Multiple testing correction in a meta-analysis of all adverse events recorded in large long-term randomised trials of statin therapy



Enti Spata

St Catherine's College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Long Vacation Term 2022

Abstract

Background: Randomised trials have shown that statin therapy reduces the risk of major vascular events (ie, heart attacks, strokes and coronary revascularisation procedures) without any increase in the risk of nonvascular causes of death or of site-specific cancer, but does produce a small increase in the risk of muscle pain or weakness, diabetes and, possibly, haemorrhagic stroke. Although statins are widely prescribed, there are concerns that they might have a range of other side effects. This DPhil addresses these concerns through an individual-participant-data meta-analysis of all recorded adverse events (AEs) in all large, long-term, randomised, double-blind trials of statin therapy, taking into consideration the substantial challenges related to multiple hypotheses testing (MHT).

Methods: Double-blind randomised trials of statin therapy with at least 1,000 participants and a scheduled treatment duration of at least 2 years were included. Individual participant data on all AEs reported in 19 trials of statin vs placebo (123,940 randomised participants) and 4 trials of a more intensive versus a less intensive statin regimen (30,724 randomised participants) were analysed. A literature review identified potentially relevant MHT methods and simulation studies were done to assess their expected performance (ie, control of false positive [FP] and false negative results). Adverse event data were organised and coded according to a common medical dictionary based upon the Medical Dictionary for Regulatory Activities (MedDRA). Under the selected strategy, inverse-variance-weighted meta-analyses of the effects on all AEs were performed using time-to-event analyses for the first occurrence of each outcome among participants randomly assigned into each trial.

Results: The literature review identified 6,569 eligible papers, of which 337 were included for full text review. Five MHT methods (Holm, Hochberg, Hommel, Benjamini & Hochberg and Mehrotra & Adewale) were selected as the most appropriate based on their statistical control of type I and II error rates. Simulation analyses identified the Mehrotra & Adewale method, which controls the false discovery rate (FDR), as the most suitable as it resulted in a low

expected number of FP results whilst maintaining reasonable statistical power to detect any real effects of statin therapy. Blinded (ie, using a 'shuffled' treatment allocation) meta-analyses of the trials of statin vs placebo were first done, which confirmed that no correction for MHT resulted in 179 (4.4%) FP findings while the Mehrotra & Adewale MHT method led to zero FPs. Unblinded analyses (ie, using the actual treatment allocation) of all non-lipid-related AEs then confirmed the already known beneficial effects of statins on major vascular events. They also showed that statin therapy was FDR-significantly associated with a reduced risk of having an arteriogram or an angiogram procedure. Subsequent analyses which ignored the already known beneficial or harmful effects of statin therapy, very rare outcomes and irrelevant overlap in tests, identified two new benefits of statin therapy: reductions in the risk of peripheral embolism and thrombosis (361 [0.6%] vs 451 [0.7%], rate ratio [RR] 0.72, 95% confidence interval [CI]: 0.61−0.83, p=0.0012), and carotid endarterectomy (47 [<0.1%] vs 83 [0.1%], RR 0.57, 95% CI: 0.40−0.80, p= 0.0013). However, statins were also FDR-significantly associated with an increased risk of hepatobiliary investigations (1,349 [2.2%] vs 1,025 [1.7%], RR 1.32, 95% CI: 1.22–1.43, p<0.0001) and the extremely rare outcome parosmia (18 vs 2, p=0.00036). The meta-analyses of the 4 trials of more vs less statin revealed that more intensive statin therapy was associated with an FDR-significantly reduced risk of coronary artery disorders (3,518 [22.9%] vs 3,779 [24.6%], RR 0.92, 95% CI: 0.88−0.96, p= 0.00025), general system disorders (5,098 [33.1%] vs 5,682 [37.1%],RR 0.87, 95% CI: 0.84−0.91, p<0.0001), any surgical and medical procedures (4,067 [26.4%] vs 4,333 [28.3%],RR 0.93, 95% CI: 0.89−0.97, p= 0.00059) and the rare AE of renal therapeutic procedures (30 [0.2%] vs 57 [0.4%], RR 0.54, 95% CI: 0.35−0.82, p= 0.0038). More intensive statin regimens were also FDR-significantly associated with an increased risk of hepatobiliary investigations (502 [3.3%] vs 253 [1.6%], RR 1.95, 95% CI: 1.69–2.25, p<0.0001) and high blood creatine phosphokinase (223 [1.4%] vs 146 [1.0%], RR 1.51, 95% CI: 1.24−1.86, p<0.0001).

Conclusion: These findings highlight the importance of using appropriate statistical methods to control for multiple testing. The results confirmed the already known benefits of statin

therapy while identifying some new potential benefits and harms. However, the nature and frequency of the newly detected harms confirm that the benefits of statin therapy greatly outweigh the potential harms.

Acknowledgements

The main acknowledgement is to all the participants who took part in the included trials, as well as the investigators who conducted those trials, without whom this research would not be possible.

I would like to thank my supervisors, Professor Jonathan Emberson, Dr Christina Reith and Professor Colin Baigent, for their precious time, patient guidance and valuable support and encouragement during the past five years.

I have also been supported by the wider Vascular Overviews Group, who are Secretariat Members of the Cholesterol Treatment Trialists' collaboration. In particular, I would like to thank Lisa Blackwell, Kelly Davies, Heather Halls, Lisa Holland and Kate Wilson.

I would like to thank the British Heart Foundation, the Medical Research Council and the Nuffield Department of Population Health, University of Oxford, for providing funding to enable me to undertake this DPhil.

Lastly, I would like to thank my wife for her patience and support, to whom this thesis is dedicated.

Statement of contribution to work

I was responsible for conceiving and developing the research questions addressed in this thesis.

The standardization of the individual patient data (IPD) received from the multiple trials into a common format ready for meta-analyses and the harmonization of the adverse event coding systems used by the individual trials into a single coding system was achieved in collaboration with other members of the Cholesterol Treatment Trialists' Secretariat in Oxford. In particular, I was co-responsible (together with Lisa Blackwell) for developing the statistical methods used for the standardization of the IPD and the programs created to harmonize the AEs. Also, I developed the SAS programs used to standardize the IPD from 17 (out of 28) trials, which were then validated by at least one research associate.

I performed all aspects of the literature review of the methods for multiple hypotheses testing, as well as the design and implementation of the simulations analyses. I designed the various analyses and conducted all the statistical analyses shown. Finally, I produced all the tables and figures, interpreted the findings and wrote-up all aspects of the study presented in this thesis.

Papers published

To date two papers have arisen as part of this work:

1. Harmonisation of large-scale, heterogeneous individual participant adverse event data from randomised trials of statin therapy. *Clinical Trials 2022; 10.1177/174-07745221105509*. On behalf of the Cholesterol Treatment Trialists' Collaboration: Enti Spata*, Lisa Blackwell*, Kelly Davies, Heather Halls, Lisa Holland, Kate Wilson, Charlie Harper, Alistair Roddick, Nathan Samuel, Will Stevens, Karl Wallendszus, David Preiss, Anthony Keech, John Simes, Rory Collins, Colin Baigent, Jonathan Emberson†, Christina Reith† (*equal first author and †equal last author).

2. Effect of statin therapy on muscle symptoms: an individual participant data meta-analysis of large-scale, randomised, double-blind trials. *Lancet 2022; 400: 832–45.* On behalf of the Cholesterol Treatment Trialists' Collaboration: Christina Reith*, Colin Baigent*, Lisa Blackwell, Jonathan Emberson, Enti Spata, Kelly Davies, Heather Halls, Lisa Holland, Kate Wilson, Jane Armitage, Charlie Harper, David Preiss, Alistair Roddick, Anthony Keech, John Simes, Rory Collins (*These authors contributed equally).

List of abbreviations

AE	Adverse Event
CE	Clinical Events
CDISC	Clinical Data Interchange Standards Consortium
CI	Confidence Interval
CHD	Coronary Heart Disease
CM	Co-Medication
CRF	Case Report Form
CSP	Case Study Protocol
CSR	Case Study Report
CV	Cardiovascular
CVD	Cardiovascular Disease
CTT	Cholesterol Treatment Trialists'
DD	Death Details
DM	Demographics
ERFC	Emerging Risk Factors Collaboration
EX	Exposure
FDA	Food and Drug Administration
FDR	False Discovery Rate
FDP	False Discovery Proportion
FDX	False Discovery Exceedance
FP	False Positive
FWER	Familywise Error Rate
HLGT	Higher Level Group Term
HLT	Higher Level Term
HO	Healthcare Encounters
ICD	International Classification of Diseases
IQR	Interquartile Range
IPD	Individual Participant Data
LB	Laboratory
LDL	Low Density Lipoprotein
LDL-C	Low Density Lipoprotein Cholesterol
LLT	Lowest Level Term
MedDRA	Medical Dictionary for Regulatory Activities
MH	Medical History
MHT	Multiple Hypothesis Testing
MVE	Major Vascular Events
NEC	Not Elsewhere Specified
PR	Procedures
PT	Preferred Term
RCT	Randomised Controlled Trial
RR	Relative Risk
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SMQ	Standardised MedDRA Queries
SOC	System Organ Class
SU	Substance Use

SV	Subject Visits
UK	United Kingdom
VS	Vital Signs

Table of contents

1 Introduction 20
1.1 The need for large-scale randomised controlled trials and meta-analyses 20
1.2 LDL cholesterol and cardiovascular disease 21
1.2.1 Epidemiological evidence 21
1.2.2 HMG-CoA reductase inhibitors ('statins') 23
1.2.3 Use of statins worldwide 23
1.3 The Cholesterol Treatment Trialists' (CTT) Collaboration 24
1.3.1 CTT efficacy findings: Effects of statin therapy on major vascular events 24
1.3.2 CTT safety findings: Effects of statin therapy on cancer and non-vascular mortality 26
1.3.3 Effects of statin therapy on other outcomes 27
1.3.3.1 AEs probably or possibly associated with statin therapy 29
1.3.3.2 AEs with unknown association with statin therapy 29
1.4 Interpreting statistical evidence 30
1.4.1 The null hypothesis 30
1.4.2 Type of null hypothesis 30
1.4.3 Limitations of the null hypothesis 31
1.4.3.1 False positive (the 'type I' error, or 'alpha') 31
1.4.3.2 False negative (the 'type II' error, or 'beta') 31
1.4.4 Multiple hypothesis testing (MHT) 32
1.5 Concepts of multiple hypotheses testing 32
1.5.1 Family of hypotheses 32
1.5.2 Controlling for statistical errors 33
1.5.2.1 Familywise Error Rate 33
1.5.2.2 False Discovery Rate 33
1.5.3 Correlation of hypotheses in MHT 34
1.5.3.1 Assumption free 34
1.5.3.2 Independent 34
1.5.3.3 Correlated 35
1.6 Aims of this thesis 35
2 Literature review of MHT methods 42
2.1 Summary 42
2.2 Aim of this chapter 42

2.3 Characteristics of MHT methods 43
2.3.1 Type of null hypothesis 44
2.3.2 Control of statistical error rate 44
2.3.2.1 FamilyWise Error Rate 44
2.3.2.2 False Discovery Rate 45
2.3.2.3 Other Error Rates 46
2.3.3 Correlation of hypotheses 47
2.3.3.1 Assumption free 48
2.3.3.2 Independent 48
2.3.3.3 Correlated 49
2.3.4 Ordering of multiple hypotheses 50
2.3.4.1 No ordering required 51
2.3.4.2 Data-driven ordering 51
2.3.4.3 Pre-specified ordering 51
2.3.4.4 Unequal weighting of hypotheses 52
2.3.4.5 MHT methods with pre-specified ordering *and* unequal weights 52
2.3.5 Adaptive approaches 53
2.3.6 Bayesian approaches to MHT 54
2.3.6.1 Analysis of AE data in clinical trials 54
2.3.7 Summary of characteristics for MHT methods suitable to use in CTT 56
2.4 Systematic search of MHT methods 56
2.4.1 Search strategy 56
2.4.2 Search results 57
2.5 Review of all suitable MHT Methods 57
2.5.1 Methods controlling the Familywise Error Rate 58
2.5.1.1 Bonferroni's method 58
2.5.1.2 Holm's 'step-down' method 58
2.5.1.3 Hochberg's 'step-up' method 59
2.5.1.4 Rom's 'step-up' method 60
2.5.1.5 Hommel's 'two-step' method 60
2.5.1.6 FWER MHT methods in practice 61
2.5.2 Methods controlling the False Discovery Rate 61
2.5.2.1 Benjamini & Hochberg 'step-down' method 61
2.5.2.2 Benjamini & Liu's 'step-down' method 62

2.5.2.3 Mehrotra & Heyse method 63
2.5.2.4 Mehrotra & Adewale's method 63
2.5.2.5 FDR MHT methods in practice 64
2.6 Summary of findings 64
3 CTT data processing and statistical methods 70
3.1 Summary 70
3.2 Data harmonisation methodology 70
3.2.1 Step 1: Review of trial IPD 71
3.2.2 Step 2: Creation of 'Meta' data dictionary—use of global standards 71
3.2.3 Step 3: Standardization of AEs to a common medical dictionary 73
3.2.3.1 Medical Dictionary for Regulatory Activities (MedDRA) 73
3.2.3.2 CTT: AE MedDRA mapping approach 74
3.2.4 Specification of outcomes for hypothesis-testing 75
3.3 Correlation structure 76
3.3.1 Correlated effect estimates by construction 76
3.3.2 Correlated effect estimates due to similar underlying pathology 76
3.4 Populations of interest 77
3.5 Simulation analyses 78
3.5.1 Selection of the family of hypotheses 78
3.5.2 Simulated scenarios 80
3.5.3 Algorithm development 81
3.5.4 MHT methods applied 83
3.6 Statistical methods for meta-analysis 85
3.6.1 Time to event 85
3.6.2 Meta-analysis 85
3.6.3 Multiple testing 86
3.7 Synopsis 86
4 Simulation analyses 95
4.1 Summary 95
4.2 Introduction 95
4.3 Methods 96
4.4 Simulation results for trials of statin vs placebo 97
4.4.1 Data 97
4.4.1.1 CTT AE database 97

4.4.1.2 Distribution of events across the MedDRA tree 98
4.4.2 No correction for multiple testing 98
4.4.2.1 Distribution of false positives results when the null hypothesis true 98
4.4.2.2 Statistical power of detecting true effects 100
4.4.3 Comparison of MHT methods 101
4.4.3.1 Distribution of false positives results 101
4.4.3.2 Statistical power of detecting true effects 102
4.4.3.3 Effect of selection of family on statistical power 104
4.4.3.4 Summary for comparison of MHT methods 106
4.4.4 Multiple testing framework in CTT 107
4.4.4.1 Recommendations on testing strategy 107
4.4.4.2 Statistical properties of Mehrotra & Adewale (FDR) method 108
5 Meta-analyses of the double-blind trials of statin vs placebo 128
5.1 Summary 128
5.2 Introduction 128
5.3 Methods 130
5.3.1 Data 130
5.3.2 Analysis 130
5.3.2.1 Statistical analysis 130
5.3.2.2 Multiple testing 131
5.3.2.3 Analyses of AEs 131
5.4 Results 133
5.4.1 Data 133
5.4.2 Analyses of AEs blinded to treatment allocation 134
5.4.3 Unblinded analyses of all non-lipid related AEs 134
5.4.4 Unblinded analyses of all AEs with prior unknown association with statins 136
5.4.5 Unblinded analyses of all AEs with prior unknown association with statins after excluding very rare outcomes and avoiding irrelevant overlap in tests 137
5.4.6 Further exploratory analyses 137
5.5 Discussion 138
6 Meta-analyses of the double-blind trials of more vs less intensive statin therapy 152
6.1 Summary 152
6.2 Introduction 152
6.3 Methods 153

6.3.1 Data 153
6.3.2 Analysis 153
6.3.2.1 Statistical analysis 154
6.3.2.2 Multiple testing 154
6.3.2.3 Analyses of AEs 154
6.4 Results 155
6.4.1 Data 155
6.4.2 Analyses of AEs blinded to treatment allocation 156
6.4.3 Unblinded analyses of all non-lipid related AEs 156
6.4.4 Unblinded analyses of all AEs with prior unknown association with statins 157
6.4.5 Unblinded analyses of all AEs with prior unknown association with statins after excluding very rare outcomes and avoiding irrelevant overlap in tests 158
6.4.6 Further exploratory analyses 158
6.5 Discussion 159
7. Discussion 174
7.1. Summary 174
7.2. Overview of findings 174
7.3. Strengths 179
7.3.1. Large-scale, double-blinded randomised trials 179
7.3.2. IPD meta-analyses of all AEs reported in large-scale trials 180
7.3.3. Validation of methodologies applied 180
7.3.4. Rapid assessment for hypotheses related to statin therapy 181
7.4. Limitations 182
7.5. Future work 183
7.6. Novelty 184
7.6.1. Ability to address public health concerns 184
7.6.2. Methodological novelties 186
7.7. Conclusion 186

List of boxes

Box 2.1: Potential inferences when simultaneously testing m hypotheses 66

Box 3.1: CDISC SDTM (V3.2) domains created for the completion of the CTT AE project including appropriate modifications 87

Box 3.2: The five-tiered hierarchical structure of Medical Dictionary for Regulatory Activities (MedDRA) 88

Box 3.3: Hypothetical example of a meta-analyses of 10 AEs reported through the MedDRA hierarchical levels and their associated p-values 89

List of tables

Table 1.1: Characteristics of the various type of statins used in clinical practice 37

Table 1.2: Characteristics of the 28 major statin trials 38

Table 2.1: Search strategy for MHT methods 68

Table 2.2: Characteristics of multiple testing methods suitable to use in CTT 69

Table 3.1: Characteristics of 23 double-blinded and 5 open-label statin trials and data provided 90

Table 3.2: Hierarchical levels of MedDRA version 20, organized alphabetically by System Organ Class 92

Table 3.3: Number of total and unique adverse event terms mapped across the hierarchical levels of MedDRA (V20.0) 93

Table 4.1: Trials of statin vs placebo: Distribution of the total number of events across the three analysable MedDRA hierarchical levels, organized alphabetically by System Organ Class 111

Table 4.2 (a): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 6 events are tested across the MedDRA tree resulting in a total of 2,963 tests 112

Table 4.2 (b): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 50 events were tested across the MedDRA tree resulting in a total of 1,325 tests 113

Table 4.2 (c): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 100 events were tested across the MedDRA tree resulting in a total of 977 tests 114

Table 4.2 (d): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 1000 events were tested across the MedDRA tree resulting in a total of 238 tests 115

Table 4.3 (a): Statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 6 events are tested across the MedDRA tree resulting in a total of 2,963 tests 116

Table 4.3 (b): Simulated statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 50 events were tested across the MedDRA tree resulting in a total of 1,325 tests 117

Table 4.3 (c): Simulated statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 100 events were tested across the MedDRA tree resulting in a total of 977 tests 118

Table 4.3 (d): Simulated statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 1000 events were tested across the MedDRA tree resulting in a total of 238 tests 119

Table 4.4: Distribution of FPs for various simulated scenarios when applying various MHT methods and the following multiple testing strategy resulting in 1,027 tests: (i) test any Preferred Term with at least 50 events and (ii) test any term higher up the MedDRA hierarchical tree which reports 50% extra information† 120

Table 5.1: Characteristics of the participating trials 142

Table 5.2: Distribution of the total number of events across the three analysable MedDRA hierarchical levels, organized alphabetically by System Organ Class (19 trials, 123,940 randomised participants) 143

Table 6.1: Characteristics of the participating trials 163

Table 6.2: Distribution of the total number of events across the three analysable MedDRA hierarchical levels, organized alphabetically by System Organ Class (4 trials, 30,724 randomised patients) 164

List of figures

Figure 1.1: Effects on major vascular events per 1.0 mmol/L reduction in LDL cholesterol in randomised trials of statin therapy among people with different presenting characteristics 40

Figure 1.2: Proportional reductions in risks of major vascular events per mmol/L reduction in LDL cholesterol during each year of scheduled statin treatment, in randomised trials of routine statin therapy versus no routine statin use 41

Figure 2.1: Review process of all eligible articles for multiple testing methods 69

Figure 3.1: Data analysis flow diagram 94

Figure 4.1: Trials of statin vs placebo: Distribution of the total number of events across the MedDRA hierarchal levels 121

Figure 4.2: Trials of statin vs placebo: under the null hypothesis, simulation of the distribution of false positives when no correction for multiple testing is applied (ie, p-value<0.05); 122

Figure 4.3: Expected power to reject the null hypothesis (at the 5% level) when statin therapy is assumed to increase the risk of adverse events by between 5% and 900% (ie, RR of between 1.05 and 10), and if no correction for multiple testing is applied 123

Figure 4.4: Relationship between the relative risk and the smallest number of events needed for such relative risk to be declared 'significant' at the 5% or 1% level 124

Figure 4.5: Trials of statin vs placebo: breakdown of expected power by number of events in each test, in the extreme situation whereby all adverse events are increased between 5% to 900% (ie, relative risk of 1.05 to 10.00); two statistical methods are applied for handling multiple testing: no correction and FDR correction 125

Figure 4.6: Trials of statin vs placebo: Mathematical relationship between the total number of events and the statistical power for detecting an increase in risk of 10%, 25%, 50% and 100% (ie, RR of 1.10, 1.25, 1.50 and 2.0) 126

Figure 4.7: Trials of statin vs placebo: breakdown of expected power by number of events in each test when applying various MHT methods and the following multiple testing strategy (resulting in 1,027 tests): (i) test any Preferred Term with at least 50 events and (ii) test any term higher up the MedDRA hierarchical tree which reported 50% extra information†; simulated scenario: null hypothesis (ie, no effect) assumed randomly in 95% of tests and alternative hypothesis (ie, true effect exist) assumed in 5% of tests 127

Figure 5.1: Distribution of total number of events across the MedDRA hierarchal levels (19 trials, 123,940 randomised patients) 144

Figure 5.2: All adverse events; BLINDED to treatment allocation (19 trials, 123,940 randomised patients) 145

Figure 5.3: UNBLINDED analyses of all non lipid−related adverse events (19 trials, 123,940 randomised patients) 146

Figure 5.4: FDR significant effects of statin therapy on non- lipid-related adverse events (19 trials, 123,940 randomised patients) 147

Figure 5.5: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (19 trials, 123,940 randomised patients) 148

Figure 5.6: FDR significant effects of statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (19 trials, 123,940 randomised patients) 149

Figure 5.7: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (19 trials, 123,940 randomised patients) 150

Figure 5.8: FDR significant effects of statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (19 trials, 123,940 randomised patients) 151

Figure 6.1: Distribution of total number of events across the MedDRA hierarchal levels (4 trials, 30,724 randomised patients) 165

Figure 6.2: All adverse events; BLINDED to treatment allocation (4 trials, 30,724 randomised patients) 166

Figure 6.3: UNBLINDED analyses of all non lipid−related adverse events (4 trials, 30,724 randomised patients) 167

Figure 6.4: FDR significant effects of more statin therapy on non- lipid-related adverse events (4 trials, 30,724 randomised patients) 168

Figure 6.5: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (4 trials, 30,724 randomised patients) 169

Figure 6.6: FDR significant effects of more statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (4 trials, 30,724 randomised patients) 170

Figure 6.7: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (4 trials, 30,724 randomised patients) 171

Figure 6.8: FDR significant effects of more statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (4 trials, 30,724 randomised patients) 172

Figure 6.9: Meta-analyses of the adverse events identified to be increased with statin or more statin therapy at the 5% FDR level 173

1 Introduction

1.1 The need for large-scale randomised controlled trials and meta-analyses

Although there are examples of some interventions that have strikingly beneficial effects on major disease outcomes, the majority of interventions have only moderate effects and hence are impossible to evaluate without careful study. Any clinical study whose main objective is to assess moderate treatment effects must therefore ensure that any biases and any random errors that are inherent in its design are both substantially smaller than the anticipated effect size (1). The only way to guarantee the avoidance of moderate biases is to randomise—with allocation concealment to prevent any foreknowledge of the next treatment assignment—while the only way to guarantee avoidance of moderate random errors is to study large numbers of outcomes. Large-scale randomised controlled trials (RCTs) are therefore typically needed when comparing the effect(s) of any new treatment (eg, a new drug) with a control group (eg, an established treatment or no treatment).

The process of randomisation avoids moderate biases because it results in groups of patients that differ from each other only by the play of chance with respect to their risks of suffering all types of health outcome, so observed differences in rates of health outcomes can generally be attributed causally to differences in study treatment (2). Furthermore, masking assignment of study treatment with a matching dummy placebo tablet, as is done in 'double-blind' RCTs (whereby neither the trial participant nor their caring doctor(s) knows what their randomised treatment assignment is), minimises the differential assessment of events between the study treatment groups. In contrast, non-randomised observational studies (such as analyses of health-care databases) in which an outcome is compared between individuals who received the treatment of interest and those who did not, may be subject to moderate systematic errors (ie, biases), and hence produce misleading results (1, 2). Even when attempts are made to

adjust for such biases (eg, propensity score matching) this cannot be *guaranteed* to remove any biases entirely.

While it may sometimes be possible to study outcomes sufficiently through a very large single RCT, meta-analysis, in which data from several RCTs are combined using statistical methods that preserve the like-with-like comparisons within each trial of the outcome under investigation (1, 2), is often required to achieve sufficiently small random errors. This is often the case where the effect of a treatment on an outcome is moderate and too few cases of it have occurred in any individual trial to assess the effect sufficiently reliably.

Meta-analysis can be used to amalgamate either individual participant data (IPD) or aggregate (tabular) data. IPD meta-analysis is more resource intensive than tabular meta-analysis, but it has several advantages over the tabular approach since it allows harmonisation of outcomes across the multiple trials, as well as, more detailed analyses, for example, in assessing treatment effects either over time or in particular types of patient.

RCTs and meta-analyses have been extensively used to research cardiovascular disease, notably in the field of lipidology and studies of blood cholesterol.

1.2 LDL cholesterol and cardiovascular disease

1.2.1 Epidemiological evidence

Cardiovascular disease (CVD) is the leading cause of adult mortality and morbidity worldwide (3, 4). An aetiological relationship between blood cholesterol concentrations and risk of cardiovascular (CV) morbidity and mortality has been established by observational, randomised and genetic studies, as described below.

Meta-analyses of data from prospective observational studies, such as those done by the Prospective Studies Collaboration (5) and the Emerging Risk Factors Collaboration (ERFC)

(6), have proven a log-linear positive association between non high-density lipoprotein cholesterol (or, approximately analogously, low-density lipoprotein [LDL] cholesterol) and the risk of atherosclerotic coronary heart disease (CHD), with no apparent threshold level below which a lower LDL cholesterol level does not associate with a lower risk. This association remains consistent across different populations (eg, various ages groups, ethnicities and by gender). Whilst a few observational studies have reported that lower concentrations of cholesterol have been associated with higher risk of all-cause mortality, particularly in the elderly (7, 8), such lower cholesterol concentrations are mainly caused by pre-existing disease (9) (a concept known as 'reverse causality').

