comment: The study was funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Elkeles 1998 (SENDCAP)	
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 5 year duration.
Participants	Type 2 Diabetes. Bezafibrate N = 81, age (mean, SD) = 50.8 ± 8.0, % female = 24.7, duration of diabetes in years (mean, SD) = 4.3 ± 4.3, % CVD = 0 Placebo N = 83, age (mean, SD) = 50.9 ± 8.1, % female = 32.5, duration of diabetes in years (mean, SD) = 5.8 ± 5.7, % CVD = 0
Interventions	Bezafibrate 400 mg/d or Placebo
Outcomes	Surrogate: common carotid artery, far wall Clinical: myocardial infarction
Notes	Funding received from British Heart Foundation, Boehringer Mannheim. Early funding was also received from the North West

	Thames Health Authority.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "a randomization list was prepared by a statistician in advance so that numbers assigned to each treatment would be approximately equal after every 10 subjects". Comment: it can be assumed that the statistician adequately generated the allocation sequence.
Allocation concealment?	YES	Quote: "subjects satisfying entry criteria were given the next consecutive number in a double-blind fashion". Quote: "in addition to the identical appearance of placebo and active study medication..." Comment: probably done.
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "all measurements were carried out by personnel who were blinded to the treatment group of the subjects". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "double-blind". Quote: "all measurements were carried out by personnel who were blinded to the treatment group of the subjects". Comment: probably done
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the subjects who were randomized (bezafibrate = 81; placebo = 83), 37 subjects were excluded because they did not meet the analysis strategy. Although not explicitly stated, it seems that the analysis strategy included all subjects who completed the trial and who had CIMT measurements available at each of the time points considered in the study. Attrition: Because of the analysis strategy chosen by the authors, attrition is not related to risk of bias. Number of subjects included in the analysis 63 (78%) bezafibrate subjects and 64 (77%) placebo subjects were included in the analysis. Overall completeness of data: Overall, 127 (77%) of subjects were included in the analysis. Overall completion rate is < 80% and completion rate in the treatment group <80%.
Incomplete outcome data addressed? Clinical outcome	NO	Exclusions: None, because of the subjects who were randomized (bezafibrate = 81; placebo = 83), all were included in the analysis strategy which was ITT. Attrition: 36 (bezafibrate = 17; placebo = 19) subjects did not complete the study but were included in the

		analysis. Overall completeness of data: Overall, 128 (78%) of subjects completed the study. 64 (79%) bezafibrate subjects completed the study and 64 (77%) placebo subjects completed the study. Overall completion rate is < 80% and completion rate in the treatment group is <80%.
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: This trial was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Hedblad 2001 (BCAPS)		
Methods	Randomized, double-blind, placebo controlled trial. Singe centre. 3 year duration.	
Participants	Carotid plaque, but no symptoms of carotid artery disease Metoprolol N = 396, age (mean, SD) = 61.1 ± 5.6, % female = 55.3, % CVD = 4 Placebo N = 397, age (mean, SD) = 61.9 ± 5.4, % female = 53.8, % CVD = 3.5	
Interventions	Metoprolol 25 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, far wall, right distal. Clinical: all-cause mortality, myocardial infarction, cerebrovascular accident	
Notes	Funding received from Astra-Zeneca Pharmaceuticals, Mölndal, Sweden	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: insufficient information provided.
Allocation concealment?	UC	Comment: insufficient information provided
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "each image was analyzed without knowledge of the subjects' randomization group". Comment: it is likely that all key study personnel and participants were blinded.
Blinding?	YES	Quote: "double-blind". Quote: "The data and safety

Clinical outcome		monitoring board, consisting of independent scientists with expertise in fields relevant to BCAPS, regularly monitored blinded outcome data". Comment: It is likely that all key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	UC	Of the subjects randomized (metoprolol = 396; placebo = 397), 10 subjects (metoprolol = 3; placebo = 7) were excluded. Although not stated explicitly, it seems that the authors provided an analysis based on the number of subjects with one follow-up measurement post baseline. Attrition: this information is not provided. Number of subjects in analysis: 393 metoprolol subjects and 390 placebo subjects were included in the analysis. Overall completeness of data: Authors do not provide information on the completion rate.
Incomplete outcome data addressed? Clinical outcome	UC	no information is provided
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported in the results. Clinical events monitoring is listed in the methods and it is likely that all events that occurred were reported.
Free of other bias?	NO	Comment: The study was funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Hodis 1998 (CLAS)	
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 10+ year duration.
Participants	Coronary artery disease. Colestipol plus niacin plus diet therapy N = 94, age (mean, SD) = 53.9 ± 4.8, % female = 0, % CVD = 100 Placebo N = 94, age (mean, SD) = 54.5 ± 4.8, % female = 0, % CVD = 100
Interventions	Colestid 30 g/d plus niacin 3-12 g/d plus diet therapy. The target diet included intake of <125 mg cholesterol per day, and 22% of total energy from fat. 10% of fat was indicated to be polyunsaturated and 4% was indicated to be saturated. Diet therapy consisted of maintaining a 7-day food record prior to clinic visits for review and

	counseling by a dietician during clinic visits. Subjects also received a 4-5 week evening diet counseling program Placebo plus diet therapy. The target diet included intake of <250 mg cholesterol per day, and 26% of total energy from fat. 10% of fat was indicated to be polyunsatured and 4% was indicated to be saturated. Diet therapy consisted of maintaining a 7-day food record prior to clinic visits for review and counseling by a dietician during clinic visits. Subjects also received a 4-5 week evening diet counseling program	
Outcomes	Surrogate: common carotid artery, far wall, right distal. Clinical: myocardial infarction, cardiovascular mortality	
Notes	Funding received from the National Heart, Lung and Blood Institute. This study was also referenced by Blankenhorn 1987.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "Eligible candidates were randomized to one of the two treatment groups by the statistician". Comment: It can be assumed that the statistician adequately generated the allocation sequence.
Allocation concealment?	UC	Comment: insufficient information provided.
Blinding? Surrogate outcome	YES	Quote: "Study subjects were masked to treatment assignment. Since all subjects experienced the study drugs, it was suspected that subjects knew their treatment assignment…Evaluation of the study outcome measures was carried out by staff and consultants who were masked to treatment group assignment as well as the temporal ordering of angiographic data" (Reference: Blankenhorn 1987). Comment: Although subjects may have known their treatment assignment, this knowledge was unlikely to influence outcome/outcome measurement.
Blinding? Clinical outcome	YES	Quote: ""Study subjects were masked to treatment assignment. Since all subjects experienced the study drugs, it was suspected that subjects knew their treatment assignment" (Reference: Blankenhorn 1987) Quote: "For all patient-reported events, hospital records were obtained for confirmation, and all causes of death were confirmed by hospital records and death certificates. Myocardial infarction was diagnosed by a cardiologist who was

		blinded to treatment assignment and ultrasonographic and angiographic end point measures". Comment: Although subjects not blinded to treatment, lack of blinding unlikely to influence outcome/outcome measurement. Blinding of outcome assessors probably done.
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the subjects who were randomized (colestipol-niacin + diet = 94; placebo + diet=94) (Reference: Blankenhorn 1987), 42 were excluded because they did not meet the analysis strategy: "subjects who had completed the 2-year treatment period and had evaluable coronary and carotid arterial end points". Attritions: The authors do not report the number of subjects who met the analysis strategy, but did not have follow-up measurements at the relevant time points. Number of subjects in the analysis: 73 colestipol-niacin+diet subjects were included in the analysis and 73 diet + placebo subjects were included in the analysis. Overall completeness of data: Overall, 73 (78%) colestipol-niacin+diet subjects completed the trial and 73 (78%) placebo + diet subjects completed the trial. Overall completion rate was 78%. Overall completion rate was <80% and completion rate in the treatment group was <80%.
Incomplete outcome data addressed? Clinical outcome	NO	Exclusions: Of the subjects who were randomized (colestipol-niacin+diet = 94; placebo + diet=94) (Reference: Blankenhorn 1987), 42 were excluded because they did not meet the analysis strategy: "subjects who had completed the 2-year treatment period and had evaluable coronary and carotid arterial end points". Attritions: There were no attritions because of the way in which the analysis strategy was defined. Number of subjects in the analysis: 146 colestipol-niacin+diet subjects were included in the analysis and 73 diet + placebo subjects were included in the analysis. Overall completeness of data: Overall, 73 (78%) colestipol-niacin+diet subjects completed the trial and 73 (78%) placebo + diet subjects completed the trial. Overall completion rate was 78%. Overall completion rate was <80% and completion rate in the treatment group <80%.
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. Clinical events as specified in the methods were reported in the results.