Although the absolute difference in CHD risk associated with a given absolute difference in LDL cholesterol is greater at higher concentrations of LDL cholesterol, the proportional difference in risk associated with a given absolute difference in LDL cholesterol concentration is similar throughout the usual range of LDL cholesterol levels (5). The development of treatments targeted to reduce LDL cholesterol would therefore be expected to reverse the risk of CHD, irrespective of the starting cholesterol concentrations (without resulting in any excess in risk of all-cause mortality). This has been confirmed by randomised trials, with numerous RCTs of cholesterol lowering diets and drugs having been conducted (10, 11).

In addition, further genetic research studies using Mendelian randomisation methods (ie, a technique which provides evidence about putative causal relations between modifiable risk factors and disease, using genetic variants as natural experiments [by assuming that the variants associate with the outcome only through their effect on the risk factors examined] to reduce the risk of unmeasured confounding, often found in observational studies (12)) have shown that the positive association between raised LDL cholesterol levels and increased risk of CVD is indeed causal (13) , whilst equally demonstrating that an association with all-cause mortality is not demonstrated (ie, possibly due to lack of power) (14). For example, in recent analyses of 438,952 individuals of white ancestry from the UK Biobank, long-term exposure

to 1 mmol/L lower LDL cholesterol was associated with about a halving in the risk of developing CVD (15).

1.2.2 HMG-CoA reductase inhibitors ('statins')

Many of the pivotal RCTs of LDL cholesterol lowering used a class of drug known as HMG-CoA (3-hydroxy-3-methylglutaryl-CoA) reductase inhibitors, widely known as 'statins'. The discovery of statins was a breakthrough in the prevention of CVD. Statins reduce the production of LDL-cholesterol in the liver by about 50% or more by blocking the HMG-CoA-Reductase enzyme. **Table 1.1** presents the key characteristics of different statin types. There are numerous types of statin used in clinical practice, with lovastatin being the first type of statin approved for use in humans (initially for patients with very high risk of myocardial infarction) by the Food and Drug Administration (FDA) in 1987 (16). Seven further statins were subsequently granted marketing approvals: simvastatin in 1988, pravastatin in 1991, fluvastatin in 1994, atorvastatin in 1997, cerivastatin in 1998 (although later withdrawn due to drug safety concerns in 2001) and, finally, rosuvastatin and pitavastatin in 2003 (16). All these different drugs can be grouped into two main statin sub-types based on their structure: either fungal-derived (including lovastatin, pravastatin, and simvastatin), or synthetically produced (with atorvastatin, cerivastatin, fluvastatin, pravastatin, rosuvastatin, and pitavastatin being fully synthetic compounds) (17). In addition, two statin drugs (ie, pravastatin and rosuvastatin) can be classified as hydrophilic (ie, water soluble) while all other statins are lipophilic (ie, fat soluble).

1.2.3 Use of statins worldwide

Evidence from numerous large-scale RCTs and from meta-analyses, particularly those conducted by the Cholesterol Treatment Trialists' (CTT) Collaboration (www.cttcollaboration.org; see section 1.3), have demonstrated that statin therapy is highly effective in reducing the risk of CVD, both in terms of prevention of an initial CV event (primary prevention) and recurrence of a CV event (secondary prevention) (18). Statins are now off-

patent (ie, generic drugs), and highly cost-effective (eg, in the UK, the cost of such drugs is less than 10 pence per patient per day, see **Table 1.1**). They are widely available and used by more than 200 million people worldwide on a daily basis (19).

1.3 The Cholesterol Treatment Trialists' (CTT) Collaboration

The CTT Collaboration was established in 1994 after it was recognized that no single lipid intervention trial would be likely to have a sufficient number of outcomes (and hence statistical power) to reliably assess cause-specific mortality or outcomes in particular types of patient (20). Since then it has conducted periodic IPD meta-analyses of all large-scale (≥1,000 participants), long-term (≥2 years scheduled treatment duration) unconfounded (ie, double-blind) RCTs of statin therapy, with previous analyses involving individual level data on major vascular events (MVEs: defined as coronary deaths, myocardial infarctions, strokes, and coronary revascularisations), cancers and mortality. The CTT has sequentially reported analyses as eligible RCTs of statin therapy contributed their results to the CTT Secretariat, with the CTT database for the most recently published analyses including data from 28 major statin trials (involving approximately 175,000 randomised participants). **Table 1.2** presents the main characteristics of the participating trials.

1.3.1 CTT efficacy findings: Effects of statin therapy on major vascular events

Key findings from the CTT collaboration of the effect of statins on MVEs in different types of people and over time are presented in **Figures 1.1** and **1.2**. Figure 1.1 presents the proportional reductions in risk of MVEs per 1.0 mmol/L reduction in LDL cholesterol in different types of people, and figure 1.2 presents the proportional reductions in risk of MVEs per 1.0 mmol/L reduction in LDL cholesterol during each year of scheduled statin treatment.

The first large-scale CTT IPD meta-analysis of trials was published in 2005 and compared statin therapy to placebo or open control, and included data from about 90,000 individuals (19). The main finding was that for each 1 mmol/L reduction in LDL cholesterol caused by a

statin, the rate of MVEs was reduced by about one quarter during each year after the first. There were also beneficial effects on MVEs in the first year of treatment, but these were about half that of subsequent years (19) (**Figure 1.2**).

Standard statin regimens reduce LDL cholesterol concentrations by about a third, but regimens involving higher doses and newer type of statins can reduce LDL cholesterol concentration by about one half. However, it was uncertain (at that time [2005]) whether larger reductions in LDL cholesterol would lead to further reductions in risk of major vascular disease. The CTT Collaboration therefore extended its initial database to incorporate data from large randomised trials comparing more versus less intensive statin regimens and additional new trials comparing statin therapy with placebo or open control. The second IPD meta-analysis published in 2010 (18) therefore included data from 5 trials of more versus less intensive statin therapy (involving about 40,000 individuals) and data from 21 trials of statin versus control (involving about 130,000 participants). This analysis reinforced the earlier findings of the effects of statins on MVEs when compared with no statin, showing a 21% reduction in MVEs per 1.0 mmol/L reduction in LDL cholesterol. Further reduction in LDL cholesterol through use of a high-dose statin produced a further reduction in risk: on average allocation to high-dose statin therapy reduced LDL cholesterol by a further 0.5 mmol/L which resulted in a further 15% proportional reduction in MVEs (18).

These, and other, CTT analyses have shown that statin therapy is effective in a wide range of people (**Figure 1.1**). In particular, the proportional reduction in MVE risk per 1.0 mmol/L reduction in LDL cholesterol was similar irrespective of baseline cholesterol levels, prior history of diabetes, prior history of vascular disease or other patient characteristics such as BMI, blood pressure, gender and predicted risk of vascular disease (see **Figure 1.1**) (18, 21-24).

The absolute expected benefit from taking a statin is determined by an individual's absolute expected risk of a MVE in the absence of statin therapy and the absolute reduction in LDL

cholesterol achieved by taking the statin. Translating this into clinical practice, lowering LDL cholesterol by 2 mmol/L with an effective low-cost statin regimen (eg, atorvastatin 40 mg daily, costing about £1.0 to £2.0 per month; see **Table 1.1**) for 5 years in 10,000 patients would typically prevent MVEs from occurring in about 1000 (10%) patients with pre-existing occlusive vascular disease (secondary prevention) and in 500 (5%) patients who are at increased risk but have not yet had a vascular event (primary prevention) (23). Given that statin therapy has been shown to reduce vascular disease risk during each year it continues to be taken, larger absolute benefits would accrue with more prolonged therapy, and these benefits persist long term (18, 19).

1.3.2 CTT safety findings: Effects of statin therapy on cancer and non-vascular mortality

By 2012, the CTT Collaboration had collected detailed information on cancer incidence and cancer mortality for 175,000 people in 27 randomised trials, in which over 10,000 participants developed cancer and over 3,500 died from cancer. Consequently, it was very well placed to test not only the hypothesis of whether statins might increase (or perhaps decrease) the risk of cancer overall, but also whether statins might cause any particular type of cancer, cancer in any particular type of patient, or cancer only with longer duration of treatment (as would generally be expected for exposure to any true carcinogen). The results from these analyses were very reassuring. Reducing LDL cholesterol with a statin had no effect on newly diagnosed cancer or on death from such cancers in either the trials of statin versus placebo or in the trials of more versus less statin (25). Moreover, there was no evidence of any effect of reducing LDL cholesterol with statin therapy on cancer incidence or mortality in any particular type of patient, and no evidence that statins caused cancer with increased duration of treatment. (25).

In addition, the CTT Collaboration has collected information on cause specific mortality including different types of non-vascular mortality (ie, due to cancer, respiratory problems,

trauma or other non-vascular causes). Overall, over 5,900 patients in the trials died from a non-vascular cause (including 2,300 from a non-cancer non-vascular cause). There was no evidence that allocation to a statin regimen was associated with any effects (beneficial or harmful) on any non-vascular cause of death, or on the composite of all non-vascular deaths (2, 18). Considering all deaths together, and bearing in mind that patients in the trials were often *selected* to be at high risk of CVD, allocation to a statin or more intensive statin regimen in these trials was associated with a proportional reduction in all-cause mortality of 10%. (2, 18).

1.3.3 Effects of statin therapy on other outcomes

It is known that statins can, rarely, cause biochemical changes (eg, rises in creatine kinase) that provide evidence of significant muscle damage in association with muscle symptoms (i.e. myopathy or, in a more severe form, rhabdomyolysis) (2). However, the absolute excess risk of such complications is very small being about 1 extra case per 10,000 people treated per year for myopathy, and about 2-3 cases per 100,000 people treated per year for rhabdomyolysis). In addition, the RCTs of statins have shown that statins cause small absolute increases in the incidence of haemorrhagic stroke and of diabetes (2). However, it has been postulated that statins may have additional adverse effects, and this has led to growing confusion and concern amongst both doctors and patients, resulting in many people at high risk of vascular disease potentially endangering their health by stopping or not starting statin therapy. This is clearly a substantial public health issue.

Such concerns have arisen chiefly based on data from non-randomised observational studies which reported that statin therapy commonly cause an increased risk of a range of other adverse events, including muscle pain (known as myalgia) (26), hepatic dysfunction (27), cataracts (27), depression (28), impaired cognition (29) and acute kidney injury (30), with some such events being added to statin prescribing labels as ‘undesirable effects’. Conversely, other observational studies have suggested statin therapy may be associated

with a reduced risk of other outcomes (including respiratory disease, Parkinson's disease and arrhythmias) (31-34).

As explained in section 1.1, such concerns and queries cannot generally be addressed by non-randomised studies because such studies are susceptible to moderate (and sometimes more than moderate) systematic errors (ie, biases). Specifically, there may be differences between people who are given a statin and people who are not which are causally related to their risk of particular health outcome, but these differences are impossible to measure (sufficiently well or, even, at all) and hence impossible to eliminate through any statistical adjustment. Instead, the most robust way to assess any such treatment effect is through a meta-analysis summarising all of the available evidence captured in all large-scale RCTs of statin therapy.

As previously mentioned, the original data collected by the CTT Collaboration focused only on the effect of statin therapy on MVEs, cancer and non-vascular mortality, and so the CTT database was previously unable to assess outcomes such as those mentioned above. However, in 2016, a new programme of CTT work was initiated whereby agreement was secured from the 28 CTT collaborating trial sponsors to provide supplementary files of IPD including data on *all* adverse events (AEs), serious or otherwise, that had been recorded in each statin trial. This included AE timing, supplemented by data on treatment adherence and non-trial statin use, non-trial co-medications, laboratory results and co-morbidities that may help in interpretation of particular AEs. Based on the 2016 published protocol (35), this CTT programme of work has systematically categorised these events, and developed an analysis strategy to explore the association of statin therapy across the vast range of different AE types. Broadly speaking, such CTT outcomes can be subdivided into 2 main classes: AEs probably or possibly related to statin therapy (ie, for which there is a confirmed or widely reported prior hypothesis); and those for which there is no such reported association (constituting the vast majority of AEs).

1.3.3.1 AEs probably or possibly associated with statin therapy

Such AEs can be subdivided into three different types:

1. *Known adverse effects of statins:* As aforementioned, previous studies have demonstrated that statins cause small increases in the incidence of the serious muscle outcome myopathy (and, rarely, its more extreme form rhabdomyolysis), new-onset diabetes and, probably, of haemorrhagic stroke (2). These events were examined (by the CTT collaboration) to explore the magnitude, timing and duration of the excess risks both overall and in particular types of people.
2. *Muscle symptoms:* Since one of the main reasons given for intolerance to a statin is muscle symptoms that do not qualify as myopathy, detailed analyses considered other muscle symptoms such as muscle pain (ie, myalgia) and muscle weakness separately from myopathy, and will explore whether statins increase reported rates of muscle pain of different levels of severity (including, for example, muscle symptoms given as a reason for stopping study treatment).
3. *Other conditions previously associated with statin use:* Some observational studies have suggested that statins are associated with an excess risk of particular AEs (eg, cognitive impairment, depression, interstitial lung disease), as well as some conditions where the risk may be reduced (eg, pancreatitis). For each condition where a hypothesis has been described in the CTT 2016 protocol, the effects of statins will be quantified.

1.3.3.2 AEs with unknown association with statin therapy

The CTT programme of work assessing AEs also planned to examine the effects of statins on all other AEs (ie, including those for which there is no widely reported prior hypothesis for an association with statin therapy). Due to the fact that a very large number of AEs is expected to be examined (ie, >1000 different types of unique AEs), this last category faces considerable

challenges related to 'multiple testing'. This category of AE, and the methods devised to address the problems associated with multiple testing, is the focus of this DPhil thesis[a].

1.4 Interpreting statistical evidence

1.4.1 The null hypothesis

Statistical techniques (specifically 'frequentist' techniques) are used to estimate the magnitude of any effect(s) of a new treatment compared with an established treatment (or no treatment) and the extent to which such effect estimates could have arisen by chance. More precisely, they quantify the extent to which observed differences in outcome event rates between the treatments are compatible with some predefined 'null' hypothesis (which is a statement about the 'true' unknown effects of a treatment in a population being studied)[b]. This is often done through the calculation of a 'p-value', which is the probability that, if the null hypothesis is true, a difference in treatment effects *at least as large* as that actually seen in the data would have arisen. A small p-value therefore provides evidence against the null hypothesis, with a p-value <0.05 often accepted as being sufficiently small to allow 'rejection' of the null hypothesis in favour of the alternative hypothesis.

1.4.2 Type of null hypothesis

When comparing the effects of two interventions, the null hypotheses (denoted H_0) may be either one or two-sided. In a one-sided null hypothesis (H_0: $\mu 1 \leq \mu 2$, where a high value of μ

[a] Note that certain AEs types might only be expected to occur very rarely in trial participants, but could be informative as single cases if known to be strongly associated with drug exposure. Specific examples of such AEs including the conditions anaphylaxis and Stevens-Johnson Syndrome. Such types of rare AEs are beyond the scope of this project and will therefore not be discussed in detail in the remainder of this thesis.

[b] Two types of statistical inference exist, Frequentist and Bayesian, which are designed to answer two slightly different questions. Frequentist methods answer the question, 'What is the likelihood of the observed data given some assumption about the true treatment effect?'. By contrast, Bayesian methods answer the question, 'Given some prior beliefs and the observed data, what is the best estimate of the treatment effect?'. Frequentist methods have dominated the statistical analysis of randomised controlled trials and the meta-analysis of multiple trials, especially in the area of drug efficacy and safety, where scientists and regulatory authorities are very experienced in drawing conclusions and making policy based on such methods. Bayesian methods are used much less frequently, for reasons that are discussed in Chapter 2.

may be assumed to correspond to a 'good' outcome) the aim is to prove that one treatment is superior to another by first assuming that it is not and then hoping to gather sufficient evidence to reject that belief. By contrast, in a two-sided null hypothesis (H_0: $\mu1=\mu2$) it is initially assumed that both interventions have the same effect (which may be no effect) and gather sufficient evidence to reject that belief. Two-sided significance tests are normally preferred in clinical trials where it cannot be known for certain from the outset whether a new treatment is beneficial or harmful[c].

1.4.3 Limitations of the null hypothesis

1.4.3.1 False positive (the 'type I' error, or 'alpha')

Rejection of the null hypothesis is often done at the 5% significance level (ie, reject H_0 if $p<0.05$) but this threshold is entirely arbitrary, having seemingly been first used by Fisher in 1925 (36) (who accepted a probability of 5% that the null hypothesis would be rejected incorrectly as a limit for statistical inference). In statistics, the probability of erroneously rejecting the null hypothesis (ie, concluding that a treatment effect exists when the truth is that it doesn't, or a 'false positive' result), is known as the *type I* error or *alpha* error rate. The selection of a 5% threshold level is widely used in practice for studies where the focus is on answering reliably a *single* question (eg, in a RCT comparing a new treatment versus placebo for a particular endpoint).

1.4.3.2 False negative (the 'type II' error, or 'beta')

When performing hypothesis testing, the p-value is typically used to draw inferences about whether there is any statistical evidence to 'reject' the null hypothesis. However, failing to reject the null hypothesis does not mean that the null hypothesis is true (ie, 'absence of evidence' is not 'evidence of absence') (37). The probability of erroneously failing to reject the null hypothesis (ie, failing to conclude that a treatment effect exists when the truth is that it

[c] Note that other types of hypothesis tests exist in medical research (eg, non-inferiority and equivalence trials) but they are not the focus in this thesis

does, or a 'false negative' result) is known as the *type II* error or *beta* error rate. The 'statistical power' of a study (ie, the ability of a study to detect a true treatment effect if one truly exists) is therefore calculated as a percentage from the equation 100 x (1 – type II error probability).

1.4.4 Multiple hypothesis testing (MHT)

Many studies are designed to answer more than one question (eg, multi-arm RCTs, or RCTs with more than one endpoint) leading to a series of hypotheses which need to be tested together as a group. In such experiments, the probability of detecting at least one false positive result is inflated in line with the number of hypotheses tested. For example, when testing 20 independent hypotheses of an ineffective drug, the probability of at least one false positive result (at the 5% type I error level) would be 0.64 (since $1-0.95^{20}=0.64$), a probability that would increase to 0.94 with 100 independent tests of an ineffective drug. This **multiple hypothesis testing (MHT)** problem is common in clinical research, particularly in the era of genetic studies when the number of tests performed may be in the thousands, hundreds of thousands, or even millions.

1.5 Concepts of multiple hypotheses testing

This section introduces the main concepts of multiple hypotheses testing, including the definition of a hypothesis 'family', the two most common statistical error rates used, and a summary of the various types of hypothesis-testing correlation structure that exist. (A detailed description of all MHT concepts and characteristics of MHT methods is provided in chapter 2.)

1.5.1 Family of hypotheses

When a group of hypotheses is tested together in a clinical trial, this constitutes a *'family'* of hypotheses. When a family of hypotheses is to be tested, MHT methods are used to reduce the likelihood of any false positive results. The size of the family should determine the threshold level of significance at which any individual hypothesis within the family is rejected

or not. For example, a p-value <0.01 for each test may be considered as appropriate evidence against the null hypothesis when analysing a family of 5 hypotheses, but when assessing a family of 50 hypotheses a p-value <0.001 for each test may be deemed necessary.

1.5.2 Controlling for statistical errors

The type I and type II error rates reflect the possible errors in interpretation arising from a single hypothesis test[d], but in the context of MHT other error rates are of more interest, with the 'familywise error rate' and the 'false discovery rate' being the two most popular and highly preferred in practice. These concepts are described below.

1.5.2.1 Familywise Error Rate

The Familywise Error Rate (FWER) is the probability of *at least one* false positive result in a family of *m* tests. So, if, for example, testing a family of 20 independent hypotheses with a 5% type I error per hypothesis, the FWER would be equal to 0.64 (since, as described above, $1-0.95^{20}=0.64$). Control of the FWER is desirable when one wishes to ensure that the overall likelihood of discovering a false positive result does not exceed a selected level of significance (eg, α=0.05). For example, in RCTs, the control of FWER may be used to allow for robust conclusions about the effectiveness of a new treatment on multiple individual components of a composite primary outcome, or when testing the effect of the drug in different pre-defined sub-populations. In practice, methods that guarantee control of the FWER are generally preferred when the number of hypothesis tests in the family is relatively small (eg, <20).

1.5.2.2 False Discovery Rate

The False Discovery Rate (FDR) is the *proportion* of all significant results that are *expected* to be false positives. For example, if, when applying an FDR error of 5% when testing a family of 200 independent hypotheses, 40 of those hypotheses are rejected, then it is expected 2 of those hypotheses have been rejected incorrectly. The FDR is used when control of the number

[d] Another error rate is found in the literature when correctly rejecting the null hypothesis but for the wrong reason. Such error is known as *type III* error but it will not be considered further in this project as it is not particularly relevant to this topic.

of false positives is of more interest than the simple question of whether any false positive conclusions are drawn (38). Use of the FDR is often preferred in very large-scale MHT problems, such as genetic studies (39, 40) where thousands of tests may be performed. By contrast, the use of the FWER for such studies would be considered as a conservative approach (due to its low overall probability of identifying one or more true positive results: see later chapters).

1.5.3 Correlation of hypotheses in MHT

The correlation structure of multiple hypotheses is heavily discussed in the literature and various MHT methods have been developed to account for different types of correlations, which can be classified into three broad groups:

1.5.3.1 Assumption free

These methods control for MHT of a family of hypotheses *without any assumptions being made about their dependency structure*. As a result, they are often a preferred choice. Two examples are the methods of Holm and Benjamini & Hochberg (described in detail in chapter 2).

1.5.3.2 Independent

This approach describes a scenario whereby the hypotheses tested are *all assumed to be independent of each other* (ie, for every hypothesis in the family, the statistic used to test that hypothesis is uncorrelated with the statistics used to test every other hypothesis in the family). In the context of clinical trials, independent hypotheses would arise if there is no correlation between the treatment effects on multiple outcomes. For example, for statins, it may be reasonable to assume that their (LDL cholesterol mediated) effect on reducing the risk of myocardial infarction would be unrelated to the effect (if any) on depression or myalgia. (Of course, if the effects, beneficial or otherwise, of statins were always mediated by their effects on LDL cholesterol, then this might not be a reasonable assumption). Several examples of such independent statistics can be found in the literature (41, 42).

1.5.3.3 Correlated

Various MHT methods exist in the literature specifically designed to account for the correlation among the multiple hypotheses tested. These MHT methods can be classified into two different groups based on the statistical techniques they employ as described below.

1. *Multivariate analysis:* this group includes MHT methods designed to account for hypotheses which are highly correlated by allowing for multivariate distributions (ie, a statistical approach which accounts for the joint distribution among outcomes). Such MHT methods are used in practice, for example, when testing the effectiveness of various active therapies, or doses of a therapy, against placebo. In this case, all active therapies are generally expected to have a non-negative effect when compared with placebo.
2. *Correlation by construction:* many MHT methods are designed to organise the multiple hypotheses into specific structures which allows hypotheses to be correlated by the construction designed. Several examples exist in the literature where hypotheses are organised into a hierarchical setting (ie, grouping the hypotheses into different levels and testing them based on their level) or hypotheses are designed to be logically related (ie, when it is not possible to perform all combinations of hypotheses that can be tested). For example, the effect of a drug on non-fatal myocardial infarction would be correlated (by construction) with its effect on any (fatal or non-fatal) major coronary event.

1.6 Aims of this thesis

The current CTT programme of work assessing all effects of statin therapy provided a challenging example of where MHT needs to be considered due to the sheer number of different types of adverse event (AE) recorded in the included trials, the way they may be either grouped or considered separately (leading to correlated effect size estimates for related

categorizations) as well as the potential desire to consider trials of statin vs placebo separately from trials of high-dose vs standard-dose statin. There was therefore a need to develop a suitable statistical framework to address the MHT problem in the CTT AE project, and to pre-specify such an approach prior to any unblinding of results, so that reliable conclusions could be drawn about the possible effects of statin therapy on AEs for which there is currently no widely accepted effect.

The aims of this thesis were to develop such a framework which both: (i) controlled the number of false positive results; and (ii) did so while maintaining as much statistical power as possible to detect any real, previously unknown effects of statin therapy. The content of the subsequent chapters is given below:

- Chapter 2 describes the results of a systematic literature search done to identify all potentially relevant MHT approaches, with an overview of how each method is implemented and interpreted.
- Chapter 3 describes the data-handling approaches taken in the CTT project, including methodologies used in the harmonisation of the data and standardisation of the outcomes. In addition, it describes the methods developed for conducting simulation analyses designed to evaluate the performance of the various MHT approaches for different types of AE under different assumptions. It also describes the general statistical approaches used to perform the meta-analyses.
- Chapter 4 presents the results from the simulation analyses and identifies and justifies the MHT statistical framework to be used for the primary analysis of the AEs.
- Chapter 5 implements the MHT framework and presents analyses of the effects of statin when compared with placebo on all relevant AEs.
- Chapter 6 implements the MHT framework and presents analyses of the effects of more vs less statin therapy on all relevant AEs.
- Chapter 7 summarises the key findings, discusses the strengths and limitations of the work, and considers the implications for public health and future research.