Free of other bias?	YES	Comment: Funding was received from the National Heart, Lung and Blood Institute. It is unlikely that the funding source introduced bias

Hodis 2001 (EPAT)		
Methods	Randomized, double-blind, placebo controlled trial. 2 year duration.	
Participants	Healthy, postmenopausal women, without CHD. Estradiol N = 111, age (mean, SD) = 60.9 ± 6.7, % female = 100, % CVD = 0 Placebo N = 111, age (mean, SD) = 62.1 ± 7.1, % female = 100, % CVD = 0	
Interventions	Estradiol 1 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, far wall, right distal. Clinical: all-cause mortality, myocardial infarction, cerebrovascular accident	
Notes	Funding received from Mead Johnson Laboratories and National Institutes of Health. Parke-Davis Pharmaceutical provided study medication. Additional medications were provided by various other pharmaceutical companies.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "Computer generated random numbers were used to assign participants to unopposed estradiol or placebo". Comment: probably done
Allocation concealment?	YES	Quote: "packets of study medications were prepared in a blinded manner (to both clinical staff and participants) before the start of the study…as a new participant was determined to be eligible for randomization, the next packet in sequence in the appropriate stratum was obtained and recorded. The data coordinating centre monitored adherence to sequential assignment of medication packets". Comment: probably done.
Blinding? Surrogate outcome	YES	Quote: "participants, gynecologists, clinical staff and image analysts were blinded to treatment assignment. The data monitor and data analyst was blinded to treatment assignment until analyses were completed". Comment: probably done.
Blinding?	YES	Quote: "participants, gynecologists, clinical staff and

Clinical outcome		image analysts were blinded to treatment assignment. The data monitor and data analyst was blinded to treatment assignment until analyses were completed". Comment: probably done.
Incomplete outcome data addressed? Surrogate outcome	NO	Attritions and exclusions reported. Exclusions: Of the subjects who were randomized (Estradiol = 111; placebo = 111), 23 (estradiol = 14; placebo = 9) did not meet the analysis strategy: all randomized subjects with baseline and "one or more follow-up" measurements. Attrition: 33 (troglitazone = 14; placebo = 19) withdrew but did meet the analysis criteria and were included in the analysis. Number of subjects in efficacy analysis: 97 (87%) estradiol subjects and 102 (92%) placebo subjects. Overall completeness of data: the trial was completed by estradiol = 83 (75%) and placebo = 83 (75%) subjects. Completion rate is <80% and < 80% in the treatment group.
Incomplete outcome data addressed? Clinical outcome	NO	Attritions and exclusions reported. Exclusions: None, because analysis strategy was ITT and all randomized subjects (Troglitazone = 111; placebo=111) were included. Attritions: 56 (troglitazone = 28; placebo = 28) withdrew from the study. Overall completeness of data: the 2-year trial was completed by troglitazone = 83 (75%) and placebo = 83 (75%) of subjects. Overall, 166 (75%) subjects completed the trial. Completion rate is < 80%.
Free of selective reporting?	YES	Comment: monitoring of clinical events listed in methods and reported in results. Expected CIMT measurement as described in the method was reported in results.
Free of other bias?	YES	Quote: " investigator initiated and conducted. The authors were solely responsible for the design, conduct, data collection, data management statistical analyses, and data interpretation. The funding source played no role in these functions. Data were not reported to the funding source, and the funding source had no role in deciding whether or where the study would be submitted for publication". Comment: Drug company provided funding but authors indicate that it had no role in the design, conduct, analysis and reporting of the study. This was probably true.

Hodis 2002 (VEAPS)		
Methods	Randomized, double-blind, placebo controlled trial. 3 year duration.	
Participants	LDL-c ≥ 3.37mmol/L and no clinical signs/symptoms of CVD. Vitamin E N = 177, age (mean, SD) = 55.7 ± 9.2, % female = 52.5, % CVD = 0 Placebo N = 176, age (mean, SD) = 56.7 ± 8.6, % female = 51.2, % CVD = 0	
Interventions	Vitamin E 400 IU/d Placebo	
Outcomes	Surrogate: common carotid artery, far wall, right distal. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, cerebrovascular accident	
Notes	Funding received from National Institute on Aging and Hoffmann-La Roche, Inc.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "computer-generated random numbers were used to assign participants". Comment: probably done
Allocation concealment?	UC	Comment: insufficient information provided
Blinding? Surrogate outcome	YES	Quote: "participants, clinical staff, imaging specialists, data monitor, and statistical analysts were blinded to treatment assignment". Comment: probably done
Blinding? Clinical outcome	YES	Quote: "participants, clinical staff, imaging specialists, data monitor, and statistical analysts were blinded to treatment assignment". Comment: probably done
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the randomized subjects (vitamin E = 177; placebo = 176), 21 (vitamin E = 15; placebo = 6) were excluded because they did not meet the analysis strategy: "those who had a baseline and at least one follow-up measurement" of CIMT. Attrition: 74 subjects did not complete the study, but met the analysis strategy and were included in the analysis. Information on group specific attritions is not provided by the authors. Number of subjects in analysis: 162 (92%) vitamin E subjects and 170 (97%) placebo subjects were included in the analysis. Overall completeness of data: 258 (73%) of subjects completed the trial. Authors do not provide group-specific completion rates. Overall completion rate is < 80%.

Incomplete outcome data addressed? Clinical outcome	NO	Exclusions: None, because all randomized subjects (vitamin E = 177; placebo = 176), were included in the analysis which was ITT. Attrition: 74 subjects did not complete the study. Information about attrition in each group is not provided by the authors. Overall completeness of data: 258 (73%) of subjects completed the trial. Authors do not provide group-specific completion rates. Overall completion rate is < 80%
Free of selective reporting?	YES	Comment: CIMT measurements as described in the method were reported. Clinical events monitoring listed in methods, and it is likely that all events that occurred were reported by the authors
Free of other bias?	NO	Comment: The study was partially funding by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Hodis 2006 (TART)		
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 2 year duration.	
Participants	Diabetes. Troglitazone N = 149, age (mean, SD) = 52.4 ± 9.4, % female = 66.9, duration of diabetes in years (mean, SD) = 9.8 ± 6.2, % CVD = <10 (total study subjects) Placebo N = 150, age (mean, SD) = 52.6 ± 8.7, % female = 67.9, duration of diabetes in years (mean, SD) = 9.7 ± 6.4	
Interventions	Troglitazone 400 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, far wall, right distal. Clinical: myocardial infarction, cerebrovascular accident	
Notes	Funding received from Parke-Davis Pharmaceutical Research and National Institutes of Health. Parke-Davis Pharmaceutical provided study medication.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: insufficient information provided.
Allocation concealment?	UC	Comment: insufficient information provided.

Blinding? Surrogate outcome	YES	Quote: "participants, investigators, image analysts, and clinical and data coordinating centre staff were blinded to treatment assignment". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "participants, investigators, image analysts, and clinical and data coordinating centre staff were blinded to treatment assignment". Comment: probably done.
Incomplete outcome data addressed? Surrogate outcome	NO	Attritions and exclusions reported. Exclusions: Of the subjects who were randomized (Troglitazone = 149; placebo=150), 23 (troglitazone = 7; placebo = 16) did not meet the analysis strategy: subjects "with baseline and at least one follow-up" measurement. Attrition: 42 (troglitazone = 16; placebo = 26) withdrew but did meet the analysis criteria and were included in the anlaysis. Number of subjects in efficacy analysis: 142 (95%) troglitazone treated subjects and 134 (89%) placebo treated and were evaluable. Overall completeness of data: The 2-year trial was completed by troglitazone = 126 (85%) and placebo = 108 (72%) of subjects. Completion rate is < 80% overall.
Incomplete outcome data addressed? Clinical outcome	NO	Attritions and exclusions reported. Exclusions: None, because analysis strategy was ITT and all randomized subjects (Troglitazone = 149; placebo=150) were included. Attritions: 65 (troglitazone = 23; placebo = 42) withdrew from the study. Overall completeness of data: The 2-year trial was completed by troglitazone = 126 (85%) and placebo = 108 (72%) of subjects. Overall, 234 (78%) subjects completed the trial. Completion rate is < 80%.
Free of selective reporting?	YES	Comment: Prespecified CIMT measurement as described in the methods section was reported in the results. Monitoring of clinical events listed in methods and reported in the results.
Free of other bias?	NO	Comment: this trial was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Hodis 2009 (BVAIT)	
Methods	Randomized, double-blind, placebo controlled trial. 3 year duration.
Participants	Men and postmenopausal women ≥ 40 years old with fasting tHcy 8.5 mol/L and no clinical signs/symptoms of CVD

	B vitamins N = 254, age (mean, SD) = 61.7 ± 10.1, % female = 39.0, % CVD = 0 Placebo N = 252, age (mean, SD) = 61.1 ± 9.6, % female = 39.0, % CVD = 0	
Interventions	Folic acid plus vitamin B12 plus vitamin B6 5 mg/d plus 0.4 mg/d plus 50 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, far wall, right distal. Clinical: all-cause mortality	
Notes	Funding received from National Institute on Aging, National Institutes of Health. The vitamin supplements and placebo were received from Leiner Health Products.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "computer-generated random numbers were used to assign participants". Comment: probably done
Allocation concealment?	UC	Comment: insufficient information provided.
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "participants, clinical staff, imaging specialists, and data monitors were masked to treatment assignment". Quote: "scans were analyzed without knowledge of treatment assignment". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "double blind". Quote: "participants, clinical staff, imaging specialists, and data monitors were masked to treatment assignment". Comment: It is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the subjects randomized (b vitamins = 254; placebo = 252), 16 subjects (b vitamins = 6; placebo = 10) were excluded because they did not meet the analysis strategy: "all participants who had carotid ultrasonography at baseline and at least one follow-up visit". Attritions: authors do not provide information on the number of subjects who did meet the analysis strategy but did not complete the study. Number of subjects in analysis: 248 (98%) B-vitamin subjects and 242 (96%) placebo subjects were included in the analysis. Overall completeness of data: 143 (56%) B-vitamin subjects and 137 (54%) placebo subjects completed the trial. Overall, 280 (55%) subjects completed the trial. Overall completion rate is <80% and <80% in the treatment group.