Table 1.1: Characteristics of the various type of statins used in clinical practice

Type	Year of marketing approval	Structural mechanism		Solubility	Daily dosage in mg, by intensity (LDL-C reduction)†			Cost for 28 days regimen‡ (in Pounds)
		Synthesis¶	Chemical structure§		Low (<30%)	Moderate (30% to 50%)	High (>50%)	
Lovastatin	1987	Natural	Type I	Lipophilic	20	40 to 80	NA	NA
Simvastatin	1988	Semi-synthetic	Type I	Lipophilic	10	20 to 40	NA	1.13 to 1.30
Pravastatin	1991	Natural	Type I	Hydrophilic	10 to 20	40 to 80	NA	1,24 to 1.84
Fluvastatin	1994	Synthetic	Type II	Lipophilic	20 to 40	2X40 or 80	NA	4.18 to 19.20
Atorvastatin	1997	Synthetic	Type II	Lipophilic	NA	10 to 20	40 to 80	1.01 to 2.10
Cerivastatin*	1998	Synthetic	Type II	Lipophilic	NA	NA	NA	NA
Rosuvastatin	2003	Synthetic	Type II	Hydrophilic	NA	5 to 10	20 to 40	1.45 to 2.53
Pitavastatin	2003	Synthetic	Type II	Lipophilic	1	2 to 4	NA	NA

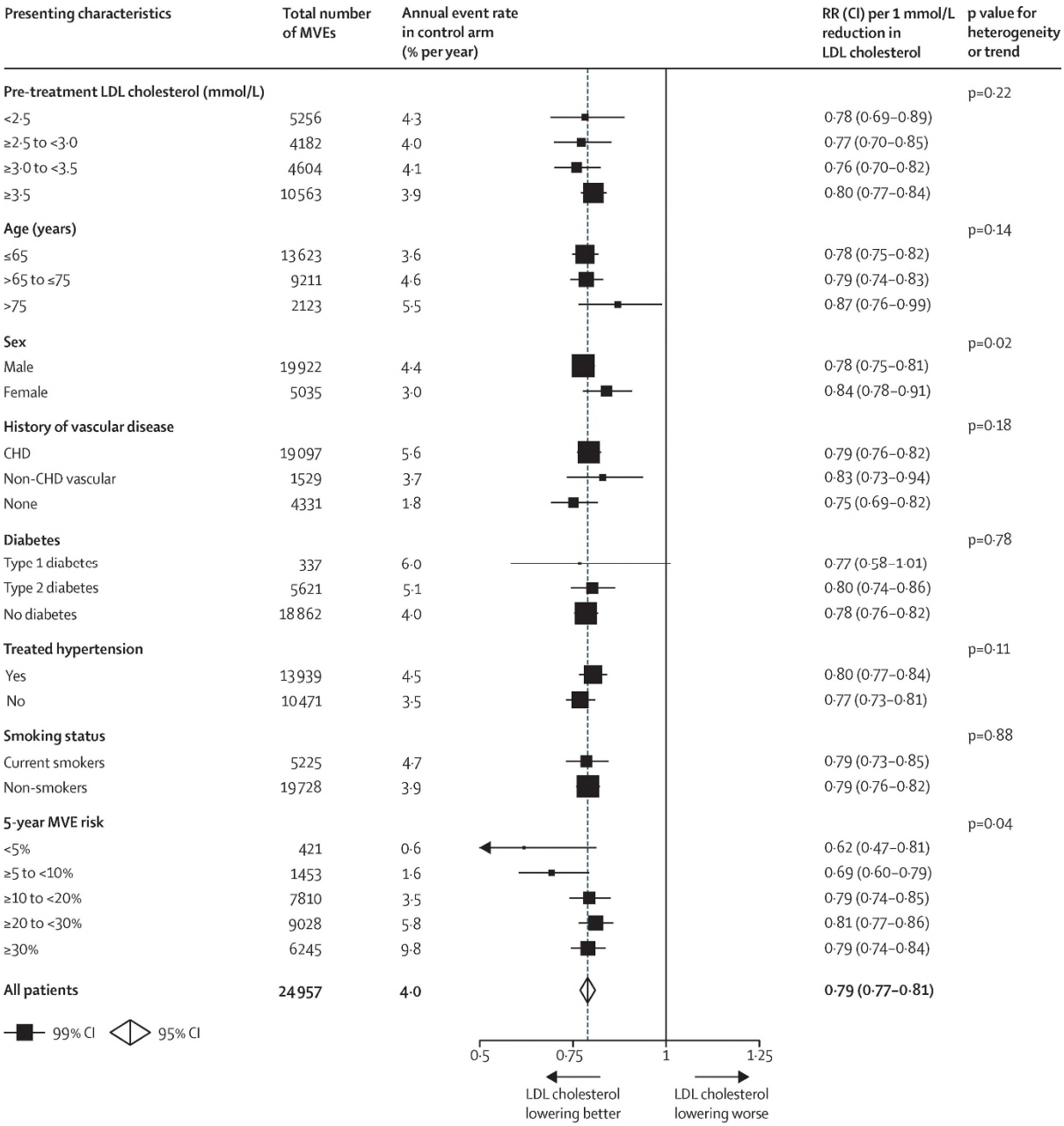
¶Natural statin are produced by fermentation of funghi; semi-synthetic statins are synthesised by the alkylation of lovastatin, replacing the 2-methylbutyrate moiety at the C-8 position of the naphthalene ring, with 2,2-dimethylbutyrate; and synthetic statins chemically synthesised, and only share the common dihydroxyheptanoic acid pharmacophore (HMG-CoA-like) moiety (43). **§**Type I statins contain a naphthalene ring; and type II statins contain a fluorophenyl group and a series of hydrophobic ring structures (43). †Based on recommendations from the American College of Cardiology and the American Heart Association (44, 45). ‡ Information about cost was based on data presented in the NHS Business Services Authority website visited in 22/06/2021 (46). *Cerivastatin was withdrawn from the market in 2001 due to drug safety concerns (ie, rare cases of rhabdomyolysis) (47). LDL-C=low-density lipoprotein cholesterol; NA=not applicable;

Table 1.2: Characteristics of the 28 major statin trials

Study	Year of publication of primary results	Number of patients	Treatment comparison (mg/day)	Median follow-up (years)	Baseline LDL-C (mmol/L)	Baseline Age (years)	Women (%)	White/ Caucasian (%)	History of Vascular disease (%)	History of Diabetes (%)	Body Mass Index (kg/m^2)
Statin vs. Placebo											
SSSS	1994	4444	S20-40 vs placebo	5.4	4.9 (0.7)	58.6 (7.1)	827 (19%)	NA	4444 (100%)	202 (5%)	25.6 (23.8 - 27.7)
WOSCOPS	1995	6595	P40 vs placebo	4.8	5.0 (0.5)	55.2 (5.5)	0 (0%)	NA	1066 (16%)	77 (1%)	25.6 (23.8 - 27.8)
CARE	1996	4159	P40 vs placebo	4.9	3.6 (0.4)	58.6 (9.3)	576 (14%)	3851 (93%)	4159 (100%)	586 (14%)	27.0 (24.6 - 29.9)
AFCAPS/TexCAPS	1998	6605	L20-40 vs placebo	5.0	3.9 (0.4)	58.2 (7.2)	997 (15%)	5860 (89%)	0 (0%)	155 (2%)	26.7 (24.7 - 29.0)
LIPID	1998	9014	P40 vs placebo	5.9	3.9 (0.8)	60.8 (8.4)	1516 (17%)	NA	9014 (100%)	782 (9%)	26.4 (24.2 - 28.9)
LIPS	2002	1677	F80 vs placebo	4.0	3.4 (0.8)	60.0 (10.0)	271 (16%)	1650 (98%)	1677 (100%)	202 (12%)	26.3 (24.2 - 28.9)
HPS	2002	20536	S40 vs placebo	5.2	3.4 (0.8)	63.5 (8.4)	5082 (25%)	19901 (97%)	17386 (85%)	5963 (29%)	27.1 (24.7 - 29.8)
PROSPER	2002	5804	P40 vs placebo	3.3	3.8 (0.8)	75.3 (3.3)	3000 (52%)	NA	2565 (44%)	623 (11%)	26.5 (24.0 - 29.3)
ASCOT-LLA	2003	10240	A10 vs placebo	3.3	3.4 (0.7)	63.2 (8.5)	1919 (19%)	9687 (95%)	1684 (16%)	2540 (25%)	28.1 (25.5 - 31.2)
ALERT	2003	2102	F40 vs placebo	5.5	4.1 (1.0)	49.7 (10.9)	715 (34%)	2039 (97%)	409 (19%)	369 (18%)	25.1 (22.7 - 28.0)
CARDS	2004	2838	A10 vs placebo	4.2	2.9 (0.8)	61.0 (8.1)	909 (32%)	2676 (94%)	106 (4%)	2838 (100%)	28.7 (25.5 - 32.2)
DDDD	2005	1255	A20 vs placebo	2.7	3.3 (0.8)	65.7 (8.3)	578 (46%)	924 (74%)	1041 (83%)	1255 (100%)	26.8 (24.4 - 30.4)
ASPEN	2006	2410	A10 vs placebo	4.0	2.9 (0.7)	60.4 (8.2)	811 (34%)	2029 (84%)	747 (31%)	2410 (100%)	29.0 (26.0 - 32.0)
SPARCL	2006	4731	A80 vs placebo	4.9	3.5 (0.6)	62.8 (11.2)	1908 (40%)	4415 (93%)	4731 (100%)	794 (17%)	27.1 (24.4 - 30.6)
CORONA	2007	4984	R10 vs placebo	2.8	3.6 (0.9)	71.7 (6.8)	1175 (24%)	NA	4984 (100%)	1475 (30%)	27.5 (22.0 - 27.5)
GISSI-HF	2008	4574	R10 vs placebo	3.9	3.1 (0.9)	67.6 (10.8)	1032 (23%)	4574 (100%)	4574 (100%)	1196 (26%)	26.6 (24.0 - 29.6)
JUPITER	2008	16714	R20 vs placebo	1.9	2.7 (0.5)	65.4 (7.9)	6374 (38%)	NA	0 (0%)	44 (0%)	27.5 (27.5 - 31.0)
AURORA	2009	2555	R10 vs placebo	3.9	2.6 (0.9)	63.7 (8.8)	969 (38%)	NA	1025 (40%)	658 (26%)	22.0 (22.0 - 27.5)
HOPE-3	2016	12705	R10 vs placebo	5.5	3.3 (0.9)	65.7 (6.4)	5874 (46%)	2546 (20%)	0 (0%)	731 (6%)	26.7 (24.1 - 29.7)
Statin vs. Other control/usual care											
POST-CABG	1997	1351	L40-80 vs L2.5-5	4.3	4.0 (0.5)	61.5 (7.3)	102 (8%)	1275 (94%)	1351 (100%)	116 (9%)	27.1 (24.9 - 29.8)
GISSI-P	2000	4271	P20 vs no treatment	1.9	3.9 (0.7)	58.9 (10.4)	587 (14%)	4271 (100%)	4271 (100%)	582 (14%)	26.2 (24.2 - 28.4)
ALLHAT-LLT	2002	10355	P40 vs usual care	4.7	3.8 (0.6)	66.4 (7.6)	5051 (49%)	5811 (56%)	4713 (46%)	3638 (35%)	29.1 (25.9 - 32.9)
ALLIANCE	2004	2442	A10-80 vs usual care	4.5	3.8 (0.7)	61.2 (8.8)	434 (18%)	2048 (84%)	2442 (100%)	521 (21%)	28.8 (26.1 - 32.2)
More intensive vs. less intensive statin (double-blind)											
PROVE-IT	2004	4162	A80 vs P40	2.1	2.6 (0.7)†	58.2 (11.2)	911 (22%)	3776 (91%)	4162 (100%)	734 (18%)	28.6 (25.7 - 32.4)
A to Z	2004	4497	S40 then S80 vs placebo then S20	2.0	2.1 (0.7) †	60.4 (10.7)	1100 (24%)	3825 (85%)	4497 (100%)	1059 (24%)	27.0 (24.5 - 30.1)
TNT	2005	10001	A80 vs A10	5.0	2.5 (0.5)	61.0 (9.0)	1902 (19%)	9410 (94%)	10001 (100%)	1501 (15%)	28.1 (24.6 - 32.2)
SEARCH	2010	12064	S80 vs S20	7.0	2.5 (0.6)	63.7 (8.9)	2052 (17%)	11854 (98%)	12064 (100%)	1267 (11%)	27.6 (25.2 - 30.4)
More intensive vs less intensive statin (open label)											
IDEAL	2005	8888	A40-80 vs S20-40	4.8	2.6 (0.8)	61.7 (9.5)	1701 (19%)	8822 (99%)	8888 (100%)	1069 (12%)	26.9 (24.7 - 29.4)

S=Simvastatin; P=Pravastatin; L= Lovastatin; F=Fluvastatin; A=Atorvastatin; R=Rosuvastatin; Age and LDL-C presented as mean(SD); BMI and eGFR presented as median(IQR); NA=Not available; SSSS=Scandinavian Simvastatin Survival Study; WOSCOPS=West of Scotland Coronary Prevention Study; CARE=Cholesterol And Recurrent Events; AFCAPS/TexCAPS=Air Force/Texas Coronary Atherosclerosis Prevention Study; LIPID=Long-term Intervention with Pravastatin in Ischaemic Disease; LIPS=Lescol Intervention Prevention Study; HPS=Heart Protection Study; PROSPER=PROspective Study of Pravastatin in the Elderly at Risk; ASCOT-LLA=Anglo-Scandinavian Cardiac Outcomes Trial-Lipid Lowering Arm; ALERT=Assessment of Lescol in Renal Transplantation; CARDS=Collaborative Atorvastatin Diabetes Study; 4D=Die Deutsche Diabetes Dialyse Studie; ASPEN=Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus; SPARCL=Stroke Prevention by Aggressive Reduction in Cholesterol Levels; CORONA=Controlled Rosuvastatin Multinational Trial in Heart Failure; JUPITER=Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin study group; GISSI-HF=Gruppo Italiano per lo Studio della Sopravvivenza nell'Insufficienza cardiaca; AURORA=A Study to Evaluate the Use of Rosuvastatin in Subjects on Regular Hemodialysis: An Assessment of Survival and Cardiovascular Events; HOPE-3=Heart Outcomes Prevention Evaluation-3 trial; ALLHAT-LLT=Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ALLIANCE=Aggressive Lipid-Lowering Initiation Abates New Cardiac Events; GISSI-P=Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico; Post-CABG=Post-Coronary Artery Bypass Graft; PROVE-IT=Pravastatin or Atorvastatin Evaluation and Infection Therapy; A to Z=Aggrastat to Zocor; TNT=Treating to New Targets; IDEAL=Incremental Decrease in End Points Through Aggressive Lipid Lowering Study Group; SEARCH=Study of the Effectiveness of Additional Reductions in AURORA, CORONA and JUPITER supplied age in categorical bands - and so the midpoint of each categorical band was used as a surrogate for baseline age in all analyses. A small number of participants in the AURORA(n=218), CORONA(n=27) and JUPITER(n=1088) trials withdrew consent for use of their data post-trial, and hence data from these participants is excluded. The ASCOT-LLA trial excludes 65 patients for whom data was not available due to protocol violations. †These two trials did not have active run-in periods; the values shown are estimated on-treatment LDL cholesterol concentrations in the standard statin group.

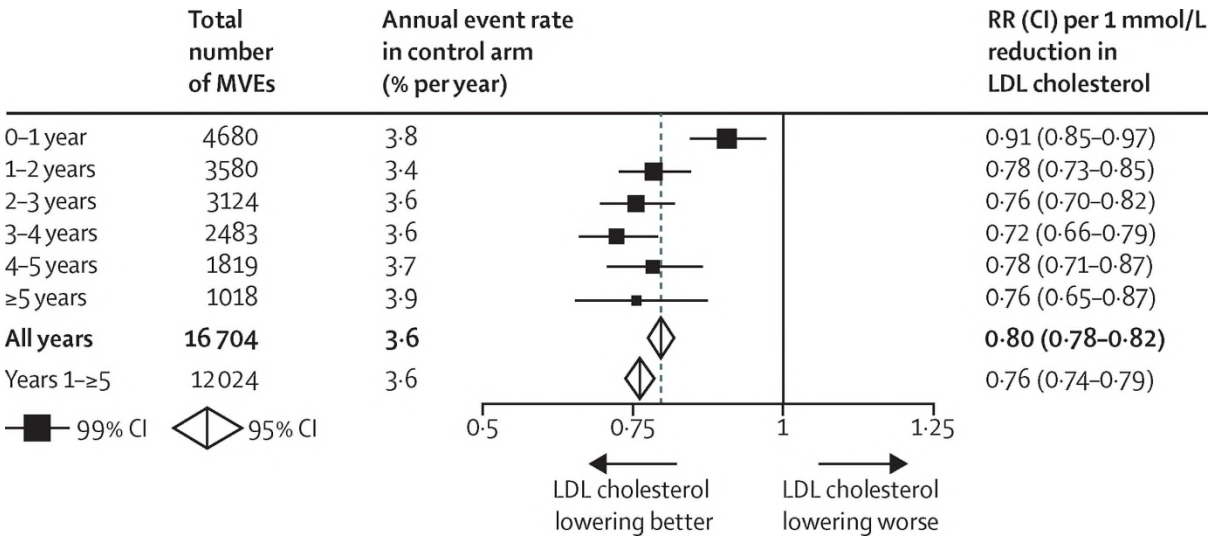
Figure 1.1: Effects on major vascular events per 1.0 mmol/L reduction in LDL cholesterol in randomised trials of statin therapy among people with different presenting characteristics



RRs are plotted for the combined comparisons of MVE rate in randomised trials of routine statin therapy versus no routine statin therapy and of more versus less intensive statin therapy, weighted per 1·0 mmol/L LDL cholesterol reduction at 1 year. The size of the squares is proportional to the numbers of events recorded (ie, statistical information) in the particular comparison. CHD=coronary heart disease. RR=rate ratio. MVE=major vascular event.

Data adapted from the CTT Collaboration website and also shown in the review paper produced in 2016 (2).

Figure 1.2: Proportional reductions in risks of major vascular events per mmol/L reduction in LDL cholesterol during each year of scheduled statin treatment, in randomised trials of routine statin therapy versus no routine statin use



For each time period, RRs weighted by trial-specific LDL cholesterol reductions at 1 year relate to participants at risk of a first post-randomisation major vascular event during the time period in the meta-analysis of trials of routine statin therapy versus no routine statin therapy. Consequently, the overall RR of 0·76 for the period after the first year indicates that risk is reduced by about one-quarter in each year that treatment continues (ie, the absolute benefits increase with increasing duration of treatment). As non-compliance to the randomly assigned treatment increased with longer duration in the trials (because of study statin therapy being stopped or because of statin therapy being started in the control group), the per mmol/L reductions based on LDL cholesterol reductions at 1 year are likely to underestimate the reductions in MVE risk per mmol/L LDL cholesterol reduction later in these trials. p<0·0001 for test of heterogeneity between RR in first year and RR in 1 -≥5 years. RR=rate ratio. MVE=major vascular event.

Data adapted from the CTT Collaboration website and also shown in the review paper produced in 2016 (2).

2 Literature review of MHT methods

2.1 Summary

This chapter describes the methods and findings of a systematic review of the literature on MHT methods, and identifies those methods potentially relevant for use in the meta-analysis of AEs with a previously unknown association with statin therapy. Of 9,040 articles identified, 2,471 were duplicates and a further 6,240 were excluded on the basis of their title or abstract, leaving 337 articles for full text review (including 8 further articles found upon close review of the references). Of the methods identified, only frequentist MHT approaches designed to allow for two-tailed hypothesis with certain types of correlation between hypotheses (ie, assumption free or specific types of correlations by construction), and which made no additional assumptions about the total number of true null hypotheses, were considered suitable for use in the CTT project. Nine such MHT methods were identified, five of which are designed to control the FWER (Bonferroni, Holm, Hochberg, Rom and Hommel) and four of which are designed to control the FDR error rate (Benjamini & Hochberg, Benjamini & Liu, Mehrotra & Heyse and Mehrotra & Adewale). Of these, five methods (three controlling the FWER and two controlling the FDR) were chosen, based on their statistical control of type I and II error rates, as the most appropriate and taken forward for further exploration through simulation analyses. These methods were: Holm, Hochberg and Hommel (controlling the FWER); and Benjamini & Hochberg and Mehrotra & Adewale (controlling the FDR).

2.2 Aim of this chapter

The aim of this chapter is to identify all potentially relevant MHT methods for use in the CTT project in the assessment of AEs with a previously-unknown association with statin therapy. In order to achieve this, a systematic search strategy was developed to identify all MHT

methods available in the literature. Then, a detailed review of those methods was performed to identify the most appropriate (ie, based on various MHT characteristics and statistical properties) MHT methods for use in the CTT project, which would be further explored through simulation analysis of the CTT AE data as described in chapter 4.

The outline of chapter 2 is as follows:

- Overview of the characteristics of available MHT methods and identification of those characteristics considered suitable for use in the CTT project.
- Description of the search strategy developed to identify all MHT methods available in the literature, and a presentation of the results from the systematic search.
- Detailed overview of how each of the MHT methods found potentially suitable to use in the CTT project is implemented and interpreted in practice (with methods organised by type of error control, starting with FWER and then FDR). Then, further review to identify the most appropriate methods based on their statistical properties (ie, control of type I and II error rates) which will then be further explored through the use of simulation studies when applied to the CTT AE database (as presented in chapter 4).

2.3 Characteristics of MHT methods

The problem of MHT is very common in medical research (and other research areas such as, for example, engineering and economics) with a wide variety of MHT methods developed to analyse different types of practical problems. Often, the MHT methods are designed specifically to account for various characteristics such as, for example, a particular type of null hypothesis or a certain type of correlation structure among the hypotheses, in order to maintain as much statistical power as possible. This section presents the various characteristics used in the development of such MHT methods, and highlights those MHT methodological characteristics considered as suitable for use in the CTT AE project.

2.3.1 Type of null hypothesis

As introduced in chapter 1, null hypotheses can be designed as either:

I. Two-tailed: hypothesis can be used when testing equality in the form of H_0: $\mu_1 = \mu_2$, (eg, if drug A is equally effective with drug B). Such a two-tailed approach is preferable for many hypothesis testing problems, and the majority of MHT methods have therefore been designed to analyse two-tailed null hypothesis.

II. One-tailed: hypothesis can be used for directional hypotheses in the form of H_0: $\mu_1 \leq \mu_2$ (eg, testing if one treatment is superior to another). Various MHT methods designed to test one-tailed null hypotheses exist in the literature with Dunnett & Gent (48) and Bauer et al. (49) presenting some methods controlling the FWER, while Gu & Shen (50) presenting an approach controlling the FDR.

Since the purpose of the CTT AE project is to detect any previously unknown effects of statin therapy, which could be either beneficial or harmful, only the use of MHT methods designed for 'two-sided' hypothesis are considered as suitable.

2.3.2 Control of statistical error rate

Unlike in single hypothesis testing where the control of the type I error rate (ie, the probability of erroneously rejecting the null hypothesis) is of interest, in the context of multiple hypothesis testing other error rates are of more interest, in particular the 'familywise error rate' and the 'false discovery rate' (see also chapter 1.4). These and some other error rates are described below, with reference to the notation employed in **Box 2.1,** which summarises the possible scenarios arising when testing a family of *m* hypotheses.

2.3.2.1 FamilyWise Error Rate

The FamilyWise Error Rate (FWER) is the probability of *at least one* false positive result in a family of *m* tests:

$$FWER = P(V \geq 1),$$

(where P denotes 'probability')

Control of the FWER is preferred in MHT problems that require strong control of the overall probability of discovering a false positive result (see also chapter 1.5.2.1). RCTs are the obvious example where the control of FWER is required to allow for robust conclusions about the effectiveness of a new treatment on multiple individual components of a composite primary outcome. Other examples of multiple testing problems encountered in RCTs where control of the FWER may be particularly desirable include when different doses of treatments in multi-arm trials are to be assessed, or when testing the effect of a drug in different pre-defined sub-populations. In practice, methods that guarantee control of the FWER are generally preferred when the number of hypothesis tests in the family is relatively small (eg, <20) (see also chapter 1.5.2.1).

2.3.2.2 False Discovery Rate

The False Discovery Rate (FDR) is the *proportion* of all 'significant' results (ie, rejected null hypotheses) that are expected to be false positives (ie, rejected incorrectly):

$$FDR=E(V/R),$$

(where E denotes the mathematically 'expected' value)

For example, if, when applying an FDR error of 5% when testing a family of 1000 independent hypotheses, 100 of those hypotheses are rejected, then 5 of those hypotheses would be expected to have been rejected incorrectly. The FDR is generally preferred in large-scale MHT problems, when control of the number of false positives is of more interest than the simple question of whether any false positive conclusions are drawn (38), for example, in genetic studies (39, 40) where tens or hundreds of thousands of tests may be performed (see also chapter 1.5.2.2). Note that the FDR equals the FWER if all *null* hypotheses tested within a family are *true* (eg, no effect of a treatment exist in any outcome tested) (51).

2.3.2.3 Other Error Rates

The problem of MHT is a multidimensional problem found in many practical applications, and numerous additional MHT approaches distinct from the FWER and FDR have been designed to control for various error rates. These error rates are sometimes defined based on the distribution of the *False Discovery Proportion (FDP)*, which is the proportion of rejected hypotheses that are false positives:

$$FDP=V/R$$

(Note that the FDR error is simply the expected value of the FDP error)

Three such error rates include:

I. ***k*-familywise error rate (*k*-FWER)** (52, 53) is a generalization of the FWER which controls the probability of at least *k* false significant results among a family of *m* tests:

$$k\text{-}FWER=P(V\geq k)$$

II. ***k*-false discovery rate (*k*-FDR)** (52, 54) is a generalization of the FDR. It is the proportion of all significant results that are expected to be false positives given that there are at least *k* false positive results:

$$k\text{-}FDR=E[V/R \mid V\geq k]$$

III. **False discovery exceedance (FDX)** (55, 56) is the probability that the proportion of false significant results (V/R) is greater than some level *c*:

$$FDX=P(V/R>c)$$

The control of k-FWER, k-FDR and FDX is found to be useful when the total number of correctly rejected hypotheses is large and their distribution (of p-values) is very different compared with the distribution of the true null hypotheses, which is uniform. In such cases, the expected value of the FDP (ie, the FDR) is not a representative parameter of the FDP (55, 57). Another instance where such methods are used is when the tail probability of the FDP is

of most interest (ie, probability of the FDP falling below a specific value, for example, less than 0.01) (55). More details about control of the FDP and the distribution of the associated p-values can be found in the literature (57, 58).

MHT methods designed to control for the k-FWER, k-FDR or FDX will not be considered further in relation to the CTT AE project (and hence are not discussed further in this document) because:

- For the purpose of this project, a large proportion of the null hypotheses (eg, >90% or >95%) tested are expected to be true due to the exploratory nature of the analyses. Thus, the use of FWER and FDR errors is preferable as the use of alternative errors will not provide any gain in the control of statistical errors or power.
- The literature has offered very little evidence or advice about the performance of such error rates in answering clinical questions (by contrast with the FDR and FWER, which tend to be preferred in practice (59)). The two main reasons such errors are not preferred in practical problems of MHT are:
    1. Methods designed to control those errors require additional decisions about the selection of appropriate values for the additional statistical parameters, such as *k* or *c*. These decisions are often unnecessary and they depend completely on the practical MHT problem, which means they cannot be generalizable.
    2. Methods which control for such errors are considered data-driven since the use of nonparametric tools, such as resampling methods, is applied due to the lack of mathematical relationships for the evaluation of the unknown parameters. The use of such nonparametric tools is considered to be an unnecessarily complicated and time consuming approach.

2.3.3 Correlation of hypotheses

As described in chapter 1.5.3, the correlation structure of multiple hypotheses is heavily discussed in the literature and various MHT methods have been developed to account for

different types of correlations. This section summarises the various types of correlation structures described in the literature and identifies those considered as potentially relevant to the CTT AE project.

2.3.3.1 Assumption free

These methods control for MHT of a family of hypotheses *without any assumptions being made about their dependency structure*. As a result, they are often a preferred choice. Two examples are the methods of Holm and Benjamini & Hochberg (described in detail in section 2.5). Since assumption free methods do not require any assumption about the correlation among the (family of) tests, they can be applied for use in any MHT problem, and hence are potentially suitable for use in the CTT AE project.

2.3.3.2 Independent

This approach describes a scenario whereby the hypotheses tested are *all assumed to be independent of each other* (ie, for every hypothesis in the family, the statistic used to test that hypothesis is uncorrelated with the statistics used to test every other hypothesis in the family). In the context of clinical trials, independent hypotheses would arise if there is no correlation between the treatment effects on multiple outcomes. For example, the effect of statins in reducing the risk of myocardial infarction may reasonably be thought to be unrelated to the effect (if any) on depression or myalgia (unless LDL cholesterol was thought to be a major causal risk factor for depression or myalgia, which it is not). Several examples of such independent statistics can be found in the literature (41, 42).

For the AEs included in the CTT database where no prior effect of statin therapy has been demonstrated or hypothesised, it is reasonable to expect no (or very low) correlation between the effects of statin on the various AEs examined. Therefore, the use of MHT and related methods designed for independent hypotheses are considered as potentially suitable for use in the CTT AE project.

2.3.3.3 Correlated

Various MHT methods exist in the literature specifically designed to account for correlation between multiple hypothesis tests. These MHT methods can be classified into two broad groups based on the statistical techniques they employ, as described below.

2.3.3.3.1 Multivariate analysis

Not to be confused with multi-*variable* analyses, MHT methods that are multivariate are designed to account for hypotheses that are highly correlated by allowing for multivariate distributions (ie, a statistical approach which accounts for the joint distribution among *outcomes*). Such MHT methods are used in practice, for example, in explanatory analyses and when simultaneously testing the effectiveness of various active therapies, or doses of a therapy, against placebo. In this case, all active therapies are expected to have a non-negative effect when compared with placebo. Several examples of MHT methods for correlated hypotheses using multivariate analysis can be found in the literature (60-62), with the Positive Regression Dependency Structure (PRDS) being the most widely known (63-65). Such MHT methods are often developed to control for MHT error rates by applying resampling techniques (eg, bootstrap or permutation) (66).

For the purpose of the CTT AE project, established 'uniformly-most-powerful'[e] methods for survival analysis (ie, log-rank test statistic) and meta-analysis (ie, inverse-variance methods) (see chapter 3.6) will be employed. Such methods do not require the use of multivariate distributions and, consequently, MHT methods designed to allow for multivariate analysis are not required for use in the CTT AE project, and so are not considered further.

2.3.3.3.2 Correlation by construction

Many MHT methods are designed to organise the multiple hypotheses into specific structures which allows hypotheses to be correlated by the construction designed. Several examples

[e] A uniformly-most-powerful method produces the statistical test (of a hypothesis) which has the most statistical power from the set of all possible tests of the same size α.

exist in the literature where hypotheses are organised into a hierarchical setting (ie, grouping the hypotheses into different levels and testing them based on their level). Such a hierarchical structure is often used to increase the statistical power of the MHT methods by reducing the number of hypotheses tested based on their hierarchical level.

For example, consider an RCT analysing the effect of a drug on a primary outcome (eg, major vascular events) which is itself a composite of multiple other outcomes of interest (eg, major coronary event, stroke, or coronary revascularisation), and suppose one wishes to control the FWER. In such a scenario, a sensible strategy might be to analyse the composite outcome first (as the initial testing level) and reject the null hypothesis (of no effect) if the associated p-value is less than 0.05 (in the direction of a treatment benefit). If the null hypothesis is not rejected (ie, the drug is not found to be effective) then one would not perform any further statistical tests. However, if it is rejected, then one would continue to test the individual components of the primary outcome, testing each against a smaller significance level (eg, an α of 0.0167 if there were three components, since 0.05/3 = 0.0167). Another example where a hierarchical structure (of hypotheses) might be employed is when analysing AEs which can be classified based on their type, anatomical or physiological system, aetiology (disease origin) or purpose. In such a scenario, a sensible MHT strategy might be to test an initial set (ie, level) of null hypotheses (ie, corresponding to the different type of AEs) and then consider further analyses (ie, testing each individual AE) *only* if the associated null hypothesis is rejected after MHT (eg FDR) adjustment. MHT methods developed to account for such inherent correlation structures are clearly relevant to the CTT AE project, and so are considered in subsequent chapters.

2.3.4 Ordering of multiple hypotheses

The majority of MHT methods are designed in such a way that they require hypotheses to be tested in a particular *order* (within a family), before controlling for the selected MHT error rate of choice, as this achieves greater statistical power when compared with MHT methods that

do not require hypothesis tests to be ordered. In general, three approaches may be considered when thinking about the ordering in which hypotheses may be tested.

2.3.4.1 No ordering required

This group includes MHT methods which do not require any ordering (within the family of hypotheses to be tested) before adjustment for MHT is done, with the simplest and most familiar such method being the Bonferroni approach (see section 2.5.5.1). Although such MHT methods are not generally preferred in practice because they provide lower statistical power than other methods, they were still considered as potentially suitable for use in the CTT AE project in the chapters that follow (not least to provide a benchmark comparison for other, superior, methods).

2.3.4.2 Data-driven ordering

MHT methods designed for data-driven ordering first require the (family of) hypotheses to be ordered based on their uncorrected p-values (ie, before controlling for MHT). Such methods are often preferred in practice because they achieve greater statistical power (compared with MHT methods that do not require any ordering) without having to pre-specify the order in which hypotheses will be tested (such methods are described below in section 2.4.4.3). MHT methods with data-driven ordering were therefore considered as potentially suitable for use in the CTT AE project.

2.3.4.3 Pre-specified ordering

This group contains MHT methods which are based on the idea that not all hypotheses tested within a family are of the same clinical 'importance'. In such a scenario, the order of hypothesis testing is *pre-specified* before any unblinding of results. This has the advantage that it can maximise statistical power if that pre-specification turns out to have been a sensible one, but has the disadvantage that potentially real effects on outcomes may fail to be detected if the pre-specification was misjudged. Nevertheless, such methods are often employed in trials, though normally when considering a fairly limited number of hypotheses. However, since the

aim of this meta-analysis is to examine a very *large* number of AEs with an unknown association with statin therapy, it was not considered an appropriate strategy for the current CT AE project, and hence such approaches are not considered further in this thesis.

2.3.4.4 Unequal weighting of hypotheses

There are particular MHT problems where not all hypotheses tested within a family are of the same importance but they still need to be tested by a MHT method. For example, consider a trial designed to have two co-primary analyses, one performed in the general population of patients and another in a pre-defined target subgroup (with patients in the target subgroup expected to experience greater treatment benefit compared with the general population). In such a scenario, both null hypotheses H_{01} (ie, no effect of treatment on general population) and H_{02} (ie, no effect of treatment in the target subgroup) need to be tested and, perhaps for regulatory filing purposes, the FWER may need to be controlled at 5%. In such a scenario, MHT methods have been developed that allow the two hypotheses to be weighted differently (based on their perceived clinical and public health importance) while still allowing each to be jointly tested. For instance, one might decide to split the overall α level *unequally* and 'spend' most of it (eg, α_1=0.04) for the most 'important' hypothesis, while still allowing the other to be tested at an alpha level of 0.01. In this way, the overall control of the FWER would be maintained at 5%.

Of course, as for the previously-described method that requires pre-specification of the *order* of tests, pre-specification of the alpha-level used for each test (in the context of a study that involves many hundreds if not thousands of tests) would be ill-judged. Consequently, methods developed for unequal weighting of hypothesis tests are not considered further in this thesis.

2.3.4.5 MHT methods with pre-specified ordering *and* unequal weights

There has been a continuous demand for developing MHT methods, which can be specifically applied to solving complex practical problems of MHT (for example, the design of complex

clinical trials), often requiring the pre-specification of *both* the ordering of hypotheses *and* the weights (alpha-levels) used in each test. Numerous such methods, often visualised using graphs (67-69), have been reported in the literature as chain (66, 70), fallback (71, 72) or gatekeeping methods (73-76). However, for the reasons outlined above in the previous two sections, such methods were not considered appropriate for the CTT AE project and are not considered further.

2.3.5 Adaptive approaches

When a procedure estimates the total number of true null hypotheses m_0 (assuming that not all hypotheses tested are true, ie, $m_0<m$) (see **Box 2.1**) and incorporates this additional parameter (m_0 is estimated based on the data provided) into the multiple hypothesis testing process, this is known as an *'adaptive'* approach. Adaptive MHT methods are more powerful than non-adaptive methods in situations when we expect a high number of false null hypotheses, but they cannot guarantee control of the error rate for which they are designed, due to the additional uncertainty generated by the estimation of the unknown parameter m_0. In particular, underestimation of m_0 can result in an underestimation of the level of error control and lead to poor performance of the procedure (38, 41). Two popular adaptive MHT methods in solving practical MHT problems have been developed by Hommel (77) for controlling the FWER and Benjamini & Hochberg (41) for controlling the FDR. (Note that the two adaptive methods mentioned above are distinct from the Hommel and Benjamini & Hochberg methods considered later in this thesis as being suitable to use in CTT.)

In the CTT AE project, the use of adaptive methods would (probably) generate more uncertainty around the control of the selected error rate rather than increasing the power of detecting true results. This is because a low proportion of the AEs to be assessed in this project are expected to be influenced by statins (because statin therapy has been so widely studied in the past, the expectation is that the vast majority of clinically important effects are already known). Consequently, adaptive approaches are not considered further in this thesis.

2.3.6 Bayesian approaches to MHT

Hypotheses testing and MHT are traditionally thought of as frequentist concepts because they seek to control frequentist statistical errors (ie, type I and II for single hypotheses and FWER or FDR for MHT). Efron (78) developed a method using empirical Bayes inferences to approach the frequentist FDR error from a Bayesian perspective. Further work has been conducted by Storey (79) to use Bayesian probability theory (and q-values[f] instead of p-values) in order to approximate the frequentist FDR error through use of posterior distributions. Numerous modifications of these methods have been developed and applied to address the problem of MHT in various genetic studies (80-83).

2.3.6.1 Analysis of AE data in clinical trials

Various Bayesian methods were identified in the literature using Bayesian models for detecting safety signals in clinical trials (84-87). Such methods use advanced statistical techniques (often using hierarchical models with 3 levels) designed to flag adverse effects by using various cut-offs of the posterior probability (ie, P[RR<1]>*k*, where *k* equals 0.8, 0.9, 0.95, etc.) in order to achieve the control of the frequentist FDR error (eg, FDR<5%). Often, simulation analysis is used to select the appropriate probability cut-off. To date, such Bayesian models have been developed to identify only *adverse* effects of active drugs (ie, similar to MHT methods for 'one-sided' tests) but not to identify both beneficial and harmful effects. Although various models have been developed to analyse different type of data (ie, using Binomial distributions for event counts and Poisson distributions for time to event data) (84, 86), no Bayesian model has been developed to accounts for MHT in a meta-analysis of AEs.

[f] The q-value is a Bayesian analogue of the p-value that incorporates multiple testing correction.

The development of a Bayesian model is achievable (by incorporating the 'meta-analytical' effect and two-sided probability cut-offs) for the analyses of the CTT AE data, but such an approach will not be further considered in this thesis for the following reasons[g]:

1. *Interpretation of results*: Frequentist methods have dominated the statistical analysis (including analysis of MHT), especially in the area of drug efficacy and safety, and scientists and regulatory authorities are experienced in drawing conclusions and making policy based on such methods. Since statins are one of the most highly-used drugs worldwide, it is important that the meta-analysis of the CTT AE data produces clear answers which are simple and easy to understand using statistical methods that are widely accepted. Since the use of Bayesian methods (including meta-analysis and MHT) have scarcely been applied and validated in large-scale studies, the use of such methods might lead to confusion and additional misinterpretation of the results.
2. *Prior distribution(s)*: Bayesian methods are designed to combine the likelihood of data together with some prior beliefs (ie, distribution) about the hypotheses in order to draw inference (ie, posterior distributions). Such methods can be more powerful than frequentist methods mainly when the prior information is informative (and agreed upon). Since the aim of this meta-analysis is to analyse AEs with no prior knowledge of association with statin therapy, the use of Bayesian methods is not expected to be more powerful than the frequentist method. Also, the use of a Bayesian approach might lead to additional unnecessary scepticism about the subjectivity of the priors.

[g] In addition, a Bayesian approach would involve some additional technical challenges. In particular, analysing the effects of statin therapy on AEs using Bayesian methods would require further development of the existing models to both account for the meta-analytical effect and to identify both beneficial and harmful effects of statins (by using two-sided probability cut-offs). The development of a new model would require additional assumptions (ie, about prior distributions) and decisions (ie, selection of appropriate probability cut-offs to control for MHT) which could increase the likelihood of generating estimation errors (ie, unreliable posterior distributions).

2.3.7 Summary of characteristics for MHT methods suitable to use in CTT

The MHT methods considered potentially-suitable for use in the CTT project, and to be assessed in more detail later in the thesis, were chosen with the following criteria in mind:

I. Allow assessment of two-tailed null hypotheses (to ensure previously-unknown beneficial *or* harmful effects of statins may be identified).
II. Control either the FWER or the FDR.
III. Not specifically designed for multivariate statistics.
IV. Consider all hypotheses tested as being of equal importance.
V. Not requiring pre-specification of the expected number of true null hypotheses.
VI. Specifically developed for frequentist inference.

2.4 Systematic search of MHT methods

2.4.1 Search strategy

In order to comprehensively assess all currently available MHT methods, a systematic search was performed. Full details of the search strategy, including search terms used, are provided in **Table 2.1**. The databases searched were MEDLINE, EMBASE and SCOPUS with an unrestricted start date and an end date of 7th December 2017. The search was limited to English language full-text articles which mentioned in their title either multiple hypothesis testing methods or one of the following: a) one of the two most scientifically appropriate error rates when considering multiple hypothesis testing (the FWER and the FDR); b) the generic False Discovery Proportion (FDP); and c) the False Discovery Exceedance (FDX). Articles which included in their title ‘multiple myeloma’, ‘multiple sclerosis’ or ‘multiple choice’ were excluded because the search strategy included the word ‘multipl$’. EndNote V7 was used to extract and review all the references returned by the search strategy.

2.4.2 Search results

The search returned 9,040 titles and abstracts for review (see **Figure 2.1**). Of these, 2,471 were duplicates, 5,627 were excluded on the basis of their title alone (with the main reasons for exclusion being the inclusion of words 'multipl$' and 'test' in another context such as multiple regression, multiple sites, laboratory test samples, antibodies or PCR test, etc.), and a further 613 were excluded based on their abstract (with most of those articles describing the *application* of an existing MHT method). The remaining 329 articles were included for full-text review, together with 8 additional articles found upon close review of their associated references. Of these 337 articles which were reviewed in detail, 60 were excluded as an application of MHT methods and 222 articles were excluded based on specific characteristics of the MHT methods for one of the following reasons (explained in detail in section 2.3):

- 78 articles reported methods (or review of methods) developed to control MHT error rates other than the FWER or the FDR; and
- 144 articles reported methods (or review of methods) specifically designed for: i) one-tailed null hypotheses, ii) multivariate analyses, iii) pre-specified ordering of hypotheses, iv) adaptive approaches, or v) Bayesian approaches.

Of the remaining 55 articles, 9 described MHT methods and 46 were reviews of those (and other) MHT methods under different data scenarios (synthetic or real). **Table 2.2** presents the 9 eligible MHT methods with a detailed description and review of those methods presented below in section 2.5.

2.5 Review of all suitable MHT Methods

This section presents a detailed review of all multiple hypothesis testing methods that were found suitable for use in the CTT AE project (as described above in section 2.4.7), organised by the error rate they control (FWER or FDR). The literature review identified 9 such methods,

5 of which control the FWER and 4 of which control the FDR (**Table 2.2**). These methods are described in detail below.

2.5.1 Methods controlling the Familywise Error Rate

2.5.1.1 Bonferroni's method

The Bonferroni method (88) controls the FWER by simply dividing the α level of significance by the total number of hypotheses tested within the family. For example, for a 'family' with *m* hypotheses, each individual hypothesis is rejected (ie, declared 'significant') if and only if its p-value is less than $\boldsymbol{\alpha/m}$. Bonferroni is the simplest and easiest method applied for handling MHT problems. However, it is very conservative because although it controls the FWER, it does so in a way that unnecessarily reduces the statistical power to detect true positives (ie, effects of a drug).

2.5.1.2 Holm's 'step-down' method

For Holm's method (89), all hypotheses H_i are first organised based on the values of the uncorrected p-values in a step-down order going from the smallest to the largest p-value ($p_1 \leq p_2 \leq \ldots \leq p_m$). Starting with the smallest p-value (p_1), one compares p_j with $\boldsymbol{\alpha/(m-j+1)}$. If p_j is less than this value, the hypothesis is rejected and the next smallest p_j is examined. One continues until the first non-rejection, after which no remaining hypotheses are rejected. For example, for a family of three hypotheses with $p_1 \leq p_2 \leq p_3$ and a FWER of 0.05, Holm's method starts with (the hypothesis related to) p_1 which is compared against the value 0.0167 (ie, 0.05/3). If $p_1 \leq 0.0167$, then the corresponding hypothesis is rejected and the next hypothesis (ie, related to p_2) is examined against value 0.025. Otherwise, all subsequent hypotheses are accepted[h] without further test.

Holm's method has been proven to be uniformly more powerful than Bonferroni because it never rejects fewer hypotheses than Bonferroni (90). This means that Holm's method is

[h] Note that the null hypothesis is never truly 'accepted', but this phraseology helps describe the process more easily than the more accurate term 'not rejected'.

always able to detect more true positives (eg, true effects of a drug) while maintaining the same number of false positives as Bonferroni. Holm's is also the only method (other than Bonferroni) that controls the FWER without requiring additional assumptions about the correlation structure of the hypotheses (see section 2.4.3). It is included as an example of an *assumption free* method (91, 92) allowing for any distributional assumptions made about the test statistics.

2.5.1.3 Hochberg's 'step-up' method

Like Holm, Hochberg's method requires that all hypotheses are first ordered according to their p-values $p_1 \leq p_2 \leq \ldots \leq p_m$ (93). By contrast with Holm's method however, Hochberg is a 'step-up' method in the sense that it starts with the largest of the uncorrected p-values (p_m) rather than the smallest. If $p_m \leq \alpha$, then all hypotheses are rejected. If not, p_m is not rejected, and the next step is to move to p_{m-1} and check this against $\alpha/2$. This method involves moving from higher to lower p-values, rejecting the null hypothesis if $\boldsymbol{p_i \leq \alpha/(m-i+1)}$ (after which all remaining hypotheses are rejected).

For example, for a family of three hypotheses with $p_1 \leq p_2 \leq p_3$ and a FWER of 0.05, Hochberg starts with (the hypothesis related to) p_3 which is compared against the value 0.05. If $p_3 \leq 0.05$, then all three hypotheses are rejected without further testing. If not, the next hypothesis (ie, related to p_2) is examined against the value 0.025 and in case it is not rejected (ie, $p_2 > 0.025$) the last hypothesis (ie, related to p_1) is examined against the value 0.0167.

Although Hochberg is designed to control the FWER only when the statistical tests are independent (see section 2.4.3.2) (94, 95), various simulation studies have shown that this method can control the FWER for particular examples of non-independent hypotheses (96).

Simulation studies have also shown that Hochberg's method has a power advantage over Holm's method when testing a low number (eg, <10) of hypotheses (97), but this advantage is usually very small (98).

2.5.1.4 Rom's 'step-up' method

As for Holm's and Hochberg's methods, Rom's method (99) also requires that all hypotheses are first ordered according to their p-values $p_1 \leq p_2 \leq \ldots \leq p_m$. As a 'step-up' method (similar to Hochberg), it starts with the largest of the uncorrected p-values (p_m). If $p_m \leq \alpha$, then all hypotheses are rejected. If not, p_m is not rejected, and the next step is to move to p_{m-1} and check this against $3\alpha/4$. This method involves moving from higher to lower p-values, rejecting the null hypothesis as soon as $\boldsymbol{p_{m+1-i} \leq \alpha(i+1)/2i}$, after which all remaining hypotheses are rejected.

For example, for a family of three hypotheses with $p_1 \leq p_2 \leq p_3$ and a FWER of 0.05, Rom's method starts with (the hypothesis related to) p_3 which is compared with value 0.05. If $p_3 \leq 0.05$, then all hypotheses are rejected without further testing required. Otherwise, the next hypothesis (ie, related to p_2) is examined against value 0.0375 and if not rejected (ie, $p_2 > 0.0375$) the last hypothesis (ie, related to p_1) is examined against value 0.0333.

Some simulation studies have shown that Rom's method offers a small power advantage over Hochberg's, but sometimes it does so by not fully controlling the FWER error rate, especially when testing correlated hypotheses (eg, when testing the effectiveness of a treatment on various outcomes which are correlated due to similar underlying pathology) (see section 2.3.3.3) (99).

2.5.1.5 Hommel's 'two-step' method

Unlike the four (FWER) methods described above which can be classified as 'single-step' methods (since they are designed to reject the null hypotheses based on a single mathematical formula), Hommel's method can be considered as a two-step approach since it requires two mathematical conditions to be satisfied before hypotheses are rejected (100). With Hommel's method, all hypotheses are initially ordered according to their p-values $p_1 \leq p_2 \leq \ldots \leq p_m$. Then, as a first step, Hommel starts with the largest of the uncorrected p-values (p_m) and it identifies the largest integer j for which $\boldsymbol{p_{m-j+k} > k\alpha/j}$ (with k ranging from *1* to j). If no

such *j* exists then Hommel rejects all null hypotheses. Otherwise, it moves to the second step and rejects all null hypotheses with $\boldsymbol{p_i \leq \alpha/j}$. Similar to Hochberg, Hommel is designed to control the FWER for independent statistical tests (see section 2.3.3.2).

Simulation studies have shown that Hommel's method is typically about as powerful as Hochberg's (101, 102), and only sometimes slightly more powerful, with this power improvement generally being considered negligible given the test's complexities.

2.5.1.6 FWER MHT methods in practice

When reporting research results, Holm's procedure is often recommended as the first choice, due to the scepticism that other methods cannot guarantee 'assumption free' control of FWER (90, 92, 95, 103), but Hochberg's method is also widely used, especially in clinical trials. By contrast, Hommel's method tends to be used less often (perhaps given its added complexities with little benefit) while Rom's method does not seem to be used widely (probably because it does not always exert *perfect* control of the FWER). Finally, it is widely-recognised that Bonferroni's method should be avoided (as other methods are much more powerful whilst retaining control of the FWER).

2.5.2 Methods controlling the False Discovery Rate

2.5.2.1 Benjamini & Hochberg 'step-down' method

As for several of the FWER methods described above, Benjamini & Hochberg's method (38) first requires all hypotheses to be ordered based on their uncorrected p-values $p_1 \leq p_2 \leq \ldots \leq p_m$. As a 'step-down' method, it starts with the largest of the uncorrected p-values (p_m). If $p_m \leq \alpha$, then all hypotheses are rejected. If not, p_m is not rejected, and the next step is to move to p_{m-1} and check this against $2\alpha/m$. This method involves moving from higher to lower p-values, rejecting the null hypothesis as soon as $\boldsymbol{\alpha j/m}$, after which all remaining hypotheses are rejected.

For example, for a family of three hypotheses with $p_1 \leq p_2 \leq p_3$ and a FDR of 0.05, Benjamini & Hochberg starts with (the hypothesis related to) p_3 which is compared with value 0.05. If $p_3 \leq 0.05$, then all hypotheses are rejected without further testing required. Otherwise, the next hypothesis (ie, related to p_2) is examined against value 0.033 and in case it is not rejected (ie, $p_2 > 0.033$), the last hypothesis (ie, related to p_1) is examined against value 0.0167.

Benjamini & Hochberg can be considered as an 'assumption free' method since it can be used for both independent and dependent hypotheses, without requiring additional distributional assumptions about the test statistics or p-values. Although, on the face of it, this method looks very similar to Hochberg, it is designed to control the FDR error rate (by applying a different threshold criteria).

2.5.2.2 Benjamini & Liu's 'step-down' method

Benjamini & Liu (104) designed a method which controls the FDR under the assumption of independent statistics. This method also requires hypotheses to be ordered based on their uncorrected p-values: $p_1 \leq p_2 \leq \ldots \leq p_m$. Starting with the smallest p-value (p_1), one compares p_j with $\mathbf{1} - [\mathbf{1} - \mathbf{min}\left(\mathbf{1}, \frac{\mathbf{m\alpha}}{\mathbf{m-i-1}}\right)]^{\mathbf{1/(m-i+1)}}$. If p_j is less than this value, the hypothesis is rejected and the next smallest p_j is examined. One continues until the first non-rejection, after which no remaining hypotheses are rejected.

For example, for a family of three hypotheses with $p_1 \leq p_2 \leq p_3$ and a FDR of 0.05, Benjamini & Liu's method starts with (the hypothesis related to) p_1 which is compared with value 0.01695 (ie, $1 - [1 - \min\left(1, \frac{3x0.05}{3}\right)]^{1/3}$). If $p_1 \leq 0.01695$, then the corresponding hypothesis is rejected and the next hypothesis (ie, related to p_2) is examined against value 0.03823 (ie, $1 - [1 - \min\left(1, \frac{3x0.05}{2}\right)]^{1/2}$)). Otherwise, all subsequent hypotheses are accepted without further tests.

Simulation studies have shown that Benjamini & Liu's method is slightly more powerful than Benjamini & Hochberg when there are a small number of hypotheses (ie, less than 8), but it has less power when a larger number of hypotheses is tested (104).

2.5.2.3 Mehrotra & Heyse method

Mehrotra & Heyse (105) developed a method which allows hypotheses to be organised into two hierarchical levels (eg, allowing for grouping adverse events into body organ systems) (see section 2.3.3.3). The steps of the algorithm for this method are described below:

1. Let p_{ij} be the p-value of the hypothesis (ie, adverse event) j, j=1,2,…,k_i, which belong to the group (ie body organ system) i, i=1,2,…,s.
2. For each group i identify the most significant results. This is the minimum p-value p_{i*} which is used as the representative p-value for group i.
3. Start with group p-values p_{i*} i=*1,2,…,s* and use the Benjamini & Hochberg's method to detect significant p-values at FDR level equal to α_1.
4. Then, for *each* group where the smallest p-value is declared significant, examine separately the individual p-values p_{ij} by re-using the Benjamini & Hochberg's method to detect significant p-values at the same FDR level (for all groups) equal to α_2.