Incomplete outcome data addressed? Clinical outcome	NO	Exclusions: Of the subjects randomized (b vitamins = 254; placebo = 252), none were excluded because the analysis included all randomized subjects and was ITT. Attritions: 226 (b vitamins = 111; placebo = 115) subjects did not complete the trial but were included in the analysis. Overall completeness of data: 143 (56%) B-vitamin subjects and 137 (54%) placebo subjects completed the trial. Overall, 280 (55%) subjects completed the trial. Overall completion rate was < 80% and completion rate in treatment group was < 80%
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported in the results. Clinical events monitoring was listed in the methods and it is likely that all outcomes that occurred were reported
Free of other bias?	YES	Comment: The trial was funded by the National Institute on Aging, National Institutes of Health and a vitamins company provided the placebo and vitamin pills. It is unlikely that the funding source introduced bias

Howard 2008 (SANDS)	
Methods	Randomized, open-label, blind-to-endpoint, trial. Multicentre. 3 year duration.
Participants	Type 2 diabetes. Aggressive treatment N = 252, age (mean, range) = 55.0 (54 – 57), % female = 66.3, duration of diabetes in years (mean) = 9.2 % CVD = 0 (49 of the total study subjects had CVD at baseline, but were excluded from all analyses) Standard treatment N = 247, age (mean, range) = 57.0 (56.0 – 58.0), % female = 64.8, duration of diabetes in years (mean) =8.7, % CVD = 0
Interventions	Aggressive treatment: for hypertension management, the primary goal of therapy was SBP 115 mm Hg or lower and secondary goal was DBP of 75 mm Hg or lower. For lipid management, the goals of therapy were LDL-c 70 mg/dL or lower and non-HDL-c 100 mg//Dl or lower. Standard treatment: for hypertension management, the primary goal of therapy was SBP 130 mm Hg or lower and secondary goal was DBP of 85 mm Hg or lower. For lipid management, the goals of therapy were LDL-c 100 mg/dL or lower and non-HDL-c 130 mg//Dl or lower.

	For hypertension management, treatment regimen for SBP was as follows: Step 1 included ACE inhibitors or ARBs in the case of intolerance to ACE inhibitors. Step 2 included hydrochlorothiazide. Step 3 included CCBs. Step 4 included beta-blockers. Step 5 included alpha-blockers or other vasodilators. DBP management was based on physician discretion. For lipid management, treatment regimen for LDL-c was as follows: lifestyle modification. Statin treatment was initiated if lifestyle modification was unsuccessful. Combination treatment with ezetimibe was initiated if treatment continued to remain unsuccessful. Non-HDL-c management was achieved with fish oil, fenofibrate or niacin.	
Outcomes	Surrogate: common carotid artery, far wall, right and left distal. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from National Heart, Lung and Blood Institute and National Institutes of Health. Study medications were donated by First Horizon Pharmacy, Merck and Co and Pfizer Inc. This study was also referenced by Russell 2006.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "the urn method" was used to randomize the subjects. Comment: probably done
Allocation concealment?	UC	Comment: insufficient information provided
Blinding? Surrogate outcome	YES	Quote: "field clinicians who delivered the intervention were not blinded; however, research assistants, ultrasound technicians and readers, and all core laboratory personnel were blinded to study assignment". Comment: it is likely that lack of blinding did not influence outcomes or outcome measurements and that most key study personnel involved in outcome assessment were blinded. Because of the nature of the intervention, it would be difficult to blind the field clinicians.
Blinding? Clinical outcome	YES	Quote: "field clinicians who delivered the intervention were not blinded; however, research assistants, ultrasound technicians and readers, and all core laboratory personnel were blinded to study assignment". Quote: "Medical records for all hospitalizations and outpatient coronary revascularization procedures were reviewed centrally by a panel of physician adjudicators

		blinded to treatment assignment". Comment: Probably done. Also, it is unlikely that lack of blinding would influence outcomes or outcome measurement and outcome adjudicators were blinded.
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: None, because all randomized subjects (aggressive treatment = 252; standard treatment = 247) were included in the analysis which was ITT. Attrition: 46 subjects (aggressive treatment = 28; standard treatment = 18) were not available for the final CIMT measurement but were included in the analysis. Information about attritions at 12 month measurement time was not available. Number of subjects in the analysis: 252 (100%) aggressive treatment subjects and 247 (100%) standard treatment subjects were included in the analysis. Overall completeness of data: For the final CIMT measurements, aggressive treatment = 224 (89%) and standard treatment = 229 (93%) were included in the trial. Overall completion rate for CIMT measurements was > 80% and completion rate for CIMT measurements in the treatment group was > 80%.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: None, because all randomized subjects (aggressive treatment = 252; standard treatment = 247) were included in the analysis which was ITT. Attrition: authors indicate that 4 subjects were lost to follow-up and 8 (aggressive treatment = 3; standard treatment = 5) subjects did not complete the study. The authors do not indicate whether the 4 subjects who were lost to follow-up were a subset of the 8 who did not complete the study. Information on clinical events was known for all of these subjects. Overall completeness of data: 252 (100%) aggressive treatment subjects and 247 (100%) standard treatment subjects provided complete data. Overall completion rate was > 80% and completion rate in the treatment group was > 80%
Free of selective reporting?	YES	Comment: CIMT measurements as described in methods were reported. Clinical events monitoring was listed in methods and reported in results.
Free of other bias?	YES	Comment: Funding was provided by the National Heart, Blood and Lung Institute and pharmacologic agents were donated by several drug companies. It is unlikely that the funding source introduced bias

Kastelein 2007 (RADIANCE 1)		
Methods	Randomized, double-blind, trial. Multicentre. 2 year duration.	
Participants	Familial hypercholesterolemia Atorvastatin plus torcetrapib N = 450, age (mean, SD) = 46.8 ± 12.0, % female = 52.4 Atorvastatin N = 454, age (mean, SD) = 45.2 ± 12.9, % female = 48.9	
Interventions	Atorvastatin dose was established during the atorvastatin-only run in phase of 6-14 weeks. The dose was titrated to achieve LDL concentrations in accordance with NCEP guidelines based on individual cardiovascular risk. The medication was administered at 20, 40 or 80 mg and titrated at 4 week intervals for a maximum of 3 visits. The titrated daily dose was found on average to be 56.6 mg in each group. Torcetrapib 60 mg/d or Placebo	
Outcomes	Surrogate: common carotid artery, near and far wall, right and left. Clinical: cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from Pfizer. This study was also referenced by Kastelein 2007.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "patients were randomly assigned". Comment: probably done because another report including some of the investigators and following the same design has described adequately generated allocation sequence (Kastelein 2007, Bots 2007).
Allocation concealment?	YES	Comment: No information provided but probably done because another report including some of the same authors has described adequately concealed allocation (Kastelein 2007, Bots 2007).
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "patients and study personnel were unaware of study-group assignments, laboratory measurements, and carotid imaging findings". Comment: probably done
Blinding? Clinical outcome	YES	Quote: "double blind". Comment: It is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate	YES	Exclusions: Of the subjects who were randomized (atorvastatin + torcetrapib = 450; atorvastatin = 454), 54 (atorvastatin + torcetrapib = 27; atorvastatin = 27) were excluded because they did not meet the analysis strategy:

outcome		subjects who "underwent ultrasonography of the carotid artery at least once both at baseline and at follow-up". Attritions: unclear. Number of subjects in analysis: 423 (94%) atorvastatin + torcetrapib subjects and 427 (94%) of atorvastatin subjects. Overall completeness of data: the trial was completed by atorvastatin+torcetrapib = 387 (86%) and atorvastatin = 391 (86%) of subjects. Overall completion rate is > 80% and > 80% in the treatment group.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: None, because all randomized subjects (atorvastatin + torcetrapib = 450; atorvastatin = 454) were included in the analysis strategy which was ITT. Attritions: 126 (atorvastatin + torcetrapib = 63; atorvastatin = 63) did not complete the trial but were included in the analysis strategy. Overall completeness of data: the trial was completed by atorvastatin+torcetrapib = 387 (86%) and atorvastatin = 391 (86%) of subjects. Overall completion rate is > 80% overall and > 80% in the treatment group.
Free of selective reporting?		Comment: Prespecified CIMT measurements as described in the method section were reported in the results. Clinical events monitoring mentioned in protocol, and it is likely that all events that occurred were reported.
Free of other bias?	NO	Comment: The study was supported by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Kastelein 2008 (ENHANCE)	
Methods	Randomized, double-blind trial. Multicentre. 2 year duration.
Participants	Familial hypercholesterolemia. Simvastatin plus ezetimibe N = 357, age (mean, SD) = 46.1 ± 9.0, % female = 46.5 Simvastatin plus placebo N = 363, age (mean, SD) = 45.7 ± 10.0, % female = 50.7,
Interventions	Simvastatin 80 mg/d plus ezetimibe 10 mg/d or Simvastatin 80 mg/d plus placebo
Outcomes	Surrogate: combined common carotid artery, carotid bulb and internal carotid artery, far wall, right and left. Clinical: cardiovascular mortality, myocardial infarction, stroke
Notes	The trial was supported by Merck and Schering-Plough.

Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "randomization, which was based on computer-generated codes". Comment: probably done
Allocation concealment?	YES	Quote: "randomization...provided to the clinical centers by a central randomization service". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "the study consisted of three periods: a screening phase, a single-blind placebo run-in period…a double-blind study period with a scheduled duration of 24 months". Quote: "Sonographers are also blinded to treatment assignment" (Reference: Kastelein 2005). Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind". Quote: "the study consisted of three periods: a screening phase, a single-blind placebo run-in period…a double-blind study period with a scheduled duration of 24 months". Comment: It is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions and attritions reported. Exclusions: Of the subjects who were randomized (simvastatin + ezetimibe = 357; simvastatin = 363), 78 (simvastatin + ezetimibe = 35; simvastatin = 43) were excluded because they did not meet the criteria of the analysis strategy. The authors have not made the criteria clear, although it seems that subjects were required to have at least one follow-up measurement post baseline. Attritions: 27 (simvastatin + ezetimibe = 6; simvastatin = 21) did not complete study, but did meet the criteria for the analysis strategy and were included in the analysis. Number of subjects in the efficacy analysis: 322 (90%) simvastatin + ezetimibe subjects and 320 (88%) simvatatin subjects were included in the analysis. Of note, the data used for the analyses in this review were based on observed data at each of the time points, and not on the data meeting the analysis strategy; however, the number of subjects at each of the CIMT measurement time points was not provided, and for this reason, the number of subjects used for the efficacy analysis was taken as the numerator. Overall completeness of data: the trial was completed by simvastatin + ezetimibe = 316 (86%) and simvstatin =

		299 (82%) of subjects; Completion rate is > 80% overall and for the treatment group.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: None, because all randomized subjects (simvastatin+ezetimibe = 357; simvastatin = 363) were included in the analysis which was ITT . Attritions: 105 (simvastatin + ezetimibe = 41; simvastatin = 64) did not complete study, but did meet the criteria for the analysis strategy and were included in the analysis. Overall completeness of data: the trial was completed by simvastatin + ezetimibe = 316 (86%) and simvstatin = 299 (82%) of subjects. Overall, 615 (85%) subjects completed the trial. Completion rate is > 80% overall and for the treatment group.
Free of selective reporting?	YES	Comment: Prespecified CIMT measurement as described in the methods section was reported in the results. Monitoring of clinical events listed in methods and reported in the results.
Free of other bias?	NO	Comment: The study was sponsored by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Lonn 2009 (STARR)	
Methods	Randomized, double-blind, placebo controlled trial with a 2X2 factorial design. Multicentre. 3 year duration.
Participants	Impaired glucose tolerance and/or impaired fasting glucose. Ramipril N = 715, age (mean, SD) = 54.2 ± 10.9, % female = 58.2, % CVD = 0 Ramipril placebo N = 710, age (mean, SD) = 54.5 ± 11.0, % female = 54.3, % CVD = 0 Rosiglitazone N = 709, age (mean, SD) = 53.9 ± 10.8, % female = 54.3, % CVD = 0 Rosiglitazone placebo N = 716, age (mean, SD) = 54.6 ± 10.9, % female = 56.2, % CVD = 0
Interventions	Ramipril 5 mg/d. After 2 months, the dose was increased to 10 mg/d. After 1 year, the dose was increased to 15 mg/d or Ramipril placebo. Rosiglitazone 4 mg/d. After 2 months, the dose was increased to 8 mg/day or Rosiglitazone placebo

Outcomes	Surrogate: common carotid artery, far wall, left and right. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from the Heart and Stroke Foundation of Canada, Sanofi-Aventis, GlaxoSmithKline, and King Pharmaceuticals	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "a computerized telephone randomization system" was used for randomization. Comment: probably done.
Allocation concealment?	YES	Quote: "concealed, computerized telephone randomization system". Comment: probably done
Blinding? Surrogate outcome	YE S	Quote: "double-blind". Quote: "three certified readers unaware of treatment assignment, performed all measurements". Comment: it is likely that all key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind". Quote: "all events were adjudicated by cardiologists and endocrinologists blinded to the study medications (Reference: Dagenais 2008). Comment: it is likely that all key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Of the subjects who were randomized (ramipril = 715; placebo = 701), 169 subjects (ramipril = 78; placebo = 91) were excluded because they did not meet the analysis strategy: "participants with at least 2 adequate CUS examinations after the baseline scan, to allow reliable slope estimates". Attritions: Authors do not report the number of subjects who were evaluable but did not complete the trial. Number of subjects in analysis: 637 (89%) ramipril subjects and 619 (87%) placebo subjects were included in the analysis. Overall completeness of data: 1418 (approx 100%) subjects completed the trial. Completion rate in each group is not provided. Overall completion rate > 80%
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: None, because all randomized subjects (ramipril = 715; placebo = 701) were included in the analysis which was ITT. Attritions: 7 subjects did not complete the trial. Authors do not provide information on the group-specific attritions. Overall completeness of data: 1418 (approx 100%) subjects completed the trial.

		Completion rate in each group is not provided. Overall completion rate > 80%
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported in the results. Clinical events as specified in the methods (Reference:Dagenais 2007 were reported in the results.
Free of other bias?	NO	Comment: The study was partially funded by several drug companies. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Magliano 2006 (MAVET)		
Methods	Randomized, double-blind, placebo controlled trial. 4 year duration.	
Participants	Cigarette smokers. Vitamin E N = 205, age (mean, SD) = 63.5 ± 6.1, % female = 56.6, , % CVD = ≥ 6.3 Placebo N = 204, age (mean, SD) = 63.5 ± 5.8, % female = 52.5, % CVD ≥ 6.9	
Interventions	Vitamin E 500 IU/day or Placebo	
Outcomes	Surrogate: common carotid artery, right. Clinical: all-cause mortality, cardiovascular mortality	
Notes	Funding received from National Health and Medical Research Council of Australia.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "the randomization schedule…used an allocation scheme with a random permuted block size of eight". Comment: Probably done
Allocation concealment?	YES	Quote: "The randomization schedule, which was kept at a remote site". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-blind" Quote: "neither study staff, nor participants were aware of treatment allocation". Quote: "images were analyzed,…, by a single assistant blinded to both the treatment allocation and visit number". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "double-blind" Quote: "neither study staff, nor participants were aware of treatment allocation". Comment: it is likely that all key study personnel and participants were blinded.

Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Of the subjects who were randomized (vitamin E = 205; placebo = 204), at twelve months, 38 (vitamin E = 19; placebo = 19) were excluded because they did not meet the analysis strategy. The authors based the analysis on the number of subjects with data available at the relevant time points. Attritions: because the analysis was performed on the number of subjects with available data, attrition was not related to risk of bias for CIMT measurements. Number of subjects in analysis: At 12 months, 186 (91%) vitamin E subjects and 185 (91%) placebo subjects were included in the analysis. Overall completeness of data: At twelve months 371 (91%) of subjects were included in the analysis. Overall completion rate was > 80% at 12 months and completion rate in the treatment group >80% at 12 months.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: Of the subjects who were randomized (vitamin E = 205; placebo = 204), none were excluded because all randomized subjects were included in the analysis which was ITT. Attritions: 76 subjects (vitamin E= 34; placebo = 42) subjects did not complete the trial but were included in the analysis. Overall completeness of data: 171 (83%) vitamin E subjects and 162 (79%) placebo subjects completed the trial. Overall, 333 (81%) subjects completed the trial. Overall completion rate was > 80%.
Free of selective reporting?	YES	Comment: CIMT as specified in the methods section was reported in the results. Clinical events monitoring was listed in the methods and it is likely that all clinical events that occurred were reported.
Free of other bias?	YES	Comment: The study was funded by the National Health and Medical Research Council of Australia. It is unlikely that the funding source introduced bias

Mazzone 2006 (CHICAGO)	
Methods	Randomized, double-blind trial. Multicentre. 6 year duration.
Participants	Type 2 Diabetes. Pioglitazone N = 232, age (mean, SD) = 58.9 ± 7.8, % female = 36.6, duration of diabetes in years (mean, SD) = 7.8 ± 7.2, % CVD = ≥ 6.9 (based on data for baseline myocardial infarction) Glimepiride N = 230, age (mean, SD) = 59.8 ± 8.1, % female = 36.0,

	duration of diabetes in years (mean, SD) = 7.5 ± 7.0, % CVD = ≥ 15.1 (based on data for baseline myocardial infarction	
Interventions	Pioglitazone 15-45 mg/d Glimepiride 1-4 mg/d	
Outcomes	Surrogate: common carotid artery, right and left Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from Takeda Pharmaceuticals North America Inc, Lincolnshire, Ill. A grant from the National Heart, Lung and Blood Institute was also used towards this study.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: insufficient information provided.
Allocation concealment?	UC	Comment: insufficient information provided
Blinding? Surrogate outcome	YES	Quote: "double-blind" Quote: "images were blinded according to visit and treatment group and forwarded to a single blinded expert reviewer". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "double-blind" Quote: "A clinical end point committee adjudicated events contributing to the composite clinical end points in a blinded fashion". Comment: probably done
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the subjects who were randomized (pioglitazone = 232; glimepiride = 230), 101 (pioglitazone = 57; glimepiride = 44) were excluded because they did not meet the analysis strategy: subjects who received the intervention as assigned and "who had both a baseline and one qualifying post-baseline CIMT image". Attritions: At twelve months, there were no subjects who did not have follow-up measurements. Number of subjects in analysis: 175 pioglitazone subjects and 186 glimepiride subjects were included in the analysis. Overall completeness of data: For the 12 month data, 175 (75%) pioglitzone and 186 (81%) glimepiride subjects completed the trial. Overall, 361 (78%) of subjects completed follow-up at 12 months. Overall completion rate was < 80% at 12 months.
Incomplete outcome data	NO	Exclusions: Of the subjects who were randomized (pioglitazone = 232; glimepiride = 230), 4 (pioglitazone

addressed? Clinical outcome		= 2; glimepiride = 2) were excluded because they did not meet the analysis strategy: subjects who received the intervention as assigned. Attritions: 135 (pioglitazone = 72; glimepiride = 63) subjects did not complete the study but did meet the analysis strategy and were included in the analysis. Number of subjects in analysis: 230 pioglitazone subjects and 228 glimepiride subjects were included in the analysis. Overall completeness of data: Overall 158 (68%) pioglitazone subjects and 165 (72%) glimepiride subjects completed the study. Overall 323 (70%) of subjects completed the study. Overall completion rate < 80% and completion rate is <80% in the treatment group.
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: This trial was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Meuwese 2009 (CAPTIVATE)		
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 2 year duration.	
Participants	Familial hypercholesterolemia. Pactimibe N = 448, age (mean, SD) = 55.5 ± 8.5, % female = 36.6, % CVD = 97 (refers to medical history of CVD) Placebo N = 443, age (mean, SD) = 54.7 ± 8.5, % female = 41.1, % CVD = 97 (refers to medical history of CVD)	
Interventions	Pactimibe 100 mg/d or Placebo	
Outcomes	Surrogate: common carotid, carotid bifurcation, internal carotid, near and far walls, right and left. Clinical: cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from Daiichi Sankyo Pharma Development.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "randomization was performed by using random permuted blocks within strata". Comment: Probably done because another report including some of the investigators has described adequately generated allocation sequence (Bots 2007).

Allocation concealment?	YES	Comment: No information provided but probably done because another report including some of the same authors has described adequately concealed allocation (Bots 2007).
Blinding? Surrogate outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Of the subjects who were randomized (pactimibe = 448; placebo = 444), 176 subjects were excluded (pactimibe = 101; placebo = 75) because they did not meet the analysis strategy: subjects for whom scans were available "at least 40 weeks apart". Number of attritions: At 12 months, there were no subjects who met the analysis strategy but did not complete the trial. Number of subjects in analysis: 347 pactimibe subjects (77%) and 369 (83%) placebo subjects were included in the analysis. Overall completeness of data: For the 12 month data on the surrogate outcome, there were pactimibe = 347 (77%) and placebo = 369 (83%) subjects in the trial. For the 12 month data, the overall completion rate was 80%.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: Of the subjects who were randomized (pactimibe = 448; placebo = 444), 11 subjects were excluded (pactimibe = 5; placebo = 6) because they did not meet the analysis strategy: subjects who were not lost to follow-up, subjects who had valid lipid values and postbaseline cardiovascular endpoints. Number of attritions: 82 (pactimibe = 43; placebo = 39) did not complete the study. Number of subjects in analysis: 443 pactimibe subjects (99%) and 438 (99%) placebo subjects were included in the analysis. Overall completeness of data: 400 (89%) pactimibe subjects and 399 (90%) placebo subjects completed the study. Overall completion rate was 89%. Completion rate in pactimibe group > 80% and completion rate in placebo group was >80. Overall completion rate was > 80% and completion rate in the treatment group > 80%

Free of selective reporting?	YES	Comment: prespecified CIMT measurements as described in the methods section were reported in the results. Clinical events monitoring listed in the methods and it is likely that all events that occurred were reported.
Free of other bias?	NO	Comment: The study was funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Nanayakkara 2007 (ATIC)		
Methods	Randomized, double-blind, placebo controlled trial. Multicentre. 2 year duration.	
Participants	Mild to moderate chronic kidney disease Pravastatin plus vitamin E plus homocysteine lowering N = 47, age (mean, SD) = 54.0 ± 11.0, % female = 48.9, % CVD = 0 Placebo N = 46, age (mean, SD) = 52.0 ± 13.0, % female = 37.0, % CVD = 0	
Interventions	Pravastatin 40 mg/d. After 6 months vitamin E (alpha-tocopherol acetate) 300 mg/d was added. After an additional 6 months, homocysteine lowering therapy was added and consisted of folic acid 5 mg/d, pyridoxine hydrochloride 100 mg/d, cyanacobalamin 1 mg/d or Placebo. After 6 months an additional placebo was administered. After an additional 6 months matching placebos were administered.	
Outcomes	Surrogate: carotid, right. Clinical: cardiovascular mortality, myocardial infarction	
Notes	Funding received from The Dutch Kidney Foundation and Bristol-Myers Squibb	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "randomization was carried out centrally by means of a computer generated sequence involving randomized blocks of 4". Comment: probably done
Allocation concealment?	YES	Quote: "Concealed enveloped were kept by 1 hospital pharmacist". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "all ultrasound measurements at each visit were performed by a single observer who was blinded to the treatment allocation". Comment: it is likely that key personnel and study

		participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double blind". Comment: it is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Of the subjects who were randomized (pravastatin = 47; placebo = 46), at 12 months, 13 subjects (pravasatin = 7; placebo = 6) were excluded because they did not meet the analysis strategy. Although not explicitly stated, it seems that the authors based the analysis on the number of subjects for which observed data was available at each of the time points. Number of attritions: While the authors report the number of subjects who did not complete the trial, this information is not related to risk of bias in CIMT measurements because it seems that the authors based the CIMT analysis on the number of subjects for whom data was available. Number of subjects in analysis: For the 12 month data, there were pravastatin = 40 (85%) and placebo = 40 (87%) subjects in the trial. At 12 months, the overall completion rate was 86%. At 12 months, the overall completion rate and the completion rate in the treatment group was > 80%.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: None, because all randomized subjects (pravastatin = 47; placebo = 46) were considered in the analysis which was ITT. Attritions: 19 (pravastatin = 11; placebo = 8) subjects did not complete the study. Overall completeness of data: 36 (77%) pravastatin subjects and 38 (83%) placebo subjects completed the study. Overall completion rate was 80% (79.57%). Completion rate in pravastatin group < 80%.
Free of selective reporting?	YES	Comment: Prespecified CIMT measurements as described in the method section were reported in the results. Clinical events monitoring not mentioned in methods, but it is likely that all events that occurred were reported
Free of other bias?	NO	Comment: The trial was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Salonen 1995 (KAPS)	
Methods	Randomized, double-masked, placebo controlled trial. Single centre. 3

	year duration.	
Participants	Diabetes. Pravastatin N = 224, age (mean, SD) = 57.3 ± 4.3, % female = 0, % CVD = 8.9 (based on baseline myocardial infarction) Placebo N = 223, age (mean, SD) = 57.5 ± 4.4, % female = 0, % CVD = 6.3 (based on baseline myocardial infarction)	
Interventions	Pravastatin 40 mg/d at bedtime or Placebo	
Outcomes	Surrogate: common carotid artery, far wall, right and left. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ and the Academy of Finland.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "the randomization scheme was generated by a KAPS biostatistician". Comment: it can be assumed that the statistician adequately generated the allocation sequence.
Allocation concealment?	YES	Quote: "placebo and pravastatin tablets looked identical". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-masked". Comment: It is likely that key study personnel and participants were blinded. In a later study by the same authors, blinding was clearly described. (Salonen 2000)
Blinding? Clinical outcome	YES	Quote: "double-masked". Comment: It is likely that key study personnel and participants were blinded. In a later study by the same authors, blinding was clearly described. (Salonen 2000)
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Of the subjects randomized (pravastatin = 224; placebo = 223), 23 (pravastatin = 12; placebo = 11) were excluded because they did not meet the analysis strategy: all randomized subjects with "any follow-up measurements regardless of their compliance with treatment." Attritions: Overall, 16 (4%) subjects met the analysis strategy but did not complete the trial. The authors do not report the number of subjects, if any, who met the analysis strategy but did not have measurements at the relevant follow-up time. Number of subjects in analysis: 212 pravastatin subjects and 212 placebo subjects were included in the analysis. Overall

		completeness of data: Overall 208 (93%) pravastatin subjects and 200 (90%) placebo subjects completed the trial. Overall, 408 (91%) of subjects completed the trial. Overall completion rate is > 80% and completion rate in the treatment group >80%.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: Of the subjects randomized (pravastatin = 224; placebo = 223), although not explicitly stated, it seems that none of the randomized subjects were excluded from the analysis because it seems that the analysis was ITT. Attritions: Overall, 39 subjects (pravastatin = 16; placebo = 23) did not complete the trial. Overall completeness of data: Overall 208 (93%) pravastatin subjects and 200 (90%) of placebo subjects completed the trial. Overall, 408 (91%) of subjects completed the trial. Overall completion rate is >80% and completion rate in each group is >80%
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. All important clinical events reported in the results.
Free of other bias?	NO	Comment: Funding was received from a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Salonen 2003 (ASAP)	
Methods	Randomized, double-blind (for first 3 years), placebo controlled trial. 6 year duration.
Participants	Hypercholesterolemia Vitamin E plus vitamin C N = 390, age (range) = 45 – 69 (total study subjects), % female = 50.8, , % CVD = 15.6 Placebo N = 130, , % female = 50.8, % CVD = 12.3
Interventions	Vitamin E plus vitamin C: Subjects in the supplements group received one of three supplement regimens for the first three years, followed by the same regimen during the last 3 years of the study. For the first three years, subjects received (i) d-alpha tocopherol 91 mg twice per day after a meal. After 3 years, subjects received d-alpha tocoperol plus slow release ascorbic acid in a single tablet or (ii) slow-release ascorbic acid 250 mg twice per day after a meal. After 3 years subjects received, d-alpha tocoperol plus slow release ascorbic acid in a single tablet or (iii) d-alpha tocoperol plus slow release ascorbic acid in a single tablet.

	During the last three years, all subjects received d-alpha tocoperol plus slow release ascorbic acid in a single tablet. Placebo: Subjects in the placebo group received placebo for 6 years.	
Outcomes	Surrogate: common carotid artery, far wall, right and left. Clinical: all-cause mortality	
Notes	Funding received from the Academy of Finland. Vitamin supplements were received by Ferrosan A/S, Denmark. This study was also reported by Salonen 2000.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "women were randomized" Comment: Probably done because another report from the authors has provided a description of an adequately generated allocation sequence (Salonen 1995)
Allocation concealment?	YES	Quote: "all tablets were identical in appearance, size and color". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-masked". Quote: "all IMT measurements were carried out by one very experienced technician who was blinded to the supplementation status". Comment: probably done
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: It is likely that key study personnel and participants were blinded
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Of the subjects who were randomized (supplements = 390; placebo = 130), 80 were excluded (supplements= 55; placebo = 25) because they did not meet the analysis strategy: "all subjects for whom ultrasound examination was available at the end" (referring to end of 6 yr f/u). Attritions: the authors do not provide the number of subjects for whom CIMT measurements were not available at the relevant time point. Number of subjects in analysis: 335 supplemented subjects and 105 placebo subjects were included in the analysis. Overall completeness of data: Overall, 335 (86%) supplemented subjects and 105 (81%) placebo subjects completed the study. Overall, 85% of subjects completed the study. Overall completion rate is >80% and completion rate for the treatment group is >80%
Incomplete	YES	Exclusions: There were no exclusions because of the

outcome data addressed? Clinical outcome		subjects who were randomized (supplements = 390; placebo = 130), all met the analysis strategy which was ITT. Attritions: 80 subjects did not complete the study but met the analysis strategy. Overall completeness of data: Overall, 335 (86%) supplemented subjects and 105 (81%) placebo subjects completed the study. Overall, 85% of subjects completed the study. Overall completion rate is >80% and completion rate for the treatment group is >80%
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: Funding was received from a supplements company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Sawayama 2002 (FAST)		
Methods	Randomized trial. 2 year duration.	
Participants	Asymptomatic hypercholesterolemia. Probucol N = 82, age (mean, SD) = 65.7 ± 12.9, % female = 69.5, % CVD = 11 Pravastatin N = 83, age (mean, SD) = 65.6 ± 14.8, % female = 73.5, % CVD = 18.1 Diet N = 81, age (mean, SD) = 67.1 ± 13.4, % female = 63.0, % CVD = 13.6	
Interventions	Probucol 500 mg two times each day, after meals or Pravastatin 10 mg once each day, after the evening meal Diet	
Outcomes	Surrogate: common carotid artery, far wall, right and left proximal to bifurcation. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, cerebrovascular accident	
Notes	Funding received from the Japanese Ministry of Education, Science and Culture	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "randomization was done by the minimization method". Comment: probably done
Allocation concealment?	UC	Comment: insufficient information provided.

Blinding? Surrogate outcome	YES	Quote: "All examinations were performed by one trained physician who had no knowledge of the clinical history and risk factor profile of the subjects". Comment: it is likely that the physician was not aware of the treatment assignment. Participants' knowledge of intervention unlikely to influence the outcome or outcome measurement.
Blinding? Clinical outcome	YES	Comment: there is no specific information about blinding with respect to target events, but the outcomes and outcome measurement are not likely to be influenced by lack of blinding.
Incomplete outcome data addressed? Surrogate outcome	YES	Exclusions: Although not explicit, it seems that there were no exclusions from the subjects who were randomized (probucol = 82; pravastatin = 83; control = 81). Attrition: A total of 34 subjects (21%) in the two intervention groups (probucol and pravastatin) did not complete the study. The number of attritions in each group is not stated. It is unclear if the subjects who did not complete the study were included in the analysis- the authors do indicate the analyses were ITT, but no detailed definition of the analysis strategy is provided. Overall completeness of data: the trial was completed by control = 81 (100%) and probucol/pravastatin (combined) = 131 (79%) of subjects. Completion rate is > 80% overall.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: Although not explicit, it seems that of the subjects who were randomized (probucol = 82; pravastatin = 83; control = 81) there were no exclusions . Attrition: A total of 34 subjects (21%) in the two intervention groups (probucol and pravastatin) did not complete the study. Overall completeness of data: the trial was completed by control = 81 (100%) and probucol/pravastatin (combined) = 131 (79%) of subjects. Overall, 212 (86%) subjects completed the study. Completion rate is > 80% overall.
Free of selective reporting?	YES	Comment: monitoring of clinical events listed in methods and reported in results. Expected CIMT measurement as described in the method was reported in results.
Free of other bias?	YES	Quote: "This study was supported by a grant from the Japanese Ministry of Education, Science, and Culture,

		Tokyo, Japan". Comment: the funding source is unlikely to bias the study.