With this method, a resampling statistical technique (nonparametric bootstrap) is used to specify the appropriate values of α_1 and α_2. In their paper, Mehrotra & Heyse suggest pre-specifying that $\alpha_2=\alpha$ (ie, 0.05) and then employing the bootstrap resampling technique to determine the largest data-dependent α_1 ($\leq\alpha_2$) that ensures the overall FDR$\leq\alpha$. Alternatively, they suggest using bootstrap techniques to study possible fixed choices of α_1 and α_2, for example, $\alpha_1=\alpha_2/2$. Mehrotra & Heyse's method is also known as a Double FDR approach, since it controls the FDR error rate at two (hierarchical) levels.

2.5.2.4 Mehrotra & Adewale's method

Mehrotra & Adewale (106) is based on the Mehrotra & Heyse's method with the difference that this new method applies just one double-FDR adjustment to the family of all p-values

collected from all the groups detected as significant in step 1 (and not a separate FDR adjustments to each group flagged as significant in step 1 as done in Mehrotra & Heyse method). In addition, the Mehrotra & Adewale method does not require the bootstrap resampling technique used in Mehrotra & Heyse method and it makes the plausible assumption that $\alpha_1=\alpha_2=\alpha$ (ie, α=0.05). Simulation studies have shown that the Mehrotra & Adewale provides better control of the FDR compared with the Mehrotra & Heyse, which is sometimes problematic from an FDR control perspective (106). In addition, Mehrotra & Adewale's method has been shown to be slightly more powerful than Benjamini & Hochberg's method for different simulation scenarios (for flagging adverse experiences) while maintaining similar control of the FDR (106).

2.5.2.5 FDR MHT methods in practice

Benjamini & Hochberg, the most well-known MHT method, is widely used in MHT medical research studies. Since most of the MHT methods designed to control the FDR are used to test a large number of hypotheses (eg, >50), Benjamini & Liu is not preferred in practice as it does not provide any gain in power compared with Benjamini & Hochberg. By contrast, the methods by Mehrotra & Heyse and Mehrotra & Adewale are specifically designed to analyse safety data, and have consequently been used in several previous MHT applications.

2.6 Summary of findings

This chapter has summarised the approaches, strengths and limitations of all MHT methods that exist in the literature and could, *theoretically,* be used in the CTT AE project. From a total of 6,569 articles identified by the systematic review (plus 8 articles found by hand searching of references), 337 were included for full text review. From these articles, 9 MHT methods were identified as potentially suitable to use in the CTT AE project with 5 methods designed to control the FWER, and 4 the FDR error rate.

From the 5 FWER methods, Bonferroni is the least powerful method while Rom's method may not fully control the FWER (especially when testing independent hypotheses). For the remaining 3 FWER methods (Holm, Hochberg and Hommel), it is not obvious from the literature which would perform best when applied to the CTT AE data (and hence all 3 are considered later).

From the 4 FDR methods, Benjamini & Liu's method is less powerful than Benjamini & Hochberg's when testing (a family with) more than eight hypotheses. The method by Mehrotra & Heyse has been reported to be problematic (from an FDR control perspective) whereas the method by Mehrotra & Adewale is reported to resolve this by performing an FDR correction to a single family containing all p-values reported in all the groups (ie, body system organs) flagged as significant in step 1. Simulation studies (found in the literature) indicate that Mehrotra & Adewale's method results in similar control of the FDR compared with Benjamini & Hochberg's, and may be slightly more powerful than Benjamini & Hochberg's method in particular scenarios. However, it is unclear which of these two FDR methods (Mehrotra & Adewale or Benjamini & Hochberg) would perform better when applied to the CTT AE data (and hence both are considered later).

Box 2.1: Potential inferences when simultaneously testing *m* hypotheses

	Decision of the MHT procedure		**Total**
	Reject the null hypothesis	**Fail to reject the null hypothesis**	
Null hypothesis is true	**V** *'false positive'*	**U** *correct inference drawn*	m_0
Null hypothesis is false	**S** *correct inference drawn*	**T** *'false negative'*	m-m_0
	R	m-**R**	m

Table 2.1: Search strategy for MHT methods

Database	Search strategy*
Medline and Embase *(via OVID)*	Only articles titles including: 1. (method$ or procedur$ or test$) 2. multipl$ 3. (test$ or hypothes$ or comparison$ or correction$ or endpoint$ or outcome$) 4. (false discover$ rate$) or FDR 5. (familywise error$ rate$) or FWE$ 6. (false discover$ proportion$) or FDP 7. (false discover$ exceedance) or FDX 8. (scleros$ or myelom$ or choice) 9. ((1 and 2 and 3) or 4 or 5 or 6 or 7) and not 8
Scopus	*TITLE(((false discover* exceedance) OR (false discover* proportion*) OR (familiwise error rate) OR (false discovery rate) OR FDR OR FWE OR FDP OR FDE OR ((method* OR procedur* OR test*) AND (multipl*) AND (test* OR hypothes* OR comparison* OR correction* OR endpoint* OR outcome*))) AND NOT (scleros* or myelom* or choic*))*

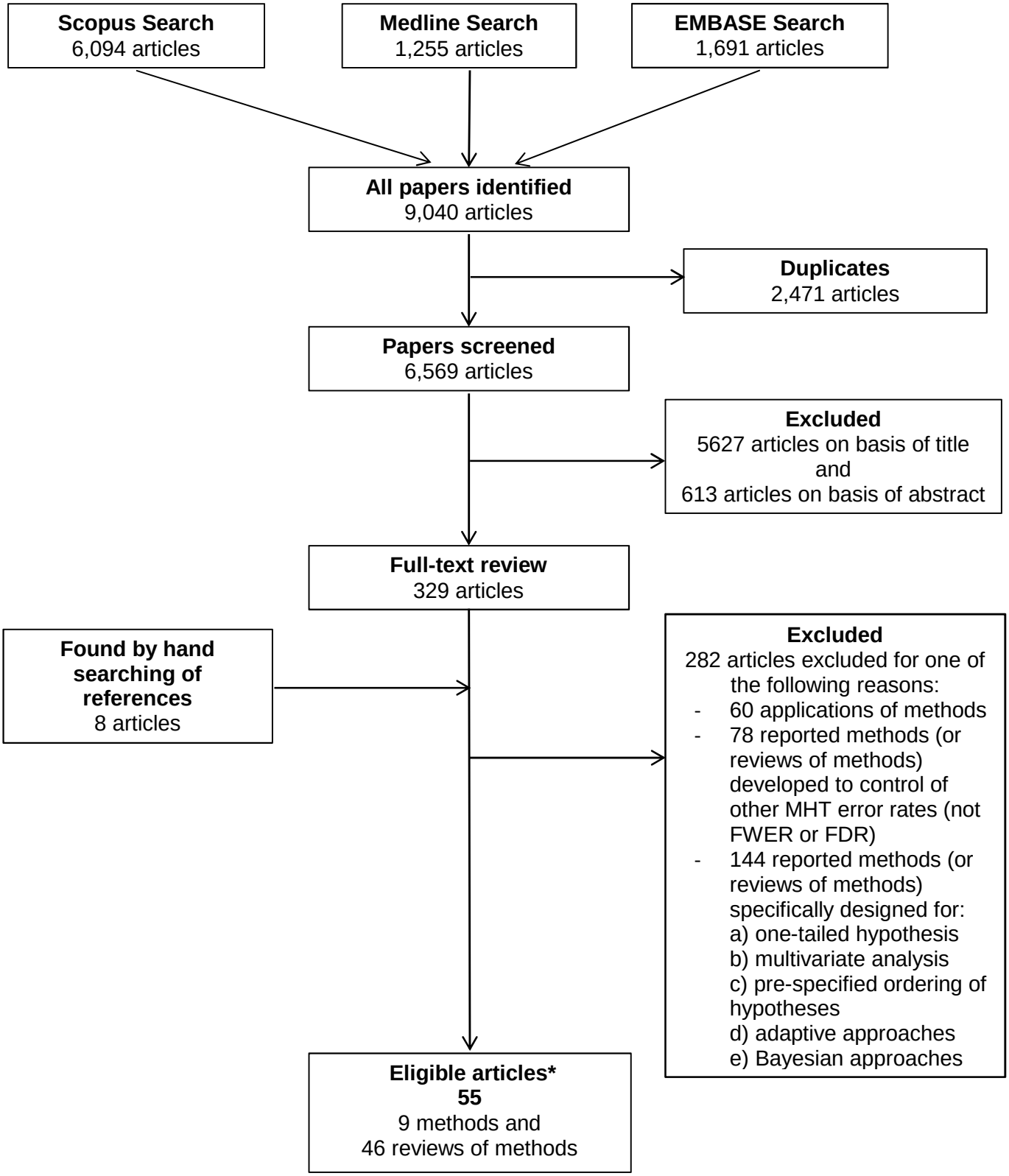
*Search strategy was limited to English articles; start date: unrestricted; end date: 07/12/2017

Table 2.2: Characteristics of multiple testing methods suitable to use in CTT

Multiple testing method	Error control	Hypotheses ordering	Hypotheses dependency
Bonferroni	FWER	Not required	Assumption free
Holm	FWER	Step-down	Assumption free
Hochberg	FWER	Step-up	Independent
Rom	FWER	Step-up	Independent
Hommel	FWER	Two-step	Independent
Benjamini & Hochberg	FDR	Step-down	Assumption free
Benjamini & Liu	FDR	Step-down	Independent
Mehrotra & Heyse	FDR	Two-step	Correlated by construction
Mehrotra & Adewale	FDR	Two-step	Correlated by construction

FWER=Familywise error. FDR=False discovery rate. Step-down methods order the hypotheses based on p-values from the smallest to the largest. Step-up methods order the hypotheses based on p-values from the largest to the smallest. Two-step methods require the first step to perform a global or group testing of hypotheses and then, if rejected, move onto the second step to sort the individual hypotheses in a step-down or step-up order and test them. Assumption free methods do not require any assumptions being made about their dependency structure. Independent means that hypotheses tested are all assumed to be independent of each other. Correlated by construction methods organise the multiple hypotheses into specific structures which allows hypotheses to be correlated by the construction designed.

Figure 2.1: Review process of all eligible articles for multiple testing methods



*Articles which report methods or (reviews of methods) which were considered as suitable to use in the CTT AE project.

3 CTT data processing and statistical methods

3.1 Summary

The aim of this chapter is to describe the data processing and statistical methods used in the CTT analyses of the effects of statins on all AEs with a previously unknown effect. This includes describing: the data harmonisation methodology developed for taking the heterogeneous data received from the trialists and transforming it into a single, common-format analysable database, in particular, the methods established for standardizing the AEs received (from the multiple trials) into a common medical dictionary; results from examining the correlation structure between the AEs, and the impact of this structure on the statistical analyses performed; the population of interest when performing both the subsequent simulation analyses and the actual meta-analyses; the design of the simulation studies developed to evaluate the expected performance of the different MHT testing options is then provided; and the statistical methods used to ultimately perform the meta-analysis. The methods described for the harmonisation of IPD and standardization of AEs have recently been published in Clinical Trials (107).

3.2 Data harmonisation methodology

The large volume and heterogeneity of data collected in the 28 CTT-eligible statin trials (which were done across multiple countries and over several decades, and involved varying levels of collaboration between academic institutions and the statin manufacturers) raised considerable challenges for the collation, organization and analyses of the AE data they captured. For the current systematic analyses of all AE data, therefore, an approach was needed to harmonise the IPD and standardise the outcomes into a single analysable database capable of addressing the key research aims. The CTT Secretariat in Oxford (which included the author of this thesis) therefore developed an integrated data harmonisation framework involving

extensive correspondence with collaborating trialists regarding data queries and coding assumptions. This framework is summarised in **Figure 3.1** and involved four key steps prior to performing any analyses of the data. These four steps are described below.

3.2.1 Step 1: Review of trial IPD

The initial step was to collect and review the data provided by the trialists. 'Data' is a broad term in this context, including not only the datasets provided but also any other relevant documentation. For each trial, we requested the following: information related to how AEs were sought, including high-level documentation such as trial protocols; blank copies of the trial case report forms and information on any coding systems used; any available existing tabulations of adverse event data (eg, in published articles or elsewhere), and all individual-level datasets containing information on AEs (including the time from randomisation to each AE); the timing of and reasons for stopping study treatment; co-medication data; and data for relevant laboratory values, physical measurements (eg, body-mass index, blood pressure) or lifestyle characteristics (eg, smoking).

Table 3.1 summarises the volume of IPD harmonised into the CTT AE database, including >840 datasets, approximately 30,000 variables, about 1.2 million AEs (with >45,000 different types of unique AEs) and over 38 million records in total. The volume of the data for each trial was driven more by the structure of the questionnaires for each trial and how trialists had structured their own data rather than just the number of participants in the trial (with smaller trials sometimes generating far more data than larger ones). Most trials provided a combination of raw data (ie, exactly as collected) and derived data (ie, whereby data cleaning had been applied by the trialists, such as when pre-specified outcomes had been clinically adjudicated).

3.2.2 Step 2: Creation of 'Meta' data dictionary—use of global standards

A 'meta' data-dictionary was created which fully described the conversion of the data into a single analysable dataset, thereby increasing the transparency of how the data was handled.

Much of the strategy developed to convert these heterogeneous data sets was based on methods developed by the Clinical Data Interchange Standards Consortium (CDISC) (https://www.cdisc.org/) (108), an organisation which develops high quality global data standards to transform incompatible clinical research data formats, inconsistent methodologies and diverse perspectives into a common framework. Specifically, a system based upon the CDISC Study Data Tabulation Model (SDTM) for organizing and formatting data was developed. SDTM includes a total of 46 data 'domains' and a large number of potential study variables across these domains. Based on the requirements of the CTT AE programme of work (as defined in the study protocol) (35), a streamlined version of SDTM was devised, with 13 of the 46 available domains selected for use (**Box 3.1**). These 13 selected SDTM domains were: Adverse Events (AE), Clinical Events (CE), Co-Medication (CM), Death Details (DD), Demographics (DM), Exposure (EX), Healthcare Encounters (HO), Laboratory (LB), Medical History (MH), Procedures (PR), Substance Use (SU), Subject Visit (SV) and Vital Signs (VS). Within each of these selected domains, only a subset of the potentially available variables (ie, only those considered as relevant for the aims of this project) was then chosen. These variables were supplemented, where necessary, with additional 'CTT-specific' variables. For example, a variable which allowed identification of AEs resulting in active treatment being stopped was added. A detailed description of the (meta-) data dictionary is presented in **Appendix Table 3.1**.

Although the *general* approach to data organization and manipulation was based on CDISC standards, these standards were not followed rigidly. For example, in some cases it was deemed necessary to create only a subset of the SDTM domains related to AEs because the trial had already provided all of their AEs in a single dataset (and so there was no need to split this into multiple outcome domains only to then merge them together again at a later stage). By contrast, some trials reported their AEs in multiple datasets (eg, in one trial, there were 13 datasets [including a mixture of raw, derived, or adjudicated datasets] containing the totality of AEs collected from the various case reports forms). In this case, the use of multiple domains

for reporting the adverse events (ie, AE, CE, DD and PR) was considered a more appropriate choice because of the design of the case report forms (and the data generated by those forms).

3.2.3 Step 3: Standardization of AEs to a common medical dictionary

The aim of the current CTT AE project was to develop a strategy for, and then implement, individual participant data meta-analyses of the effects of statin therapy on all AEs. In this context, an 'AE' is defined as any post-randomisation medical occurrence reported for a participant, regardless of attribution to study treatment (ie, regardless of whether classed as related to any allocated study treatment or not).

Generally, only a small subset of AEs (often pre-defined as outcome measures in the study protocol) reported for each trial had been adjudicated by the original trialists, with the majority of other AEs being unadjudicated and captured using a variety of medical coding dictionaries, or as uncoded "free text". **Table 3.1** presents the total and unique number of AE terms reported for each trial and any associated medical dictionary used. Sixteen trials reported their AEs through various versions of the Medical Dictionary for Regulatory Activities (MedDRA) (eg, V11 or V13), while 9 trials used a different dictionary (eg, ICD-9 or ICD-10) for AE reporting. The development of a tool to standardise all AEs reported across multiple dictionaries into a single medical dictionary was therefore required.

3.2.3.1 Medical Dictionary for Regulatory Activities (MedDRA)

To categorise the adverse events captured by the trials a coding system based upon the Medical Dictionary for Regulatory Activities (MedDRA: https://www.meddra.org/) was utilised (109). MedDRA is an internationally recognised standardised medical terminology widely used for regulatory communication and evaluation of data pertaining to medical products. MedDRA employs the following five-level hierarchical structure for coding AEs (from very general to very specific) as described in **Box 3.2** and summarised below:

i. System Organ Classes (SOC) (eg, cardiac disorders)

ii. High Level Group Terms (HLGT) (eg, coronary artery disorders)

iii. High Level Terms (HLT) (eg, ischaemic coronary artery disorders)

iv. Preferred Terms (PT) (eg, myocardial infarction)

v. Lowest Level Terms (LLT) (eg, heart attack or anterior myocardial infarction)

Note that every PT has an identical LLT underneath it as well as similar medical terms or synonyms and lexical variants (eg, abbreviations or different word order). Further details can be found at: https://www.meddra.org/how-to-use/basics/hierarchy.

MedDRA is a very rich dictionary, including a total of >77,000 LLTs. **Table 3.2** presents the number of terms at each of the hierarchical levels of MedDRA, organised alphabetically by SOC. MedDRA Version 20.0 was used for the CTT project, since this was the version in place at the time of the project starting. The process for mapping the trial-reported AE terms to MedDRA categories is described below. Note that all such mapping was done blind to treatment allocation, using the AE datasets identified as being most relevant to the project (for example, derived AE datasets typically 'trumped' corresponding raw ones).

3.2.3.2 CTT: AE MedDRA mapping approach

CTT AE terms (including those given as reasons for stopping study treatment as determined from the SDTM EX domain) were coded to MedDRA version 20.0A program was devised to ensure that AE terms that directly mapped to an existing MedDRA v20.0 term would be included in analyses not only of that one category, but also of all relevant higher AE categories. Initially, a direct match was sought at the LLT level, but if no exact match was found, matches were attempted at higher MedDRA levels (initially at the HLT level, and if that failed then at the HLGT level). For terms that did not directly encode in this way, mapping was done manually by a research team overseen by clinicians. Initially, such terms were checked for obvious errors such as simple typographical errors or changes in word ordering that precluded direct coding. If such basic corrections still failed to yield a direct MedDRA match, a more

detailed manual mapping assessment was undertaken, following a coding guidance document developed for the project based upon the MedDRA coding guidance. Such detailed assessment was particularly required when events had been coded as 'free text'. To ensure consistency, all AEs were coded by at least 2 members of the research team, with any discrepancies resolved by discussion. In addition, periodic clinician sub-sampling of all other encoded AEs was also undertaken.

3.2.3.2.1 Distribution of CTT mapping

Table 3.3 presents the total number of AEs mapped directly across the hierarchical levels of MedDRA V20.0, as well as those requiring manual mapping. Of the 1.2M AE terms, about 780,000 (66%) directly auto-coded to a LLT (of which approximately 12,000 were unique terms). Less than 1% of AEs directly auto-coded only at the higher HLT or HLGT levels. For those AEs requiring manual mapping, again the majority of terms were coded at the LLT level. The proportion of AE terms undergoing auto-encoding versus manual mapping varied considerably between different trials depending on the original coding system used and the structure of the trial case report forms, with auto-encoding ranging from more than 90% of trial AEs to less than 5%.

3.2.4 Specification of outcomes for hypothesis-testing

Using the SDTM-based (meta-) data dictionary (see section 3.2.2) and the mapping tool based on MedDRA V20 (see section 3.2.3), the highly diverse and large CTT trial specific data items were successfully harmonised into a single analysable AE database. As described in chapter 1.6, the main aim of this project is to analyse any AE where no prior hypothesis exists for an association with statin therapy. A list of the AEs where either a known effect of statins has been established or a specific hypothesis was reported in the CTT 2016 protocol (35) is presented in **Appendix Table 3.2**; such AEs were not considered in the formal MHT analyses shown later (although several of them *are* shown in some analyses later as 'positive-control' tests to validate the coding approach used). The effect of statin therapy on AEs was *not* to be

done systematically across different types of patient because this would result in substantial additional hypothesis testing.

3.3 Correlation structure

A key challenge of this project is to account for the inherent correlation structure of the data (ie, relationships between the AEs examined) and the impact of this on the statistical analysis. The correlation structure of the CTT AE data can be thought of as separating into two types of correlation, as described below.

3.3.1 Correlated effect estimates by construction

We use the term 'correlation by construction' to describe the situation whereby the estimated effects of statin therapy on a set of outcomes are all correlated due to the fact that the outcomes overlap in their definitions (due to the hierarchical structure of MedDRA). For example, an analysis of the PT 'Colitis' would (to some extent) determine the results of an analysis of HLT 'Colitis (excl infective)' which in turn would determine the results of an analysis of the HLGT 'Gastrointestinal inflammatory conditions'. More generally, the results of hypothesis testing at one level (eg, PT) directly influences the results at higher levels (ie, HLT or HLGT).

3.3.2 Correlated effect estimates due to similar underlying pathology

Statins reduce LDL cholesterol and consequently the risk of atherosclerotic coronary heart disease (see chapter 1.2). Correlated effect estimates due to the underlying mechanism of the action of statins would therefore be expected for any AE for which atherosclerosis has a causal role. For example, the protective effect of statin therapy on risk of myocardial infarction would be expected to be correlated with the effect on hospitalization for unstable angina (due to the

similar underlying pathology of these two outcomes (9). Although some MHT methods do exist to deal with such situations (ie, the 'multivariate' methods described in chapter 2.3.4.3), the expectation was that, once the AEs known to be affected by statins were excluded (eg, coronary outcomes), any such correlation due to similar underlying pathology among the remaining AEs would be small (if present at all).

3.4 Populations of interest

This thesis aims to identify any previously undetected effects of statin therapy using AE data from randomised trials incorporated within the CTT collaboration. The most up-to-date CTT database includes data from 28 trials involving 181,971 participants, 23 trials involving 154,664 participants randomised to statin vs placebo or open control and 5 trials involving 27,307 participants randomised to a more intensive vs a less intensive statin regimen (**Table 3.1**).

Only double-blinded trials were considered for analyses because of the risk of reporting bias in open-label trials, sometimes referred to as the *nocebo* effect (whereby subjective AEs may be more likely to be attributed to a treatment simply because the patient knows they are taking it) (110). A particular example of the nocebo effect is provided by the ASCOT-LLA trial of atorvastatin 10 mg daily vs placebo. During the double-blind randomised main phase of the trial, there was no significant excess in reports of muscle-related AEs among those allocated atorvastatin. However, in a subsequent post-trial follow-up phase where participants knew whether they were taking a statin or not, reports of muscle-related AEs were more common among those taking a statin (111).

9 This type of correlation is distinct from correlated outcomes due to common risk factors (eg, age, blood pressure and smoking are all common risk factors for both the risk of myocardial infarction and the risk of haemorrhagic stroke), or from treatment effects which are positively correlated due to the nature of the hypotheses.

Of the 28 randomised trials, 5 were open-label and were excluded from the later analyses; the remaining 23 double-blinded trials include 19 double-blind trials of statin vs placebo involving 123,940 participants (the main analysis population used for hypothesis-testing) and 4 double-blind trials of a more intensive vs a less intensive statin regimen involving 30,724 participants (a secondary analysis population in which any apparent effects may be dose-related).

3.5 Simulation analyses

In chapter 2, a review of the literature on MHT methods was presented with the aim of identifying all MHT methods potentially-appropriate for use in the CTT AE project. In total, 5 MHT methods were selected: 3 controlling the FWER (Holm, Hochberg and Hommel) and 2 controlling the FDR (Benjamini & Hochberg and Mehrotra & Adewale). In order to explore the expected performance of these 5 different methods (with respect to controlling 'false positive' and 'false negative' results) when applied to the various types of AEs within the CTT database, a set of simulation studies was designed (the results of which are given in chapter 4). This section describes the design of and rationale for these simulation studies.

3.5.1 Selection of the family of hypotheses

As described in chapter 1, the selection of an appropriate family of hypotheses is important in MHT theory because it determines the threshold level of significance at which any individual hypothesis within the family is rejected or not. Unlike when reporting safety results in individual clinical trials, where it may be reasonable to perform hypothesis testing at only a single level of MedDRA (eg, the HLT level), in the current CTT AE project it was considered desirable to perform hypothesis testing at all MedDRA levels, for the following reasons:

- *Richness of the dataset:* When combining all of the trials, the CTT AE database is extremely rich in terms of the number of outcomes and the number of patients reporting

each outcome. Consequently, reliable hypothesis-testing even at the granular PT level is possible (and desirable).

- *Heterogeneity between the trials in how AEs were recorded.* The trials contributing to this meta-analysis reported their AEs across different levels of MedDRA. Performing analyses at only one MedDRA level would risk some trials effectively being excluded from the analysis of some outcome types.

- *Transparency:* Planning to systematically examine *all* AEs (providing enough of them have been reported for a statistical test to be meaningful: see next chapter), irrespective of where they feature in the MedDRA hierarchy, is the only way to guarantee refuting any future suggestion that the assessments performed were deliberately selected and/or incomplete in a way to manipulate the results.

With the above considerations in mind, the 'families' considered for evaluation in the simulation studies were created according to the following principles:

- The PT level was to be selected as the lowest level of MedDRA for performing statistical tests (since events at the even more granular LLT level simply represent different synonyms or lexical variants of the PTs).

- For higher levels of MedDRA (ie, HLT and HLGT), any term with the same number of events as a term in the level immediately beneath it was not to be tested again. Indeed, only terms which contributed 'substantial additional statistical information' over and above a term immediately beneath it were to be tested. (The definition of what constitutes 'substantial additional statistical information' is presented and justified in the next chapter.)

- The inclusion of each MedDRA term reported at the three 'analysable' levels (ie, PT, HLT and HLGT) for multiple testing was to be based on the total number of events reported across both study arms (blinded to treatment allocation). For example, testing

any term with at least 10 or 100 events (given the caveats mentioned above about avoiding irrelevant overlap in tests at different levels).

- Although analyses at the SOC level may be insensitive (albeit very well powered) they are shown later for completeness (and there are only 27 of them so adding them doesn't materially change the power for any more sensitive test at a different MedDRA level).