Smilde 2001 (ASAP)		
Methods	Randomized, double-blind, trial. Two-centre. 2 year duration.	
Participants	Familial hypercholesterolemia. Atorvastatin N = 162, age (mean, SD) = 48.5 ± 10.5 (total study subjects), % female = 60 (total study subjects), % CVD = 23.3 (total study subjects) Simvastatin N = 168	
Interventions	Atorvastatin 40 mg/d, after 4 weeks the dose was increased to 80 mg/d or Simvastatin 20mg/d, after 4 weeks the dose was increased to 40 mg/d	
Outcomes	Surrogate: common carotid artery, near and far wall, right and left distal. Clinical: all-cause mortality, cardiovascular mortality	
Notes	Funding received from Parke-Davis B V Netherlands. This study is also referenced by Smilde 2000.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YE S	Quote: "randomisation was done from a computer-generated sequence" Comment: probably done
Allocation concealment?	YES	Quote: "concealed in sequentially numbered, sealed, opaque envelopes, and kept by the hospital pharmacist of the two centers". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "The images stored on disk are read by independent readers blinded to any information on the patient" (Reference: Smilde 2000). Comment: it is likely that blinding of participants and key study personnel was done.
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded
Incomplete outcome data addressed? Surrogate outcome	YES	Attritions and exclusions reported: Exclusions: Of the subjects who were randomized (atorvastatin =162; simvastatin = 168), 5 (atorvastatin = 2; simvastatin = 3) were excluded because they withdrew before treatment and were not considered in the analyses. Attrition: At 12 months, 26 (atorvastatin = 10; simvastatin = 16) subjects had missing data and were not included in the analysis. Number of subjects in efficacy analysis: 150 (93%)

		atorvastatin subjects and 149 (89%) simvastatin subjects. Overall completeness of data: For the 12 month surrogate outcome data, there were atorvastatin = 150 subjects in the trial and simvastatin = 149 subjects in the trial. Completion rate at 12 months is > 80% overall and for the treatment group. both groups.
Incomplete outcome data addressed? Clinical outcome	YES	Attritions and exclusions reported. Exclusions: Of the subjects who were randomized (atorvastatin =162; simvastatin = 168), 5 (atorvastatin = 2; simvastatin = 3) were excluded because they did not meet the analysis strategy: subjects who received treatment. Attrition: 45 (atorvastatin = 19; simvastatin = 26) subjects did not complete the study but met the analysis strategy and were included in the analysis. Number of subjects in analysis: 160 (99%) troglitazone treated subjects and 165 (98%) simvastatin treated subjects. Overall completeness of data: The trial was completed by atorvastatin = 141 (87%) and simvastatin = 139 (83%) subjects. Overall, 280 (85%) subjects completed the trial. Completion rate is > 80% overall and in the treatment group.
Free of selective reporting?	YES	Comment: monitoring of clinical events listed in methods and reported in results. Expected CIMT measurement as described in the method was reported in results.
Free of other bias?	NO	Quote: "this study was supported by Parke Davis". Comment: research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Terpstra 2004 (ELVERA)	
Methods	Randomized, double-blind trial. 2 year duration.
Participants	Mild to moderate hypertension without previous treatment. Amlodipine N = 81, age (mean, SD) = 67.0 ± 4.0, % female = 53.1, duration of hypertension (mean, SD) = newly diagnosed, % CVD = 0 Lisinopril N = 85, age (mean, SD) = 67 ± 4, % female = 36.5, duration of hypertension (mean, SD) = newly diagnosed
Interventions	Amlodipine 5mg/d, after 6 weeks the dose was increased to 10 mg/d Lisinopril 10mg/d, after 6 weeks the dose was increased to 20 mg/d
Outcomes	Surrogate: common carotid, far wall Clinical: all-cause mortality
Notes	Funding received from Pfizer BV
Risk of bias	

Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: insufficient information provided
Allocation concealment?	UC	Comment: insufficient information provided.
Blinding? Surrogate outcome	YES	Quote: "double-blind". Quote: "all images were saved on S-VHS tape and analyzed off-line throughout the study by an analyst who was unaware of the patients' characteristics". Comment: It is likely that the key study personnel and participants were blinded .
Blinding? Clinical outcome	YES	Quote: "double-blind" Comment: it is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions and attritions reported. At 12 months, of the subjects who were randomized (amlodipine = 81; lisinopril = 85) 37 (amlodipine = 17; lisinopril = 20) were excluded because they did not meet the analysis strategy: valid CIMT "reading at baseline and at least one valid observation after 1 and 2 years." Attrition: For the 12 month surrogate outcome data, there were no subjects who did not complete the study. Number of subjects in the efficacy analysis: At 12 months, 64 (79%) amlodipine and 65 (76%) lisinopril subjects were included in the efficacy analysis. Overall completeness of data: for the 12 month data, there were amlodipine = 64 (79%) and lisinopril = 65 (76%) subjects in the trial. Overall completion rate at 12 months is 77%. Overall completion rate at 12 months is <80% and <80% in the treatment group
Incomplete outcome data addressed? Clinical outcome	NO	Exclusions and attritions reported. None, because all randomized subjects (amlodipine = 81; lisinopril = 85) were included in the analysis strategy which was ITT. Attrition: Overall, 46 subjects (amlodipine = 24; lisinopril = 22) did not complete the study. Overall completeness of data: the trial was completed by amlodipine = 57 (70%) and lisinopril = 63 (74%) subjects. Overall completion rate is < 80% overall and <80% in the treatment group.
Free of selective reporting?	YES	Comment: Prespecified CIMT measurement as described in the methods section was reported in the results. Clinical events not mentioned in methods, but it is likely

		that all events that occurred were reported
Free of other bias?	NO	Comment: The study was sponsored by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

van Vonderen 2009		
Methods	Randomized trial. Subjects from a single centre or referred from other centres participating in a multicentre trial. 2 year duration.	
Participants	antiretroviral naïve male patients with indicate to initiate cART NVP/LPV/r N = 18, age (median, range) = 44.8 (39.7 – 54.7), % female = 0, % CVD = 11.1 ZDV/3TC/LPV/r N = 19, age (mean, SD) = 38.0 (35.0 – 43.1), % female = 0, % CVD = 15.8	
Interventions	NVP 200 mg twice per day plus LPV/r 533/133 mg twice per day (NRTI sparing regimen) or ZDV/3TC 300/150 mg twice per day plus LPV/r 400/100 mg twice per day (NRTI containing regimen)	
Outcomes	Surrogate: common carotid arteries, right and left Clinical: all-cause mortality, myocardial infarction	
Notes	Funding received from Abbott International and Boehringer Ingelheim	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "a treatment allocation sequence was generated using the minimization variable BMI". Comment: probably done
Allocation concealment?	YES	Quote: "at the central study coordinating centre, a treatment allocation sequence was generated". Comment: probably done
Blinding? Surrogate outcome	YES	Quote: "all CIMT and arterial stiffness measurements were performed by a single investigator in a blinded fashion". Comment: blinding of participants is unclear, but it is likely that lack of participant blinding would not have influenced the outcome or outcome measurements.
Blinding? Clinical outcome	YES	Comment: information on blinding with respect to clinical outcomes is not provided. It is likely that lack of blinding would not have influenced the outcomes or outcome measurements.
Incomplete	YES	Exclusions: None of the subjects who were randomized

outcome data addressed? Surrogate outcome		(nvp/lpv/r = 18; zdv/3TC/LPV/r = 19), were excluded because the analysis included all randomized subjects and was ITT. Number of attritions: 1 subject (nvp/lpv/r = 1) did not complete the study but authors do not state if this subject had surrogate outcome data available. Overall completeness of data: 17 (94%) nvp/lpv/r subjects and 19 (100%) zdv/3tc/lpv/r sujects completed the study. Overall completion rate was 97%. Completion rate in nvp/lpv/r group > 80% and completion rate in zdv/3TC/LPV/r group was >80.
Incomplete outcome data addressed? Clinical outcome	YES	Exclusions: None of the subjects who were randomized (nvp/lpv/r = 18; zdv/3TC/LPV/r = 19), were excluded because the analysis included all randomized subjects and was ITT. Number of attritions: 1 subject (nvp/lpv/r = 1) did not complete the study but had outcome data available and was included in the analysis. Overall completeness of data: 18 (100%) nvp/lpv/r subjects and 19 (100%) zdv/3tc/lpv/r subjects provided complete outcome data. Overall completeness of data was 100%. Completeness of data was > 80% overall and >80% for the treatment group.
Free of selective reporting?	YES	Comment: CIMT measurements as described in the method were reported. Clinical events monitoring not listed in methods, but it is likely that all events that occurred were reported by the authors
Free of other bias?	NO	Comment: The trial was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Wiklund 2002 (ELVA)	
Methods	Randomized, double-blind, placebo controlled trial. Single centre. 3 year duration.
Participants	Hypercholesterolemia. Metoprolol N = 62, age (mean, SD) = 59.5 ± 9.9, % female = 50.0, % CVD = 0 Placebo N = 67, age (mean, SD) = 60.0 ± 9.3, % female = 48.0, % CVD = 8
Interventions	Metoprolol 100 mg/d or Placebo
Outcomes	Surrogate: common carotid artery, far wall, right. Clinical: myocardial infarction
Notes	Funding received from the Swedish Heart-Lung Foundation, the

	Swedish Medical Research Council and AstraZeneca,Mölndal, Sweden.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	UC	Comment: insufficient information provided.
Allocation concealment?	UC	Comment: insufficient information provided
Blinding? Surrogate outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Blinding? Clinical outcome	YES	Quote: "double-blind". Comment: it is likely that key study personnel and participants were blinded.
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the randomized subjects (metoprolol = 62; placebo = 67) 37 (metoprolol = 22; placebo = 15) were excluded because they did not meet the analysis criteria which was subjects with at least one follow-up ultrasound measurement. Attrition: this information is not provided. 50 (metoprolol = 27; placebo = 23) did not complete the study but this is not specific to the number of subjects who were included in the CIMT analysis. Number of subjects in analysis: 40 (65%) metoprolol subjects and 52 (78%) placebo subjects were included in the analysis. Overall completeness of data: 35 (56%) metoprolol subjects and 44 (66%) placebo subjects completed the trial. Overall completion rate <80%.
Incomplete outcome data addressed? Clinical outcome	NO	Exclusions: Of the randomized subjects (metoprolol = 62; placebo = 67) 26 (metoprolol = 15; placebo = 11) were excluded because they were withdrawn at placebo run-in. Attrition: 24 (metoprolol = 12; placebo = 12) did not complete the study but were included in the analysis. Number of subjects in analysis: 47 (76%) metoprolol subjects and 52 (84%) placebo subjects were included in the analysis. Overall completeness of data: 35 (56%) metoprolol subjects and 44 (66%) placebo subjects completed the trial.
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported in the results. Clinical events monitoring was not listed in methods, but it is likely that all events that occurred were reported.

Free of other bias?	NO	Comment: The study was partially funded by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Zanchetti 1998 (VHAS)		
Methods	Randomized, double-blind (for first six months), placebo controlled trial. Multicentre. 4 year duration.	
Participants	Hypertension. Verapamil N = 224, age (mean, SD) = 54.2 ± 6.8, % female = 46.9, duration of hypertension in years (mean, SD) = 5.2 ± 5.5, % CVD = 5.1 Chlorthalidone N = 232, age (mean, SD) = 53.9 ± 7.2, % female = 48.7, duration of hypertension in years (mean, SD) = 5.1 ± 5.7, % CVD = 4.8	
Interventions	Verapamil 240 mg/d or Chlorthalidone 25 mg/d	
Outcomes	Surrogate: common carotid, carotid bifurcation, internal carotid, far wall, right and left. Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from Knoll Farmaceutici Spa and Ravizza Farmaceutici Spa.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "patients were randomly assigned". Comment: Probably done because another report from the authors has provided a description of an adequately generated allocation sequence.
Allocation concealment?	UC	Comment: insufficient information provided.
Blinding? Surrogate outcome	YES	Quote: "double blind for the first six months, open subsequently. Quote: "all measurements were recorded by two different readers, unaware of the subject's identity, clinical condition and assigned treatment" Comment: outcome assessors were blinded and open label unlikely to influence outcome/outcome measurement.
Blinding? Clinical outcome	YES	Quote: "double-blind for the first six months, open subsequently". Quote: "the details of the cardiovascular events were verified, according to predetermined criteria, by experts blind to the randomized treatment assigned".

		(Reference: Rosei 1997) Comment: open-label unlikely to influence outcome or outcome measurement
Incomplete outcome data addressed? Surrogate outcome	NO	Exclusions: Of the subjects randomized (verapamil=244; chlorthalidone=254), 42 (verapamil = 20; chlorthalidone = 22) were excluded because they did not have satisfactory baseline ultrasound readings. A further 79 (verapamil = 38; chlorthalidone = 41) were excluded because they did not meet the analysis strategy: subjects with "ultrasound measurements taken on at least 3 different occasions over a period of at least 2 years". Attritions: the authors do not provide information on the number of subjects who met the analysis criteria but for whom CIMT measurements were not available at the relevant time point. Number of subjects in analysis: 186 (76%) verapamil subjects and 191 (75%) chlorthalidone subjects were included in the analysis. Overall completeness of data: Completion rates for the CIMT measurement or for the overall trial are not provided. Overall 377 (76%) of the randomized subjects were included in the analysis. Overall completion rate is at most <80% and overall completion rate in each group is at most <80% as well.
Incomplete outcome data addressed? Clinical outcome	UC	Exclusions: Of the subjects randomized (verapamil=244; chlorthalidone=254), 42 were excluded because they did not meet the analysis strategy: subjects with ultrasound measurements for all carotid walls considered and not just the carotid near walls. Attritions: the authors do not provide information on the number of subjects who met the analysis criteria but for whom follow-up measurements were not available. Number of subjects in analysis: 224 verapamil subjects and 232 chlorthalidone subjects were included in the analysis. Overall completeness of data: Completion rates for the trial are not provided.
Free of selective reporting?	YES	Comment: CIMT measurements as specified in the methods were reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: Funding was received from a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)

Zanchetti 2002 (ELSA)		
Methods	Randomized, double-blind trial. Multicentre. 4 year duration.	
Participants	Hypertension Lacidipine N = 1177, age (mean, SD) = 56.1 ± 7.5, % female = 45.8, duration of hypertension in years (mean, SD) = 7.0 ± 6.3 (total study subjects), % CVD = 12.8 (total study subjects) Atenolol N = 1157 age (mean, SD) = 55.9 ± 7.5, % female = 44.6,	
Interventions	Lacidipine 4 mg/d. After 1 month, if a reduction of ≥ 5 mm Hg in DBP and a DBP of < 95 mm Hg was not achieved, the dose was increased to 6 mg/d. After 3 months, HCTZ 12.5 mg/d could be added. After 6 months, HCTZ 25 mg/d could be added or Atenolol 50 mg/d. After 1 month, if a reduction of ≥ 5 mm Hg in DBP and a DBP of < 95 mm Hg was not achieved, the dose was increased to 100 mg/d. After 3 months, HCTZ 12.5 mg/d could be added. After 6 months, HCTZ 25 mg/d could be added or	
Outcomes	Surrogate: common carotid artery, near and far wall, right and left Clinical: all-cause mortality, cardiovascular mortality, myocardial infarction, stroke	
Notes	Funding received from GlaxoSmithKline Italy, Verona and Boehringer Ingelheim International. This study was also referenced by Zanchetti 1996, Zanchetti 1998, and Zanchetti 2004.	
Risk of bias		
Item	Authors' judgement	Description
Adequate sequence generation?	YES	Quote: "randomization was computer generated". Comment: probably done
Allocation concealment?	UC	Comment: insufficient information provided
Blinding? Surrogate outcome	YES	Quote: "Patients and study personnel, excluding the Safety Committee, were blinded to treatment assignment for the study duration". Quote: "Scans of any individual patient were assigned to the same reader, but the scan time-sequence was randomized so that the reader was blind to the time of recording". Comment: probably done.
Blinding? Clinical outcome	YES	Quote: "Patients and study personnel, excluding the Safety Committee, were blinded to treatment assignment for the study duration." Quote: "Incidence of fatal and nonfatal cardiovascular events…adjudicated by an Event Monitoring Committee blinded to treatment assignment". Comment: probably done
Incomplete	NO	Exclusions and attritions reported. Exclusions: Of the

outcome data addressed? Surrogate outcome		subjects who were randomized (lacidipine = 1177; atenolol = 1157), 299 (lacidipine = 154; atenolol = 145) were excluded because they did not meet the analysis strategy: all randomized subjects who "had baseline ultrasound scan and at least 1 follow-up scan, including scans performed after withdrawal". Attrition: 516 (lacidipine = 268; atenolol = 248) subjects did not complete the trial but were included in the analysis strategy. Number of subjects analysed: 1023 lacidipine subjects and 1012 atenolol subjects were included in the analysis. Overall completeness of data: 755 (64%) lacidipine subjects completed the trial and 764 (66%) atenolol subjects completed the trial. Overall, 1519 (65%)subjects completed the trial. Overall completion rate < 80% and completion rate in the treatment group was < 80%
Incomplete outcome data addressed? Clinical outcome	NO	Exclusions and attritions reported. Exclusions: None, because all subjects who were randomized (lacidipine = 1177; atenolol = 1157), were included in the analysis. Attrition: 815 (lacidipine = 422; atenolol = 393) subjects did not complete the trial but were included in the analysis strategy. Overall completeness of data: 755 (64%) lacidipine subjects completed the trial and 764 (66%) atenolol subjects completed the trial. Overall, 1519 (65%) subjects completed the trial. Overall completion rate < 80% and completion rate in the treatment group < 80%
Free of selective reporting?	YES	Comment: CIMT as specified in the methods was reported in the results. Clinical events as specified in the methods were reported in the results.
Free of other bias?	NO	Comment: The study was sponsored by a drug company. Research has shown that industry-funded trials are more likely to be associated with an overestimation of the treatment effect (Bhandari 2004)