3.5.2 Simulated scenarios

For the types of AE to be considered in the CTT AE project (ie, after removing AEs where there is a definite known benefit or hazard of statin therapy) the proportion of the remaining AEs genuinely influenced by statin therapy was expected to be low. However, under MHT theory (see chapter 2), this unknown proportion influences the statistical properties of the various MHT methods (eg, the expected number of false positives). Therefore, simulations were designed under each of the following possible scenarios:

a) No effect of statin therapy on any AE (ie, null hypothesis is true in 100% of cases)
b) Real effect of statin therapy on all AEs (ie, null hypothesis is false in 100% of cases)
c) No effect for 99% of hypotheses and real effect for the remaining 1%
d) No effect for 95% of hypotheses and real effect for the remaining 5%
e) No effect for 90% of hypotheses and real effect for the remaining 10%

For implementing scenario *a*, all hypotheses were simulated assuming a risk ratio (RR) equal to 1.00 (as described in steps 3 and 4 of section 3.5.3 below). For scenario *b*, all hypotheses were simulated assuming a series of RRs >1.00, ranging from 1.03 to 10.00. (Note that simulated scenarios assuming RRs ranging from 0.1 to 0.971 are not required because two-sided (null) hypothesis for each AEs will be tested which consequently allows for detection of both harmful and beneficial effects of statin therapy).

In order to implement scenarios *c*, *d* and *e*, each hypothesis was allocated a Bernouli distribution with probability of no effect of statin p_0 equal to 0.99, 0.95 or 0.90 for the three different scenarios respectively. All hypotheses assigned (from the Bernouli distribution) no effect of statin therapy were simulated assuming an RR equal to 1.00, while all other hypotheses, assigned (from the Bernouli distribution) a real effect of statin therapy, were simulated assuming RRs ranging from 1.03 to 10.00.

3.5.3 Algorithm development

The algorithm developed for producing the simulation analyses is described below:

1. For each of the 23 double-blind trials, combine the AEs from each contributing SDTM domain for that trial (**Figure 3.1**) into a single dataset (thus creating 23 AE datasets).

2. For each trial *i* (where i=1,…,23), extract the total number of participants experiencing each type of adverse event coded at the PT level, $\boldsymbol{C_{ij}}$ (j=1,…,n_i, where n_i is the total number of PTs recorded in trial i).

3. For each PT, sample $\boldsymbol{O_{ij}}$ from a binomial distribution with total $\boldsymbol{C_{ij}}$ and probability $\boldsymbol{p=RR/(1+RR)}$, where *RR* (the relative risk) initially takes the value 1.0 (corresponding to no effect of statin therapy on any type of event) but is subsequently allowed to take each of the following values: 1.03, 1.05, 1.07, 1.10, 1.15, 1.20, 1.30, 1.40, 1.50, 1.70, 1.90, 2.00, 2.50, 3.00, 4.00, 5.00, 7.00 or 10.00.

4. Then, move to the HLT level and extract the total number of participants affected $\boldsymbol{D_{ij}}$ (j=1,…,k_i, where k_i is the total number of HLTs recorded in trial i). To account for the correlation structure of MedDRA (ie, a participant might report multiple PTs which are grouped into the same HLT) and also the fact that some AE terms were directly coded to the HLT level, for each HLT, the number of events $\boldsymbol{E_{ij}}$ in the statin arm was then calculated as the product of the total number of events in both arms $\boldsymbol{D_{ij}}$ and the overall

probability of events occurring in the statin arms derived from all PTs reported underneath each HLT using the following formula:

$$E_{ij} = D_{ij} * \frac{\sum_j^{n_i} O_{ij}}{\sum_i^{n_i} C_{ij}}$$

Then, the same process was repeated for the HLGT level. The total number of participants reporting each type of recorded AE was extracted and then the same formula as above was applied using the PT-level data to calculate the number of events in the statin arm.

5. For every analysable term (ie, reporting at least one event) at the three MedDRA levels (PT, HLT and HLGT), calculate the trial specific ***'o-e'*** test statistic and its associated variance ***'v'*** using the following formulas:

$$o\text{-}e = E\text{-}(N_1 * C)/Z \text{ and}$$

$$v = N_1 * C/Z * (1\text{-}N_1/Z) * (Z\text{-}C)/(Z\text{-}1)$$

where ***E*** = events in statin arm, N_1 = number of participants in statin arm, ***C*** = events in both arms and ***Z*** = total number of participants randomised (in both arms) (112).

6. For the meta-analysis of trials (ie, statin vs placebo) and every analysable term, calculate the log of the relative risk as ***S/V*** with variance ***1/V*** where ***S*** is the sum over all trials of ***o–e*** and ***V*** is the sum over all trials of ***v***.

7. Calculate the associated (2-tailed) p-value by comparing the statistic S^2/V against a chi-square distribution (ie, χ^2) with *1* degree of freedom.

   Repeat steps 3 – 7 10,000 times to allow the simulation algorithm to produce accurate and robust results.

8. Select the *family* of hypotheses (see section 3.5.1) to consider for MHT analysis. Then, for each of the 10,000 repetitions, apply the six MHT approaches (i – no MHT

correction, ii – Holm (FWER), iii – Hochberg (FWER), iv – Hommel (FWER) v – Benjamini & Hochberg (FDR) and vi – Mehrotra & Adewale (D-FDR)) to detect the hypotheses flagged as significant by comparing the p-values with the critical values generated from each method. (An example of how those 6 methods are used in practice is shown in section 3.5.4 below).

9. Finally, for the six MHT approaches, produce statistical estimates (based on the results of the 10,000 simulations) to determine the distribution of false positives and false negatives.

3.5.4 MHT methods applied

Based on the theoretical strengths of the various MHT approaches (see chapter 2) six approaches were considered for assessment by simulation: (i) No correction for MHT (as a control strategy), (ii) Holm (FWER), (iii) Hochberg (FWER), (iv) Hommel (FWER), (v) Benjamini & Hochberg (FDR) and (vi) Mehrotra & Adewale (D-FDR). For demonstrative purposes (and to give a better understanding of how these six MHT approaches work in practice) **Box 3.3** shows hypothetical results for 10 specific AEs terms across three levels of MedDRA. With such results, the six approaches above would lead to the following conclusions (**Box 3.3)**.

i) **No MHT:** Five of the ten AEs would be declared 'significant' at the 5% level.

ii) **Holm:** When sorted in ascending order the p-values would be compared against the following critical values: 0.005, 0.0056, 0.0063, 0.0071, 0.0083, 0.01, 0.0125, 0.0167, 0.0250 and 0.05. With Holm's method only the AE with the smallest p-value (the HLGT of Vascular therapeutic procedures, p=0.003) would be declared significant, because its p-value is smaller than the 1st critical value above (0.005) whereas the *next* smallest p-value (of 0.006, corresponding to the HLT of Aortic therapeutic procedures) is greater than the 2nd critical value above (of 0.0056).

iii) **Hochberg:** When sorted in descending order the p-values would be compared against the following critical values: 0.05, 0.025, 0.0167, 0.0125, 0.01, 0.0083, 0.0071, 0.0063, 0.0056, 0.005. Hence, with the Hochberg method only the AE with the smallest p-value (the HLGT of Vascular therapeutic procedures, p=0.003) would be declared as significant.

iv) **Hommel**: In the first step, all p-values are sorted in descending order and compared against *0.05k/j* (with *k* ranging from *1* to *j)*. The Hommel method identifies that the maximum *j* for this example is equal to 8. Then, in the second step, the Hommel method compares all p-values with the critical value 0.0063 (ie, 0.05/8) resulting in the detection of two p-values as significant (the HLGT of Vascular therapeutic procedures with p=0.003 and the HLT of Aortic therapeutic procedures with p=0.0056).

v) **Benjamini & Hochberg:** When sorted in descending order the p-values would be compared against the following critical values: 0.05, 0.045, 0.04, 0.035, 0.03, 0.025, 0.02, 0.015, 0.01, 0.005. With Benjamini & Hochberg method the first AE which would be declared significant is the HLT of Aortic therapeutic procedures (p=0.006). Thereafter, all remaining AEs would also be declared a significant without requiring further testing (the HLGT of Vascular therapeutic procedures, p=0.003).

vi) **Mehrotra & Adewale:** In the first step, the representative (ie, minimum) p-values for each SOC were identified (0.003 and 0.027), sorted in ascending order and compared against the following critical values: 0.025 and 0.05. Since both p-values would be declared as significant, in the second step, Mehrotra & Adewale would allow all AEs underneath both SOCs to be tested therefore resulting in identical findings to the Benjamini & Hochberg method shown above (both the HLGT of

Vascular therapeutic procedures with p=0.003 and the HLT of Aortic therapeutic procedures with p=0.0056).

3.6 Statistical methods for meta-analysis

This section describes the statistical methods used for the meta-analyses of AEs presented in chapters 5 and 6. Analyses were to include all randomised participants irrespective of whether they received their allocated treatment (ie, intention-to-treat). AEs reporting at least one event were included for analyses, with a time to event analysis being performed for the first occurrence of the specific AE. The statistical analyses can be described in three steps, as below.

3.6.1 Time to event

For every AE reported in a particular trial, the logrank method was used to calculate the event rates. The logrank method is based on the times of events where, for each such time point, it calculates the observed number of events (o) in each treatment arm and the number expected (e) if there were in reality no difference between the two treatments (113). These calculations are performed each time an event occurs with the observed minus expected (o–e) statistic and its variance (v) calculated from the totals of the observed and expected events.

3.6.2 Meta-analysis

In a meta-analysis of trials, the log of the rate ratio of was calculated as S/V with variance 1/V where S is the sum over all trials of (o–e) and V is the sum over all trials of v (114, 115) and hence the 95% confidence interval calculated as $S/V \pm 1{\cdot}96/\sqrt{V}$. Then, the p-value (uncorrected for MHT) was calculated from the comparison of the test statistic (ie, S/V) against a chi-square distribution (ie, χ^2) with 1 degree of freedom.

3.6.3 Multiple testing

The description of the best (ie, most appropriate) MHT approach used to test the multiple null hypotheses of all AEs examined in the meta-analyses is shown in chapter 5, with chapter 4 used to select the best method based on various estimates (ie, distribution of false positives and statistical power).

3.7 Synopsis

This chapter has summarised the methods used for processing the data and performing the statistical analyses of the effects of statin therapy on all AEs with a previously unknown effect. The methods used for the harmonisation of the IPD were based on the global standards of the CDISC SDTM data tabulation model. In addition, MedDRA V20.0 was employed to standardise all AEs reported across the 28 trials into a common medical dictionary. Simulation methods were developed to identify the most appropriate MHT method and testing strategy for the meta-analysis of the AEs examined in this thesis. These simulations account for multiple scenarios of treatment effects and the proportion of AEs affected by statin therapy, as well as the correlation structure (ie, correlation by construction) among the AEs examined and the population of interest. The results of the simulation analyses are described in the next chapter.

Box 3.1: CDISC SDTM (V3.2) domains created for the completion of the CTT AE project including appropriate modifications

SDTM Domain Models	Description
Adverse Events [AE]	Adverse events which may be reported either as free text or a as terms created from a coding system (eg, ICD or MedDRA).
Clinical Events [CE]	Trial specific non-fatal events which were not classified as adverse events
Concomitant/Prior Medications [CM]	Concomitant and prior medications/therapies used by the participant. Examples are the concomitant medications/therapies given on an as needed basis and the usual background medications/therapies given for a condition.
Death Details [DD]	Any fatal events regardless of cause.
Demographics [DM]	DM is the parent domain for all other domains and includes a set of essential standard variables that describe each participant in the clinical study. Examples are the participants' characteristics at baseline (eg, treatment allocation, age and race).
Exposure [EX]	Reasons for stopping (or adjusting) study treatment. For CTT purposes, only adverse events resulting in stopping (or adjusting) study treatment were sought.
Healthcare Encounters [HO]	Inpatient and outpatient healthcare events (eg, hospitalizations, nursing home stay, rehabilitation facility stays and ambulatory surgery).
Laboratory Test Results [LB]	Laboratory test findings including, but not limited to hematology, clinical chemistry and urinalysis data. For CTT purposes, only the following laboratory tests were captured: 1) Alanine aminotransferase, 2) Aspartate aminotransferase, 3) Creatine kinase (Total creatine kinase), 4) Glucose in blood, 6) High Density Lipoprotein (HDL) cholesterol, 7) Low Density Lipoprotein (LDL) cholesterol, 8) Triglyceride, 9) Cholesterol (Total), 10) Creatinine conc (serum), 11) eGFR, 12) HbA1c and 13) Platelets (thrombocytes).
Medical History [MH]	The participant’s prior history at the start of the trial (eg, general medical history, gynecological history, and primary diagnosis).
Procedures [PR]	Details describing a participant’s therapeutic and diagnostic procedures (eg, coronary artery bypass graft, cataract surgery or hip replacement).
Substance Use [SU]	Substance use information that may be used to assess the efficacy and/or safety of therapies that look to mitigate the effects of chronic substance use. For CTT purposes, only smoking status and alcohol consumption at baseline were collected.
Subject Visit [SV]	Information about the timing of participant visits that is otherwise spread over domains that include the visit variables. For CTT purposes, SV domain was created only when imputation methods were required for timing variables.
Vital Signs [VS]	Measurements including but not limited to blood pressure, temperature, respiration, body surface area, BMI, height and weight. For CTT purposes, only the following measurements were captured: 1) Systolic blood pressure, 2) Diastolic blood pressure, 3) Heart Rate, 4) BMI, 5) Height, 6) Weight and 7) Waist circumference.

Box 3.2: The five-tiered hierarchical structure of Medical Dictionary for Regulatory Activities (MedDRA)

MedDRA Hierarchical Level	Description	Number of Terms
Lowest Level Term [LLT]	A reflection of how observations might be reported in practice eg a verbatim term. Each LLT is linked to only one PT.	77248
Preferred Term [PT]	A single descriptor such as a symptom, sign, disease diagnosis, investigation or procedure. Each PT has at least one LLT (itself) as well as synonyms and lexical variants (eg. abbreviations, different word order).	22499
High Level Term [HLT]	A group of PTs that are related based on anatomy, pathology, physiology, aetiology, or function.	1553
High Level Group Term [HLGT]	Related HLTs are linked to HLGTs based on anatomy, pathology, physiology, aetiology, or function.	326
System Organ Class [SOC]	The final level of the hierarchy where HLGTs are grouped by aetiology (e.g. Infections and infestations), manifestation site (e.g. Gastrointestinal disorders) or purpose (e.g. Surgical and medical procedures). There is also a SOC to contain issues pertaining to products and one to contain social circumstances.	27

Box 3.3: Hypothetical example of a meta-analyses of 10 AEs reported through the MedDRA hierarchical levels and their associated p-values

System Organ Class	MedDRA Term	P-value	Detection of AE as significant at 5% level					
			No MHT correction (Type-I)	Holm (FWER)	Hochberg (FWER)	Hommel (FWER)	Benjamini & Hochberg (FDR)	Mehrotra & Adewale
Surgical and medical procedures	HLGT: Vascular therapeutic procedures	0.003	Yes	Yes	Yes	Yes	Yes	Yes
	HLT: Aortic therapeutic procedures	0.006	Yes	No	No	Yes	Yes	Yes
	PT: Arterial stent insertion	0.049	Yes	No	No	No	No	No
	PT: Limb operation	0.410	No	No	No	No	No	No
	PT: Amputation	0.061	No	No	No	No	No	No
Gastrointestinal disorders	HLGT: Exocrine pancreas conditions	0.415	No	No	No	No	No	No
	HLT: Acute and chronic pancreatitis	0.113	No	No	No	No	No	No
	HLT: Gastrointestinal disorders NEC	0.047	Yes	No	No	No	No	No
	PT: Gastric dysplasia	0.075	No	No	No	No	No	No
	PT: Gastric haemorrhage	0.027	Yes	No	No	No	No	No

FWER=Familywise error rate. FDR=False discovery rate.

Table 3.1: Characteristics of 23 double-blinded and 5 open-label statin trials and data provided

Study	Year of publication of primary results	Number of patients	Treatment comparison (mg/day)	Median follow-up (years)	Meta-data CRF (F) CSP (P) CSR (R)	Total number of			Adverse event terms		
						Datasets	Records	Variables	Main Medical Terminology†	Total	Unique
Statin vs placebo, double-blinded trials											
4S	1994	4,444	S20-40 vs placebo	5.4	FP	5	116,502	147	MedDRA*	33,661	1,390
WOSCOPS	1995	6,595	P40 vs placebo	4.8	FP	23	671,759	509	BMS ICD9	78,673	7,978
CARE	1996	4,159	P40 vs placebo	4.9	FPR	9	1,849,516	275	BMS ICD9	91,635	753
AFCAPS/TexCAPS	1998	6,605	L20-40 vs placebo	5.0	P	5	995,156	97	MedDRA*	8,143	770
LIPID	1998	9,014	P40 vs placebo	5.9	FPR	28	1,347,182	592	ICD9	57,636	7,075
LIPS	2002	1,677	F80 vs placebo	4.0	FPR	61	537,144	1,145	ICD10	8,867	2,392
HPS	2002	20,536	S40 vs placebo	5.2	FP	28	8,152,101	392	MedDRA 14	76,986	2,378
PROSPER	2002	5,804	P40 vs placebo	3.3	FP	11	628189	485	Modified ICD9	131,598	848
ASCOT-LLA	2003	1,0240	A10 vs placebo	3.3	FP	10	710,266	466	ICD9	42,294	1,896
ALERT	2003	2,102	F40 vs placebo	5.5	FPR	79	1,069,747	5,654	ICD-9/MedDRA 5.1	27,114	3,589
CARDS	2004	2,838	A10 vs placebo	4.2	FP	17	988,712	522	MedDRA 11	26,485	1,890
4D	2005	1,255	A20 vs placebo	2.7	FP	21	302,442	632	MedDRA 11	6,758	461
ASPEN	2006	2,410	A10 vs placebo	4.0	FP	18	472,820	379	MedDRA 11	20,323	2,861
SPARCL	2006	4,731	A80 vs placebo	4.9	FP	14	900,713	512	MedDRA 11	38,811	3,539
CORONA	2007	4,982	R10 vs placebo	2.8	FPR	62	2,155,519	1,716	MedDRA 10	28,899	3,450
JUPITER	2008	16,714	R20 vs placebo	1.9	FPR	33	1,540,095	2,428	MedDRA 11.1	73,526	5,731
GISSI-HF	2008	4,574	R10 vs placebo	3.9	F	10	208,489	169	MedDRA 15.1	28,701	537
AURORA	2009	2,555	R10 vs placebo	3.9	FPR	42	724,490	1,273	MedDRA 11.1	39,512	4,013
HOPE-3	2016	12,705	R10 vs placebo	5.6	FP	170	1,415,309	6,213	MedDRA 17.1	70,651	7,631
Subtotal		**123,940**				**646**	**24,786,151**	**23,606**		**890,273**	**40,221**
Statin vs other control/usual care, open label trials											
POST-CABG	1997	1,351	L40-80 vs L2.5-5	4.3	FP	83	227,740	2,821	None	37,164	650
GISSI-P	2000	4,271	P20 vs no treatment	1.9	FP	1	4,271	795	ICD-9	2,777	228
ALLHAT-LLT	2002	10,355	P40 vs no treatment	4.7	FP	10	381,498	103	ICD-9/ICD-10	9,932	115
ALLIANCE	2004	2,442	A10-80 vs usual care	4.5	FP	15	136,557	344	MedDRA 11	3,717	1,334
Subtotal		**18,419**				**109**	**750,066**	**4,063**		**53,590**	**2,311**
More vs less statin, double-blinded trials											
PROVE-IT	2004	4,162	A80 vs P40	2.1	FP	24	603,289	1,161	ICD-9	25,600	1,551

A to Z	2004	4,497	S40 then S80 vs placebo then S20	2.0	FP	10	386,787	405	MedDRA*	4,615	561
TNT	2005	10,001	A80 vs A10	5.0	FP	20	2,676,152	624	MedDRA 11	93,583	3,735
SEARCH	2010	12,064	S80 vs S20	7.0	FP	25	7,341,699	344	MedDRA 14	61,379	1,303
Subtotal		**30,724**				**79**	**11,007,927**	**2,534**		**185,177**	**6,210**
More vs less statin, open label trials											
IDEAL	2005	8,888	A40-80 vs S20-40	5	FP	11	1,999,354	292	MedDRA 13	59,129	2,281
Subtotal		**8,888**				**11**	**1,999,354**	**292**		**59,129**	**2,281**
Total		**181,973**				**845**	**38,543,498**	**30,495**		**1,188,169**	**45,230**

†Most trials that used a medical terminology also provided some events in free text format. Some trials used more than one medical terminology.

*MedDRA version was unknown.

S=Simvastatin; P=Pravastatin; L= Lovastatin; F=Fluvastatin; A=Atorvastatin; R=Rosuvastatin

CRF=Case report forms, CSP=Clinical study protocol, CSR= Clinical study report.

4S=Scandinavian Simvastatin Survival Study; WOSCOPS=West of Scotland Coronary Prevention Study; CARE=Cholesterol And Recurrent Events; AFCAPS/TexCAPS=Air Force/Texas Coronary Atherosclerosis Prevention Study; LIPID=Long-term Intervention with Pravastatin in Ischaemic Disease; LIPS=Lescol Intervention Prevention Study; HPS=Heart Protection Study; PROSPER=PROspective Study of Pravastatin in the Elderly at Risk; ASCOT-LLA=Anglo-Scandinavian Cardiac Outcomes Trial-Lipid Lowering Arm; ALERT=Assessment of Lescol in Renal Transplantation; CARDS=Collaborative Atorvastatin Diabetes Study; 4D=Die Deutsche Diabetes Dialyse Studie; ASPEN=Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus; SPARCL=Stroke Prevention by Aggressive Reduction in Cholesterol Levels; CORONA=Controlled Rosuvastatin Multinational Trial in Heart Failure; JUPITER=Justification for the Use of Statins in Prevention: an Intervention Trial Evaluatin g Rosuvastatin study group; GISSI-HF=Gruppo Italiano per lo Studio della Sopravvivenza nell'Insufficienza cardiaca; AURORA=A Study to Evaluate the Use of Rosuvastatin in Subjects on Regular Hemodialysis: An Assessment of Survival and Cardiovascular Events; HOPE-3=Heart Outcomes Prevention Evaluation-3 trial; Post-CABG=Post-Coronary Artery Bypass Graft; GISSI-P=Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico; ALLHAT-LLT=Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ALLIANCE=Aggressive Lipid-Lowering Initiation Abates New Cardiac Events; PROVE-IT=Pravastatin or Atorvastatin Evaluation and Infection Therapy; A to Z=Aggrastat to Zocor; TNT=Treating to New Targets; SEARCH=Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine. IDEAL=Incremental Decrease in End Points Through Aggressive Lipid Lowering Study Group.

A small number of participants in the AURORA (n=218), CORONA (n=27) and JUPITER (n=1088) trials withdrew consent for use of their data post-trial, and hence data from these participants is excluded. The ASCOT-LLA trial excludes 65 patients for whom data was not available due to protocol violations.

Table 3.2: Hierarchical levels of MedDRA version 20, organized alphabetically by System Organ Class

	Number of terms at each sublevel			
System Organ Class	**Higher Level Group Term**	**Higher Level Term**	**Preferred Term**	**Lower Level Term**
Blood and lymphatic system disorders	10	32	284	1135
Cardiac disorders	9	29	335	1430
Congenital, familial and genetic disorders	19	98	1311	3462
Ear and labyrinth disorders	5	14	87	428
Endocrine disorders	7	25	187	668
Eye disorders	12	50	592	2432
Gastrointestinal disorders	21	88	847	3807
General disorders and administration site conditions	7	36	977	2412
Hepatobiliary disorders	3	14	191	650
Immune system disorders	4	16	142	457
Infections and infestations	12	149	1907	7079
Injury, poisoning and procedural complications	8	71	1128	6496
Investigations	23	106	5518	13512
Metabolism and nutrition disorders	13	46	277	939
Musculoskeletal and connective tissue disorders	10	38	444	2511
Neoplasms benign, malignant and unspecified (including cysts and polyps)	39	194	1942	8490
Nervous system disorders	18	76	938	3531
Pregnancy, puerperium and perinatal conditions	8	32	222	1633
Product issues	2	21	145	589
Psychiatric disorders	23	69	508	2341
Renal and urinary disorders	8	28	350	1202
Reproductive system and breast disorders	14	44	475	1715
Respiratory, thoracic and mediastinal disorders	10	33	519	1696
Skin and subcutaneous tissue disorders	8	47	489	2054
Social circumstances	7	20	271	635
Surgical and medical procedures	19	141	2107	4632
Vascular disorders	11	36	306	1312
Total number of potential pairwise tests*	**326**	**1553**	**22499**	**77248**

This is the minimum number of tests required (ie, assuming exactly one statistical test for each term at each level is performed).

Table 3.3: Number of total and unique adverse event terms mapped across the hierarchical levels of MedDRA (V20.0)

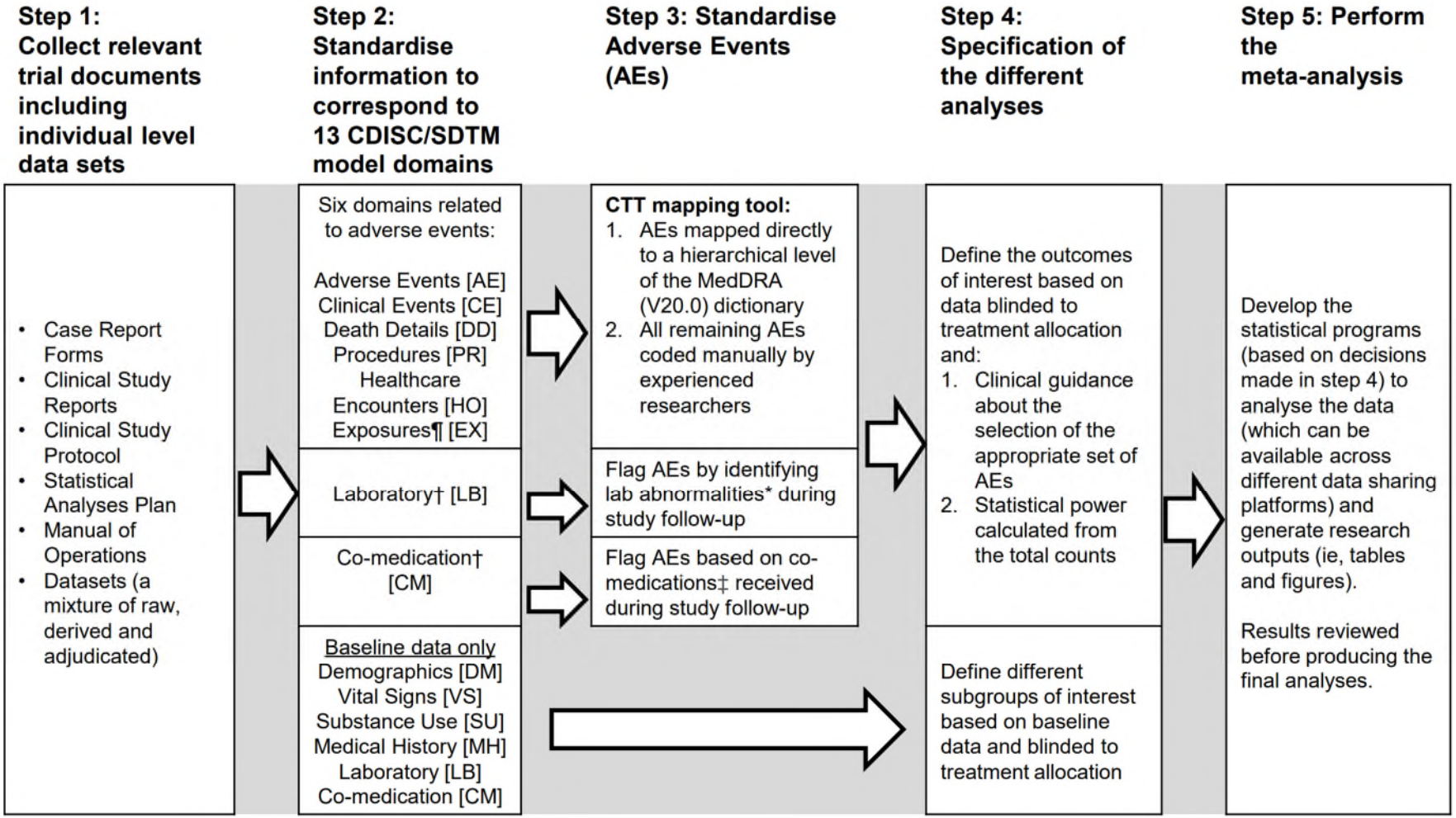
Type and MedDRA level of coding	Adverse events, N (%)	
	Total*	**Unique terms†**
Automatically-coded events		
Lower level term	779,449 (65.6%)	11,965 (26.5%)
Higher level term	6,216 (0.5%)	659 (1.5%)
Higher level group term	820 (<0.1%)	1 (<0.1%)
Subtotal	786,485 (66.2%)	12,625 (27.9%)
Manually-coded events		
Lower level term	391,338 (32.9%)	32,458 (71.8%)
Higher level term	7,571 (0.6%)	116 (0.3%)
Higher level group term	2,775 (0.2%)	31 (<0.1%)
Subtotal	401,684 (33.8%)	32,605 (72.1%)
Total	**1,188,169 (100%)**	**45,230 (100%)**

* Total across all 28 trials. Numbers represents the number of adverse event terms provided by the trialists that were either automatically or manually coded into particular MedDRA levels (lower level, higher level, and higher level group).

† As above, but counting the number of unique adverse event terms coded.

Adverse events were not coded to the MedDRA 'Preferred Term' (PT) level because each PT has an identical lower level term (LLT) available. Thus, once all LLTs have been coded, coding of PTs would provide no additional information.

Figure 3.1: Data analysis flow diagram



CDISC=Clinical Data Interchange Standards Consortium, SDTM= Study Data Tabulation Model. ¶ The exposure domain reported only treatment stops related to adverse events. † Measurements taken during study follow up. *Only laboratory abnormalities related to muscle outcomes (ie, creatine kinase) or diabetes mellitus (ie, glucose or HbA1c) were considered in the current project. ‡ Only co-medications which could lead to the identification of additional incident cases of diabetes mellitus were considered in the current project.

4 Simulation analyses

4.1 Summary

This chapter presents the results from simulation studies developed to identify the most appropriate MHT method (of the five identified as potentially suitable) for use in the CTT AE project. It also illustrates the impact (on type I and type II error) of the choice of the family of AEs considered for meta-analyses. Each of the five approaches are compared against each other as well as against an approach of not controlling for multiple testing at all. The results show very clearly how all three MHT methods that control the FWER (ie, Holm, Hochberg and Hommel) reduce the likelihood of any false–positive results to a negligible probability, but do so at a substantial cost of having low statistical power to detect any true (beneficial or harmful) effects of statin therapy. By contrast, the two MHT methods that control the FDR (ie, Benjamini & Hochberg and Mehrotra & Adewale) perform similarly, with each still greatly reducing the likelihood of false-positive results, but doing so whilst preserving reasonable statistical power for the detection of moderate treatment effects. With respect to the family of hypotheses to be tested, the simulation studies reveal how statistical power varies with both the number of patients affected by a given AE and the decisions made regarding repeatedly testing terms at higher levels of MedDRA. Based on the results of the simulation studies a recommendation is made for the preferred approach for hypothesis testing in the CTT AE project.

4.2 Introduction

The aims of the simulation analyses in this chapter are to: (i) monitor the performance of the various multiple testing methods (found in the literature as suitable for use in the CTT database) with respect to their statistical properties (ie, expected number of false positives (FPs) and statistical power) in order to identify the best method for use in these meta-analyses;

and (ii) develop an integrated MHT framework for the selection of the appropriate AEs (ie, based on the total number of events and/or the hierarchical level of MedDRA) to be considered for meta-analyses.

This chapter presents the results generated from the various simulation studies, summarises the key findings and makes recommendations for the appropriate MHT framework to use in the CTT AE project. The simulation methods are outlined in chapter 3.5.

4.3 Methods

For every PT with a non-zero event count (across all trials together) the following five scenarios were simulated: i) no effect of statin therapy on any AE (ie, null hypothesis is *true* in 100% of cases); ii) real effect of statin therapy on all AEs (ie, null hypothesis is *false* in 100% of cases); iii) no effect for 99% of hypotheses and real effect for the remaining 1%; iv) no effect for 95% of hypotheses and real effect for the remaining 5%; and v) no effect for 90% of hypotheses and real effect for the remaining 10%. The null hypothesis of no effect (of statin) was implemented by forcing the true RR to be equal to 1.00. For the alternative hypotheses of a real effect, a range of RRs was considered (from 1.03 to 10).

Within each trial, for each PT with a non-zero event count, the events in the statin arm were simulated from a binomial distribution with probability equal to RR/(1+RR). Then, for each HLT and HLGT, the number of events in the statin arm was derived from the total number of events (for each particular term) and its mathematical relationship with the simulated events from the PTs underneath it. Thereafter, for each PT, HLT and HLGT within each trial, the observed (o) minus expected (e) test statistic and its associated variance (v) was calculated.

In a meta-analysis of trials, the log of the risk ratio was calculated as S/V with variance 1/V where S is the sum over all trials of (o–e) and V is the sum over all trials of v, and hence the 95% confidence interval was calculated as $S/V \pm 1{\cdot}96/\sqrt{V}$. Then, the p-value (uncorrected for

MHT) was calculated from the comparison of the test statistic (ie, S/V) against a chi-square distribution (ie, χ^2) with 1 degree of freedom.

When performing hypothesis testing, any AE reported for fewer than six participants cannot yield a two-sided p-value <0.05 (as the most extreme result, of 0 vs 5 or 5 vs 0, would yield a two-sided p-value of 0.0625). Consequently, only MedDRA AE terms reported for at least six participants were included in the simulation analyses.

For each of the six MHT approaches (i – no MHT correction, ii – Holm , iii – Hochberg, iv – Hommel, v – Benjamini & Hochberg and vi – Mehrotra & Adewale) the number of incorrectly-rejected hypotheses (ie, false positives) and correctly-rejected hypotheses (ie, true positives) was calculated across the various simulated scenarios. The simulation analyses were repeated 10,000 times to produce robust and accurate statistical estimates.

The simulation analyses were performed only for the 19 double-blinded trials of statin vs placebo (involving 123,940 participants) because this was the population of primary interest for hypothesis-testing. However, appropriate interpretation of the findings from the simulation analyses with regards to the secondary analysis population of the 4 double-blinded trials of a more vs a less intensive statin regimen (involving 30,724 participants) are also described at the end of this chapter.

4.4 Simulation results for trials of statin vs placebo

4.4.1 Data

4.4.1.1 CTT AE database

The simulation analyses were performed using the data available in January 2020, at which point the CTT AE database included about one million AEs (with 35,000 unique terms) of which 790,000 (20,000 unique terms) had been successfully coded to MedDRA. Of these, about 480,000 were reported in the trials of statin vs placebo.

4.4.1.2 Distribution of events across the MedDRA tree

For the simulation analyses, only first events of the same type of AE were included in the analyses, which means that each participant who developed a specific type of AE (as reported through the hierarchical levels of MedDRA) contributed only once in the calculation of the 'o-e' test statistic (and its associated variance) for count data (see chapter 3.5).

The distribution of the total number of events across the three analysable MedDRA hierarchical levels (ie, PT, HLT and HLGT), organised alphabetically by the 27 System Organ Classes, is presented in **Table 4.1**. Based on the CTT AE data as of January 2020, in the trials of statin vs placebo, there were 6,116 terms out of a total of 24,378 analysable MedDRA terms (ie, sum of 22,499 PTs, 1,553 HLTs and 326 HLGTs) which reported at least one event. Of those, 4,751 (78%) reported less than 50 events, 364 (6%) had a count of between 50 and 99 events, 756 (12%) reported between 100 and 999 events and 245 (4%) reported at least 1000 events. From the total of 6,116 terms (reported across the MedDRA hierarchical tree), 4,765 (77.9%) were PTs (with only 16% and 11% of those terms reported at least 50 and 100 events respectively), 1,086 (18%) were HLTs (with 31% of those terms reported at least 100 events) and 326 (5%) were HLGTs (with 51% of those terms reported at least 100 events).

The distribution of event counts for AEs with potentially significant hypotheses across the three analysable MedDRA levels is shown in **Figure 4.1**. The median (interquartile range) number of participants experiencing an event at each MedDRA level was 27 (11 to 92) across the 2,080 PTs, 87 (24 to 356) across the 675 HLTs and 453 (94 to 1392) across the 208 HLGTs.

4.4.2 No correction for multiple testing

4.4.2.1 Distribution of false positives results when the null hypothesis true

In CTT, a very large number of hypotheses are available for testing, meaning that a large number of false positive results would be obtained if there were no correction for MHT. The simulated distribution of false positives for various types of AEs (based on event counts) when all hypotheses are assumed to be null (ie, no effect of statin therapy) and no correction for

multiple testing is applied is presented in **Figure 4.2**. When testing AEs with at least 6 events, a total of 2,963 tests were performed and the mean number of FP was 144 (SD 14). When increasing the threshold number of events for which AEs were considered to be tested to 50, 100 or 1000, the total number of tests was reduced to 1,325, 977 and 238 respectively, resulting, on average, in 72, 55 and 15 FP results respectively.

The expected probability of a type I error was found to be positively associated with the threshold number of events. For example, when testing AEs with at least 6 events, the mean (SD) type I error was equal to 4.86% (SD 0.47%), but this increased to 5.43% (SD 0.79%) when testing only AEs with at least 50 events and increased further to 6.40% (SD 2.11%) when limiting the testing to AEs with 1000 events. Although the statistical theory indicates that the probability of a type I error (when all the assumptions hold, for example, when performing the t-test for normally distributed data) is expected to be 5%, in this particular data example and for binary outcomes, the type I error is shown to deviate from the threshold level of significance for various reasons (108, 116-120) with the following two being the most important:

1) *Discrete distribution:* for binary outcomes, the total number of events influences the exact probability of a type I error (119). For example, suppose the total number of events for a particular AE is equal to 10 and the null hypothesis of no effect of treatment is of interest. In such a scenario, the p-value will be rejected at the 5% significance level only when zero, one, nine or ten events occur in one of the two arms, with the actual probability of a type I error being equal to 2.3%. The probability of a type I error converges to the threshold level (ie, 5%) only as the total number of events increase.
2) *Correlation by construction among hypotheses:* this refers to how a result of a test at a lower MedDRA level (eg, PT) affects the result of a test at a higher MedDRA level (eg, HLT or HLGT). In probability theory, such correlation is expressed as the binomial sum variance inequality (117, 120). This inequality states that the sum of independent binomial random variables is itself a binomial random variable if all the component

variables share the same success probability. However, when success probabilities differ, the probability distribution of the sum is not binomial and the effects of such correlation structure on the probability of type I error (and possibly type II error) are unknown. (Note that the binomial sum variance inequality is also expected to affect the statistical properties of various MHT methods, especially those controlling for the FDR error (118)).

4.4.2.2 Statistical power of detecting true effects

Figure 4.3 presents the expected power (calculated from the 10,000 simulations) to reject the null hypothesis at the 5% significance level when statin therapy is assumed to increase the risk of adverse events by between 3% and 900% (ie, RRs between 1.03 and 10) and when no correction for MHT is applied. For AEs reported by a low number of participants (fewer than 50 participants, mean 18 participants), there was very limited statistical power to detect any effect of statin therapy even when the effect size was very large. For example, the expected power would only be about 70% even when the true RR was assumed to be 5. The expected power increases rapidly when testing AEs with more events however. For example, for AEs with between 50 and 99 events (mean 71 events) there would be 80% power to detect RRs >1.85, for AEs with between 100 and 999 events (mean 333 events) there would be 80% power to detect RRs >1.40, and for AEs with more than 1000 events (mean 4367 events) there would be 80% power to detect RRs >1.15.

Figure 4.4 shows the mathematical relationship between the RR and the lowest number of events needed for such an RR to be declared significant at the 5% or 1% level. A RR of 2.5 would require a total of at least 19 and 44 events to be declared as significant at the 5% and 1% level respectively, while an RR of 1.5 would require at least 100 and 170 events to be declared as significant at the 5% and 1% level respectively in a population of 120,000 participants.

4.4.3 Comparison of MHT methods

Five MHT methods were identified as potentially suitable for use in the CTT AE project: three of which control the FWER (Holm, Hochberg and Hommel) and two of which control the FDR (Benjamini & Hochberg and Mehrotra & Adewale) (see chapter 2). The circumstances under which each of these five approaches concludes 'significance' or 'lack of significance' for a given AE vary (and are described in chapter 2.5) but all depend on the ordering of the uncorrected p-values. The performance of each of these MHT procedures (with respect to power and false positive results) under the various scenarios outlined in sections 4.4.3.1 and 4.4.3.2 is described below.

4.4.3.1 Distribution of false positives results

Tables 4.2 (a) to **(d)** present the mean (95% CI) number of false positives produced when using the 5 different MHT methods (compared with the approach of no correction for MHT) and when applied to: (i) all AEs with event count ≥6 (**Table 4.2 (a)**); or the subset of AEs with event counts (ii) ≥50 (**Table 4.2 (b)**), (iii) ≥100 (**Table 4.2 (c)**) or (iv) ≥1000 (**Table 4.2 (d)**).

For the MHT methods controlling the FWER (ie, Holm, Hochberg and Hommel), the mean number of FPs was ≤0.6 for all scenarios and all three methods (**Tables 4.2 (a)** to **(d)**).

For the MHT methods controlling the FDR (ie, Benjamini & Hochberg and Mehrotra & Adewale), the mean number of FPs increased with the number of true effects assumed in the four simulated scenarios, as expected by the definition of the FDR. For example, when choosing to test all AEs with at least six events and assuming true adverse effects of statin for 1% of AEs tested (ie, 29 MedDRA terms, **Table 4.2 (a)**) the expected number of FPs was less than 2 for both MHT methods, whereas when 5% of AEs were assumed to be increased by statin therapy (ie, 148 MedDRA terms), the expected number of FPs was about 5 (**Table 4.2 (a)**). However, in consensus with the statistical theory (see chapter 2.3.2), such MHT methods (ie, controlling the FDR) resulted in similar distribution of FPs with MHT methods controlling

the FWER when no effect of statin was assumed in *any* AE tested (ie, all null hypotheses were true).

In addition, for MHT methods controlling the FDR, although the expected number of FPs decreased when performing fewer tests, the expected proportion of FPs among hypotheses declared as significant was inflated notably (**Tables 4.2 (a)** to **(d)**). For example, when assuming a true effect of statin for 5% of all AEs, the expected proportion of FPs was equal to 0.2% (ie, ~5/2963) when testing all AEs (**Table 4.2 (a)**), 0.4% when testing only AEs with at least 50 events (**Table 4.2 (b)**) and >0.6% when testing AEs with at least 1000 events (**Table 4.2 (d)**). Such inflation in the proportion of FPs may be because of the correlation by construction and the statistical theory of the binomial sum variance inequality (118) discussed earlier (see section 4.4.1.1).

Overall, the distribution of FPs produced by both MHT methods controlling the FDR was similar, with the Benjamini & Hochberg method often resulting in marginally fewer FPs compared with the Mehrotra & Adewale method, with the mean difference ranging between 0.1 to 0.5 FPs (**Tables 4.2 (a)** to **(d)**).

4.4.3.2 Statistical power of detecting true effects

Tables 4.3 (a) to **(d)** present mean (95% CI) estimates of statistical power for the 5 different MHT methods (compared with the approach of no correction for MHT) to detect true increases in risk ranging from 10% (ie, RR=1.10) to 400% (ie, RR=5.00). For each method, scenarios corresponding to true effect of statin for 1%, 5% and 10% of all hypotheses are considered when applied to: (i) all AEs with event count ≥6 (**Table 4.3 (a)**), or the subset of AEs with event counts (ii) ≥50 (**Table 4.3 (b)**), (iii) ≥100 (**Table 4.3 (c)**) or (iv) ≥1000 (**Table 4.3 (d)**).

MHT methods controlling the FDR produced substantially higher statistical power estimates compared with methods controlling the FWER (**Tables 4.3 (a)** to **(d)**). For example, when testing AEs with at least 100 events, the statistical power of detecting an RR of 2.0 ranged between 81% to 82% in methods controlling the FWER and 89% to 96% in methods controlling

the FDR under the three simulated scenarios (**Tables 4.3 (c)**). The distribution of power was virtually identical for all MHT methods controlling the FWER in all simulated scenarios (**Tables 4.3 (a)** to **(d)**). Similarly, both MHT methods controlling the FDR resulted in similar statistical power in all simulated scenarios with the Mehrotra & Adewale method occasionally being slightly more powerful than the Benjamini & Hochberg method (**Tables 4.3 (a)** to **(d)**). For example, when testing AEs with at least 100 events, the statistical power of detecting an RR of 2.0 was equal to 89% for the Benjamini & Hochberg method and 90% for the Mehrotra & Adewale when true effect (of statin) was assumed for 1% of hypotheses (10 AEs) and no effect (of statin) for the remaining 99% hypotheses (967 AEs) (**Tables 4.3 (c)**).

When testing all AEs with at least 6 events (2,963 AEs), all five MHT methods had low power to detect even large treatment effects (eg, RR>3.00) because most of those AEs (1,986) were reported by less than 50 participants. For example, when assuming a true effect (of statin) for 5% of AEs (resulting in a total of 148 tests), the average expected power to detect an RR of 3.00 was 73% if no correction for MHT was applied, 39% if a MHT method controlling the FWER was applied and 51% if a MHT method controlling for FDR was applied (**Tables 4.3 (a)**).

When selecting to limit the testing of AEs with at least 50 events (see Table 4.3 (b)) and for RRs greater than or equal to 2.00, the expected power was greater than 90% if no correction for MHT was applied, 62% if a MHT method controlling the FWER was applied and ranged between 72% and 83% if a method controlling the FDR was applied (**Tables 4.3 (b)**).

In general, the statistical power of the MHT methods controlling the FDR increased substantially with the number of hypotheses for which a true effect of statin therapy was assumed to exist. (**Tables 4.3 (a)** to **(d)**). For example, when testing AEs with at least 100 events and for a true effect of 50% increase (ie, RR=1.5), the expected power of the Benjamini & Hochberg (FDR) method was equal to 57% when such RR was assumed in 1% (ie, 10 tests) of all AEs, equal to 65% when this RR was assumed in 5% (ie, 49 tests) of AEs and equal to

70% when it was assumed in 10% (ie, 98 tests) of all AEs (**Tables 4.3 (c)**). Similar patterns were observed in MHT methods controlling the FWER but of smaller magnitude (**Tables 4.3 (a)** to **(d)**).

4.4.3.3 Effect of selection of family on statistical power

The number of hypotheses tested within a family of tests determines the threshold level of significance at which any individual hypothesis within that family is rejected or not (ie, the fewer hypotheses are being tested, the higher the probability of appropriately rejecting any individual hypothesis) (see chapter 1.5.1). Consequently, the selection of the family of hypotheses has a direct impact on the statistical power of obtaining true positive results (eg, the fewer AEs are being tested, the higher the likelihood of detecting true effects of statin therapy). (Note that the selection of the family of hypotheses has no impact on the statistical power when no correction for MHT is applied because this approach simply compares each individual p-value against a standard significance level (eg, test if p-value is less than 0.05).)

Figure 4.5 presents the impact on statistical power for the Mehrotra & Adewale (FDR) MHT method when testing four different families of AEs based on the total number of events (all AEs with at least 6, 50, 100 or 1000 events) in trials of statin vs placebo, in the extreme situation whereby all AEs are increased between 5% to 900% (ie, RR between 1.05 and 10.00).

The expected power of the Mehrotra & Adewale (FDR) MHT method increased when reducing the number of hypotheses being tested (**Figure 4.5**). For example, for AEs with between 100 and 999 events (panel C in **Figure 4.5**), the expected power when testing all AEs with at least 6 events was considerably lower (eg, an absolute difference in power between 5% to 7% for RRs between 1.10 to 1.70) compared with when testing only AEs with at least 50 events (ie, comparing the red and the blue line in that figure). Also, the gain in power depends on the number of hypotheses (ie, AEs) *not* being tested when selecting a family. For example, for AEs with between 100 and 999 events (panel C in **Figure 4.5**) the strategy to increase the

threshold level of events from 6 to 50 (ie, test only AEs with at least 50 events and exclude [ie, not test] the 1,638 AEs with between 6 and 49 events) resulted in higher gain in statistical power (ie, moving from blue to red line) than the strategy to further increase this threshold level from 50 to 100 AEs (ie, test only AEs with at least 100 events and exclude the 348 AEs with between 50 and 99 events) (ie, moving from red to green line).

Similar findings were observed when using the Benjamini & Hochberg (FDR) method and also all three MHT methods controlling the FWER (ie, Holm, Hochberg and Hommel).

4.4.3.3.1 Testing across the hierarchical levels of MedDRA

All AEs reported in the CTT database were coded through the various hierarchical levels of MedDRA V20.0 (see chapter 3.2.3). It was considered desirable to perform hypothesis testing across various MedDRA levels (ie, PT, HLT and HLGT) for various reasons such as the richness of the AE database and the heterogeneity between trials in how AEs were recorded (see chapter 3.2.3.1). The MHT strategy was to start with performing hypothesis testing at the PT level and then to move higher up the MedDRA tree (ie, HLT and HLGT) and test only the hypotheses which contributed significant extra information (ie, defined based on the total number of events reported) not already captured in the previous levels (ie, PT for HLT and PT or HLT for HLGT) (chapter 3.2.3.1).

The selection of the minimum cut off percentage of extra information (ie, additional events recorded within an AE higher up the MedDRA tree compared with the term with the maximum number of events underneath it) needed in order to perform a test at the HLT or HLGT level was based on two factors as described below:

1. *Correlation by construction:* due to the hierarchical structure of MedDRA, the results of hypothesis testing at one level (eg, PT) directly influences the results at higher levels (ie, HLT or HLGT) (see chapter 3.3.1). For example, in the scenario when testing the following two AEs, an HLT with 10% extra events compared with the PT underneath it, the two results (ie, p-values) most likely will be very similar. However, in the scenario

when testing the following two AEs, an HLT with 50% or 100% extra events compared with the PT underneath it, the two results, although correlated, might be quite different.

2. *Impact on statistical power:* the mathematical relationship between the total number of events and the expected statistical power (for a single test) for detecting an increase in risk of between 10%, 25%, 50% and 100% (ie, RR of 1.10, 1.25, 1.50 and 2.0) is presented in **Figure 4.6**. It can be seen that an increase in the total number of events of 50% or more would result in a substantial increase in statistical power to detect various RRs. For example, for RRs between 1.25 and 2.00, an increase in the total number of events from 50 to 75 events resulted in an absolute increase in power of between 7% and 20%, while for RRs between 1.10 and 1.25, an increase in the number of events from 200 to 300 resulted in an absolute increase in power of between 6% to 15%.

4.4.3.4 Summary for comparison of MHT methods

The aims of the simulation studies in this chapter were: (i) to document the statistical properties (ie, number of FPs and statistical power) of the various statistical approaches to multiple testing and identify the most suitable method for use in these meta-analyses; and (ii) to develop an integrated MHT framework for the selection of the appropriate AEs (ie, based on number of events and/or the hierarchical level of MedDRA) to be considered for meta-analyses.

As expected, not correcting for MHT would result in a large number of FPs with the expected proportion of hypotheses flagged as FPs ranging between 4.9% to 6.3%, depending on the type of AEs analysed. By contrast, all three MHT methods controlling the FWER (ie, Holm, Hochberg and Hommel) were shown to produce almost no FPs (ie, always less than 1 FP) but this was achieved by producing unacceptably low statistical power to detect any real effects of statin therapy. Among the three FWER methods, Holm and Hochberg resulted in identical statistical properties with Hommel resulting in a minor power advantage (compared with the other two methods) in very rare cases.

The use of MHT methods controlling the FDR (ie, Benjamini & Hochberg and Mehrotra & Adewale) were shown to be the most appropriate to use in these meta-analyses since these methods resulted in a low number of FPs in all simulated scenarios and, in parallel, maintained reasonable statistical power to detect any large effects of statin therapy. In addition, both of these FDR methods produced very similar results, with the Mehrotra & Adewale method considered the most suitable to use in the meta-analyses of the CTT AE database because on various occasions it was marginally more powerful than the Benjamini & Hochberg approach.

With regards to the second aim of the simulation studies (ie, the selection of the appropriate types of AEs to consider for (meta-) analyses), for AEs with a very low number of events (ie, less than 50 events), the expected power was very low to detect any real effects of statin, even if such effect were of great magnitude (ie, RR equal to 3.0). In addition, the decision to not test such rare AEs was shown to increase the statistical power of detecting real effects (of statin) in all other AEs considered for testing by about 5%. Although it was desirable to test across the various hierarchical levels of MedDRA, tests at the higher levels (ie, HLT and HLGT) should be performed only if such hypotheses contribute significant extra information, for example, at least a 50% increase in the number of events.

4.4.4 Multiple testing framework in CTT

4.4.4.1 Recommendations on testing strategy

The following hierarchical multiple testing approach is proposed based on the results from simulation analyses presented above:

1. Test a PT if it reports at least 50 events. Based on simulation analyses (section 4.4.3), the expected power of AEs with less than 50 events to detect any real effects of statin was shown to be very low and the decision of not testing such AEs was shown to increase the statistical power of all other AEs selected for testing.

2. Test all HLTs where the number of events is at least 50 **and** where the number of events is at least 50% higher than for any PT underneath it.
3. Test all HLGTs where the number of events is at least 50 **and** where the number of events is at least 50% higher than for any HLT or PT underneath it.
4. Test all SOCs[j].

4.4.4.2 Statistical properties of Mehrotra & Adewale (FDR) method

As shown above, the statistical properties of the various MHT approaches depends on the correlation by construction among the hypotheses which are tested within a family (see sections 4.4.2 and 4.4.3). In order to calculate the best estimates for the distribution of FPs and statistical power, the performance of the most powerful methods identified from simulation analysis (sections 4.4.2 and 4.4.3) controlling the three types of error rates (ie, no MHT correction (type-I), Hommel (FWER) and Mehrotra & Adewale (FDR)) is monitored when testing the family of hypotheses proposed above (see section 4.4.4.1).

Table 4.4 presents the distribution of FPs under four simulated scenarios and when using three approaches to MHT: i) no correction, ii) Hommel (FWER) and iii) Mehrotra & Adewale (FDR) method. The proposed testing strategy resulted in a total of 1,027 hypotheses being tested across the hierarchical levels of MedDRA. (Note that when performing the meta-analyses of trials of statin vs placebo in chapter 5, the number of hypotheses examined would be expected to be more due to using the complete AE dataset.) The distribution of FPs for the Mehrotra & Adewale (FDR) method increased with the number of true effects of statin (section 4.4.3.1) with the expected number of FPs equal to 0.5, 1.5, 3.7 and 6.1 when assuming true effects of statins in 0%, 1%, 5% and 10% of the AEs respectively.

[j] Although SOC were not being tested when performing the simulation analyses, these terms might result in effects of statin therapy in general conditions as SOCs contain disorders by body system (ie, eye disorders), relate to aetiology (ie, infections and infestations), or serve a classification purpose (ie, surgical and medical procedures). In addition, there are only 27 SOCs and such a low number of extra tests is not expected to influence the statistical power of the family of AEs being tested.

The Mehrotra & Adewale (FDR) method is shown to substantially reduce the number of FPs compared with the approach of no MHT correction but did not eliminate them in the same way as the Hommel (FWER) method did, in the extreme situation when true effects of statin were assumed to exist in a large proportion of AEs (**Table 4.4**).

Figure 4.7 presents the breakdown of the expected statistical power by number of events in each test when applying the three MHT methods under the simulated scenario when the null hypothesis (ie, no effect) was assumed randomly in 95% of tests and the alternative hypothesis (ie, true effect exist) was assumed in 5% of tests and the multiple testing strategy proposed in section 4.4.4.1. For very rare AEs (ie, with between 50 and 99 events representing an actual risk <0.08% in trials of statin vs placebo), the Mehrotra & Adewale method produced sufficient power (ie, >80%) to detect very large effects (of statin) of 200% increase in risk (ie, RR≥3.0), while for relatively rare AEs (ie, with between 100 and 499 events and risk from 0.08% to 0.42%) this method had >80% power to detect large effects of statin of more than 70% increase in risk (ie, RR> 1.7) (**Figure 4.7**). In addition, for AEs with between 500 and 999 events (ie, risk from 0.42% to 0.83%), the Mehrotra & Adewale method resulted in good power to detect moderate effects of about 35% increase in risk (ie, RR≥ 1.35), while for more frequent AEs (ie, >1000 events, risk >0.83%) this method had sufficient power to detect mild effects of statin of 15% increase in risk (ie, RR≥ 1.15). The Mehrotra & Adewale method was shown to be more powerful than the Hommel method but less powerful than the approach of not correcting for MHT (**Figure 4.7**).

In summary, although an approach of not correcting for MHT would result in greater power of detecting true effects of statin compared with the Mehrotra & Adewale method, such an approach is extremely undesirable because it would result in a very large number of false positive results (and it would be impossible to know which, if any, were real and which were false). Also, while Hommel's (FWER) method resulted in marginally better control of the number of FPs compared with Mehrotra & Adewale (FDR), its power to detect true effects of statin was substantially lower. Hence, the Mehrotra & Adewale (FDR) is proposed to be the

most appropriate MHT method to use when performing the CTT meta-analysis of all AEs in trials of statin vs placebo (chapter 5) and more vs less statin trials (chapter 6).

Table 4.1: Trials of statin vs placebo: Distribution of the total number of events across the three analysable MedDRA hierarchical levels, organized alphabetically by System Organ Class

	Number of tests at each sublevel with total number of events between <50| 50-99|100-999|>1000		
System Organ Class	**Higher Level Group Term**	**Higher Level Term**	**Preferred Term**
Blood and lymphatic system disorders	0|3|2|1	14|4|2|1	66|3|2|1
Cardiac disorders	2|0|4|3	11|2|7|6	103|17|19|7
Congenital, familial and genetic disorders	8|1|1|0	29|1|1|0	56|1|1|0
Ear and labyrinth disorders	0|1|2|2	6|1|3|2	34|3|7|1
Endocrine disorders	4|0|2|0	17|0|4|0	39|3|3|0
Eye disorders	2|0|6|3	20|8|11|2	159|11|13|2
Gastrointestinal disorders	0|0|11|7	41|6|29|8	297|21|35|9
General disorders and administration site conditions	0|0|3|4	15|2|5|7	122|9|12|11
Hepatobiliary disorders	0|0|2|1	4|4|5|0	64|3|5|0
Immune system disorders	1|0|1|0	5|0|4|0	30|3|2|0
Infections and infestations	7|0|0|4	64|4|10|12	436|29|39|14
Injury, poisoning and procedural complications	1|0|1|3	29|4|13|4	268|20|34|2
Investigations	3|1|8|6	45|5|22|6	388|19|36|4
Metabolism and nutrition disorders	5|0|3|4	17|7|10|2	65|13|12|2
Musculoskeletal and connective tissue disorders	1|1|2|5	8|3|15|8	140|12|28|11
Neoplasms benign, malignant and unspecified (including cysts and polyps)	13|4|7|3	106|8|8|3	348|10|16|0
Nervous system disorders	7|0|6|4	33|5|15|6	241|16|25|6
Pregnancy, puerperium and perinatal conditions	3|0|0|0	3|0|0|0	3|0|0|0
Product issues	0|1|0|0	2|0|0|0	7|0|0|0
Psychiatric disorders	9|3|4|3	35|6|6|4	122|7|5|4
Renal and urinary disorders	1|1|3|2	14|2|8|3	122|7|18|1
Reproductive system and breast disorders	2|2|7|2	20|3|11|1	142|8|11|1
Respiratory, thoracic and mediastinal disorders	2|0|3|4	9|5|9|5	161|14|27|3
Skin and subcutaneous tissue disorders	1|0|4|3	24|4|14|4	164|9|22|3
Social circumstances	4|0|0|0	12|0|0|0	27|0|0|0
Surgical and medical procedures	4|2|8|2	66|5|22|2	307|10|20|2
Vascular disorders	0|0|5|4	10|1|14|3	101|6|21|2
Total number of tests	**80|20|95|70**	**659|90|248|89**	**4012|254|413|86**

MedDRA V20.0 was used to code and analyse the adverse event data.
System organ class is shown in alphabetical order.
Data freeze as of in 8 Jan 2020 was used for simulation analyses.

Table 4.2 (a): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 6 events are tested across the MedDRA tree resulting in a total of 2,963 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Number of False Positives; mean (95% CI)†
100% H_0, 0% H_1	No MHT correction (Type-I)	0	143.9 (117 - 172)
	Holm (FWER)		0.4 (0 - 2)
	Hochberg (FWER)		0.4 (0 - 2)
	Hommel (FWER)		0.4 (0 - 2)
	Benjamini & Hochberg (FDR)		0.4 (0 - 2)
	Mehrotra & Adewale (FDR)		0.5 (0 - 3)
99% H_0, 1% H_1	No MHT correction (Type-I)	29	143.0 (119 - 171)
	Holm (FWER)		0.4 (0 - 2)
	Hochberg (FWER)		0.4 (0 - 2)
	Hommel (FWER)		0.4 (0 - 2)
	Benjamini & Hochberg (FDR)		1.4 (0 - 5)
	Mehrotra & Adewale (FDR)		1.7 (0 - 6)
95% H_0, 5% H_1	No MHT correction (Type-I)	148	137.1 (112 - 166)
	Holm (FWER)		0.4 (0 - 2)
	Hochberg (FWER)		0.4 (0 - 2)
	Hommel (FWER)		0.4 (0 - 2)
	Benjamini & Hochberg (FDR)		5.0 (1 - 11)
	Mehrotra & Adewale (FDR)		5.1 (1 - 11)
90% H_0, 10% H_1	No MHT correction (Type-I)	296	130.1 (107 - 156)
	Holm (FWER)		0.3 (0 - 2)
	Hochberg (FWER)		0.3 (0 - 2)
	Hommel (FWER)		0.3 (0 - 2)
	Benjamini & Hochberg (FDR)		9.8 (3 - 18)
	Mehrotra & Adewale (FDR)		9.9 (3 - 18)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR=10.00 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.2 (b): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 50 events were tested across the MedDRA tree resulting in a total of 1,325 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Number of False Positives; mean (95% CI)†
100% H_0, 0% H_1	No MHT correction (Type-I)	0	72.0 (52 - 93)
	Holm (FWER)		0.5 (0 - 2)
	Hochberg (FWER)		0.5 (0 - 2)
	Hommel (FWER)		0.5 (0 - 2)
	Benjamini & Hochberg (FDR)		0.6 (0 - 3)
	Mehrotra & Adewale (FDR)		0.7 (0 - 4)
99% H_0, 1% H_1	No MHT correction (Type-I)	13	71.6 (53 - 95)
	Holm (FWER)		0.5 (0 - 2)
	Hochberg (FWER)		0.5 (0 - 2)
	Hommel (FWER)		0.5 (0 - 2)
	Benjamini & Hochberg (FDR)		1.6 (0 - 5)
	Mehrotra & Adewale (FDR)		2.0 (0 - 6)
95% H_0, 5% H_1	No MHT correction (Type-I)	67	68.8 (50 - 91)
	Holm (FWER)		0.5 (0 - 2)
	Hochberg (FWER)		0.5 (0 - 2)
	Hommel (FWER)		0.5 (0 - 2)
	Benjamini & Hochberg (FDR)		4.8 (1 - 11)
	Mehrotra & Adewale (FDR)		5.0 (1 - 11)
90% H_0, 10% H_1	No MHT correction (Type-I)	132	65.1 (48 - 86)
	Holm (FWER)		0.4 (0 - 2)
	Hochberg (FWER)		0.4 (0 - 2)
	Hommel (FWER)		0.4 (0 - 2)
	Benjamini & Hochberg (FDR)		8.1 (2 - 16)
	Mehrotra & Adewale (FDR)		8.2 (2 - 16)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR=10.00 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.2 (c): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 100 events were tested across the MedDRA tree resulting in a total of 977 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Number of False Positives; mean (95% CI)†
100% H_0, 0% H_1	No MHT correction (Type-I)	0	54.5 (37 - 74)
	Holm (FWER)		0.6 (0 - 3)
	Hochberg (FWER)		0.6 (0 - 3)
	Hommel (FWER)		0.6 (0 - 3)
	Benjamini & Hochberg (FDR)		0.6 (0 - 4)
	Mehrotra & Adewale (FDR)		0.8 (0 - 5)
99% H_0, 1% H_1	No MHT correction (Type-I)	10	54.2 (38 - 75)
	Holm (FWER)		0.5 (0 - 2)
	Hochberg (FWER)		0.5 (0 - 2)
	Hommel (FWER)		0.5 (0 - 2)
	Benjamini & Hochberg (FDR)		1.5 (0 - 5)
	Mehrotra & Adewale (FDR)		2.0 (0 - 6)
95% H_0, 5% H_1	No MHT correction (Type-I)	49	52.1 (36 - 72)
	Holm (FWER)		0.5 (0 - 2)
	Hochberg (FWER)		0.5 (0 - 2)
	Hommel (FWER)		0.5 (0 - 2)
	Benjamini & Hochberg (FDR)		4.0 (0 - 10)
	Mehrotra & Adewale (FDR)		4.2 (0 - 10)
90% H_0, 10% H_1	No MHT correction (Type-I)	98	49.3 (34 - 69)
	Holm (FWER)		0.5 (0 - 2)
	Hochberg (FWER)		0.5 (0 - 2)
	Hommel (FWER)		0.5 (0 - 2)
	Benjamini & Hochberg (FDR)		6.5 (1 - 14)
	Mehrotra & Adewale (FDR)		6.6 (1 - 14)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR=10.00 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.2 (d): Distribution of FPs for various simulated scenarios when applying various MHT methods; AEs with at least 1000 events were tested across the MedDRA tree resulting in a total of 238 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Number of False Positives; mean (95% CI)†
100% H_0, 0% H_1	No MHT correction (Type-I)	0	15.2 (6 - 26)
	Holm (FWER)		0.2 (0 - 3)
	Hochberg (FWER)		0.2 (0 - 3)
	Hommel (FWER)		0.2 (0 - 3)
	Benjamini & Hochberg (FDR)		0.3 (0 - 3)
	Mehrotra & Adewale (FDR)		0.4 (0 - 4)
99% H_0, 1% H_1	No MHT correction (Type-I)	2	15.1 (7 - 26)
	Holm (FWER)		0.2 (0 - 3)
	Hochberg (FWER)		0.2 (0 - 3)
	Hommel (FWER)		0.2 (0 - 3)
	Benjamini & Hochberg (FDR)		0.5 (0 - 4)
	Mehrotra & Adewale (FDR)		1.0 (0 - 5)
95% H_0, 5% H_1	No MHT correction (Type-I)	12	14.5 (6 - 25)
	Holm (FWER)		0.2 (0 - 2)
	Hochberg (FWER)		0.2 (0 - 2)
	Hommel (FWER)		0.2 (0 - 2)
	Benjamini & Hochberg (FDR)		1.4 (0 - 6)
	Mehrotra & Adewale (FDR)		1.8 (0 - 7)
90% H_0, 10% H_1	No MHT correction (Type-I)	24	13.7 (6 - 23)
	Holm (FWER)		0.2 (0 - 2)
	Hochberg (FWER)		0.2 (0 - 2)
	Hommel (FWER)		0.2 (0 - 2)
	Benjamini & Hochberg (FDR)		2.2 (0 - 7)
	Mehrotra & Adewale (FDR)		2.5 (0 - 8)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR=10.00 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.3 (a): Statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 6 events are tested across the MedDRA tree resulting in a total of 2,963 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Statistical power (%), mean (95% CI)†; Relative Risk					
			1.1	1.25	1.5	2.0	3.0	5.0
99% H_0, 1% H_1	No MHT correction (Type-I)	29	12 (0 - 25)	25 (10 - 42)	40 (23 - 59)	57 (40 - 74)	72 (55 - 88)	84 (71 - 96)
	Holm (FWER)		1 (0 - 7)	6 (0 - 16)	15 (3 - 29)	27 (12 - 43)	39 (22 - 57)	49 (32 - 66)
	Hochberg (FWER)		1 (0 - 7)	6 (0 - 16)	15 (3 - 29)	27 (12 - 43)	39 (22 - 57)	49 (32 - 66)
	Hommel (FWER)		1 (0 - 7)	6 (0 - 16)	15 (3 - 29)	27 (12 - 43)	39 (22 - 57)	49 (32 - 66)
	Benjamini & Hochberg (FDR)		1 (0 - 7)	7 (0 - 19)	17 (4 - 32)	31 (14 - 48)	45 (26 - 62)	56 (38 - 73)
	Mehrotra & Adewale (FDR)		1 (0 - 7)	7 (0 - 19)	18 (5 - 33)	31 (14 - 50)	45 (27 - 64)	57 (38 - 74)
95% H_0, 5% H_1	No MHT correction (Type-I)	148	12 (7 - 18)	26 (18 - 33)	40 (33 - 48)	57 (49 - 65)	73 (66 - 79)	84 (79 - 90)
	Holm (FWER)		1 (0 - 3)	7 (3 - 11)	15 (10 - 21)	27 (20 - 34)	39 (31 - 47)	49 (41 - 57)
	Hochberg (FWER)		1 (0 - 3)	7 (3 - 11)	15 (10 - 21)	27 (20 - 34)	39 (31 - 47)	49 (41 - 57)
	Hommel (FWER)		1 (0 - 3)	7 (3 - 11)	15 (10 - 21)	27 (20 - 34)	39 (31 - 47)	49 (41 - 57)
	Benjamini & Hochberg (FDR)		1 (0 - 4)	9 (4 - 14)	20 (13 - 27)	35 (28 - 43)	51 (42 - 59)	63 (54 - 72)
	Mehrotra & Adewale (FDR)		1 (0 - 4)	9 (4 - 14)	20 (14 - 28)	35 (28 - 43)	51 (42 - 59)	63 (54 - 72)
90% H_0, 10% H_1	No MHT correction (Type-I)	296	12 (9 - 16)	26 (21 - 30)	40 (35 - 46)	58 (52 - 63)	73 (68 - 78)	84 (80 - 88)
	Holm (FWER)		1 (0 - 2)	7 (4 - 9)	15 (12 - 19)	26 (22 - 31)	39 (34 - 44)	49 (44 - 54)
	Hochberg (FWER)		1 (0 - 2)	7 (4 - 9)	15 (12 - 19)	26 (22 - 31)	39 (34 - 44)	49 (44 - 54)
	Hommel (FWER)		1 (0 - 2)	7 (4 - 9)	15 (12 - 19)	26 (22 - 31)	39 (34 - 44)	49 (44 - 54)
	Benjamini & Hochberg (FDR)		2 (0 - 3)	10 (6 - 14)	22 (17 - 27)	38 (32 - 44)	54 (47 - 60)	67 (61 - 72)
	Mehrotra & Adewale (FDR)		2 (0 - 3)	10 (6 - 14)	22 (17 - 27)	38 (32 - 44)	54 (47 - 60)	67 (61 - 72)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR of 1.2, 1.5, 2.0, 5.0 and 10.0 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.3 (b): Simulated statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 50 events were tested across the MedDRA tree resulting in a total of 1,325 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Statistical power (%), mean (95% CI)†; Relative Risk					
			1.1	1.25	1.5	2.0	3.0	5.0
99% H_0, 1% H_1	No MHT correction (Type-I)	13	21 (0 - 47)	48 (22 - 75)	75 (50 - 100)	94 (78 - 100)	100 (94 - 100)	100 (100 - 100)
	Holm (FWER)		3 (0 - 15)	16 (0 - 40)	35 (9 - 64)	62 (35 - 89)	87 (62 - 100)	98 (88 - 100)
	Hochberg (FWER)		3 (0 - 15)	16 (0 - 40)	35 (9 - 64)	62 (35 - 89)	87 (62 - 100)	98 (88 - 100)
	Hommel (FWER)		3 (0 - 15)	16 (0 - 40)	35 (9 - 64)	62 (35 - 89)	87 (62 - 100)	98 (88 - 100)
	Benjamini & Hochberg (FDR)		3 (0 - 17)	18 (0 - 44)	41 (12 - 71)	72 (43 - 94)	94 (75 - 100)	100 (92 - 100)
	Mehrotra & Adewale (FDR)		3 (0 - 18)	19 (0 - 44)	41 (11 - 73)	72 (43 - 100)	94 (75 - 100)	100 (93 - 100)
95% H_0, 5% H_1	No MHT correction (Type-I)	67	21 (11 - 31)	48 (37 - 59)	75 (64 - 84)	94 (87 - 100)	100 (98 - 100)	100 (100 - 100)
	Holm (FWER)		3 (0 - 8)	16 (7 - 26)	36 (25 - 47)	62 (50 - 73)	87 (78 - 94)	98 (95 - 100)
	Hochberg (FWER)		3 (0 - 8)	16 (7 - 26)	36 (25 - 47)	62 (50 - 73)	87 (78 - 94)	98 (95 - 100)
	Hommel (FWER)		3 (0 - 8)	16 (7 - 26)	36 (25 - 47)	62 (50 - 73)	87 (78 - 94)	98 (95 - 100)
	Benjamini & Hochberg (FDR)		4 (0 - 10)	22 (11 - 34)	49 (35 - 61)	79 (68 - 90)	97 (92 - 100)	100 (98 - 100)
	Mehrotra & Adewale (FDR)		4 (0 - 10)	23 (12 - 34)	49 (36 - 62)	79 (68 - 90)	97 (92 - 100)	100 (98 - 100)
90% H_0, 10% H_1	No MHT correction (Type-I)	132	21 (14 - 29)	49 (40 - 57)	74 (67 - 82)	94 (90 - 98)	100 (98 - 100)	100 (100 - 100)
	Holm (FWER)		3 (1 - 6)	16 (10 - 22)	36 (28 - 44)	62 (53 - 70)	87 (81 - 92)	98 (96 - 100)
	Hochberg (FWER)		3 (1 - 6)	16 (10 - 22)	36 (28 - 44)	62 (53 - 70)	87 (81 - 92)	98 (96 - 100)
	Hommel (FWER)		3 (1 - 6)	16 (10 - 22)	36 (28 - 44)	62 (53 - 70)	87 (81 - 92)	98 (96 - 100)
	Benjamini & Hochberg (FDR)		4 (1 - 9)	25 (16 - 34)	53 (43 - 63)	83 (76 - 89)	98 (95 - 100)	100 (99 - 100)
	Mehrotra & Adewale (FDR)		5 (1 - 9)	25 (16 - 34)	53 (43 - 63)	83 (76 - 89)	98 (95 - 100)	100 (99 - 100)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR of 1.2, 1.5, 2.0, 5.0 and 10.0 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.3 (c): Simulated statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 100 events were tested across the MedDRA tree resulting in a total of 977 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Statistical power (%), mean (95% CI)†; Relative Risk					
			1.1	1.25	1.5	2.0	3.0	5.0
99% H_0, 1% H_1	No MHT correction (Type-I)	10	27 (0 - 58)	59 (25 - 89)	88 (60 - 100)	99 (89 - 100)	100 (100 - 100)	100 (100 - 100)
	Holm (FWER)		4 (0 - 20)	22 (0 - 50)	49 (17 - 82)	81 (50 - 100)	99 (88 - 100)	100 (100 - 100)
	Hochberg (FWER)		4 (0 - 20)	22 (0 - 50)	49 (17 - 82)	81 (50 - 100)	99 (88 - 100)	100 (100 - 100)
	Hommel (FWER)		4 (0 - 20)	22 (0 - 50)	49 (17 - 82)	81 (50 - 100)	99 (88 - 100)	100 (100 - 100)
	Benjamini & Hochberg (FDR)		5 (0 - 22)	26 (0 - 58)	57 (17 - 89)	89 (62 - 100)	100 (93 - 100)	100 (100 - 100)
	Mehrotra & Adewale (FDR)		5 (0 - 23)	26 (0 - 60)	58 (20 - 89)	90 (64 - 100)	100 (93 - 100)	100 (100 - 100)
95% H_0, 5% H_1	No MHT correction (Type-I)	49	26 (14 - 40)	60 (45 - 73)	87 (76 - 96)	99 (96 - 100)	100 (100 - 100)	100 (100 - 100)
	Holm (FWER)		4 (0 - 11)	22 (11 - 35)	50 (35 - 64)	81 (70 - 91)	99 (95 - 100)	100 (100 - 100)
	Hochberg (FWER)		4 (0 - 11)	22 (11 - 35)	50 (35 - 64)	81 (70 - 91)	99 (95 - 100)	100 (100 - 100)
	Hommel (FWER)		4 (0 - 11)	22 (11 - 35)	50 (35 - 64)	81 (70 - 91)	99 (95 - 100)	100 (100 - 100)
	Benjamini & Hochberg (FDR)		5 (0 - 14)	31 (16 - 46)	65 (50 - 80)	94 (87 - 100)	100 (98 - 100)	100 (100 - 100)
	Mehrotra & Adewale (FDR)		6 (0 - 14)	32 (18 - 47)	66 (50 - 80)	94 (87 - 100)	100 (98 - 100)	100 (100 - 100)
90% H_0, 10% H_1	No MHT correction (Type-I)	98	26 (18 - 36)	60 (50 - 70)	87 (80 - 93)	99 (97 - 100)	100 (100 - 100)	100 (100 - 100)
	Holm (FWER)		4 (1 - 9)	22 (14 - 30)	50 (40 - 60)	82 (74 - 89)	99 (96 - 100)	100 (100 - 100)
	Hochberg (FWER)		4 (1 - 9)	22 (14 - 30)	50 (40 - 60)	82 (74 - 89)	99 (96 - 100)	100 (100 - 100)
	Hommel (FWER)		4 (1 - 9)	22 (15 - 30)	50 (40 - 60)	82 (74 - 89)	99 (96 - 100)	100 (100 - 100)
	Benjamini & Hochberg (FDR)		6 (1 - 12)	35 (24 - 46)	70 (60 - 80)	96 (92 - 99)	100 (99 - 100)	100 (100 - 100)
	Mehrotra & Adewale (FDR)		7 (1 - 13)	36 (25 - 46)	70 (60 - 80)	96 (91 - 99)	100 (99 - 100)	100 (100 - 100)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR of 1.2, 1.5, 2.0, 5.0 and 10.0 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.3 (d): Simulated statistical power for various simulated scenarios when applying various MHT methods; AEs with at least 1000 events were tested across the MedDRA tree resulting in a total of 238 tests

Simulated scenario*	Multiple testing approach (α=5%)	True Effects	Statistical power (%), mean (95% CI)†; Relative Risk					
			1.1	1.25	1.5	2.0	3.0	5.0
99% H_0, 1% H_1	No MHT correction (Type-I)	2	66 (0 - 100)	98 (75 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Holm (FWER)		22 (0 - 100)	84 (0 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Hochberg (FWER)		22 (0 - 100)	84 (0 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Hommel (FWER)		22 (0 - 100)	84 (0 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Benjamini & Hochberg (FDR)		24 (0 - 100)	86 (0 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Mehrotra & Adewale (FDR)		23 (0 - 100)	87 (0 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
95% H_0, 5% H_1	No MHT correction (Type-I)	12	63 (30 - 89)	99 (89 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Holm (FWER)		21 (0 - 47)	83 (58 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Hochberg (FWER)		21 (0 - 47)	83 (58 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Hommel (FWER)		21 (0 - 47)	83 (58 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Benjamini & Hochberg (FDR)		28 (0 - 62)	93 (71 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Mehrotra & Adewale (FDR)		29 (0 - 64)	93 (73 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
90% H_0, 10% H_1	No MHT correction (Type-I)	24	64 (44 - 84)	99 (92 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Holm (FWER)		22 (6 - 39)	83 (67 - 96)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Hochberg (FWER)		22 (6 - 39)	83 (67 - 96)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Hommel (FWER)		22 (6 - 39)	83 (67 - 96)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Benjamini & Hochberg (FDR)		33 (11 - 56)	94 (80 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)
	Mehrotra & Adewale (FDR)		34 (11 - 58)	95 (83 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)	100 (100 - 100)

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR of 1.2, 1.5, 2.0, 5.0 and 10.0 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

†The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

Table 4.4: Distribution of FPs for various simulated scenarios when applying various MHT methods and the following multiple testing strategy resulting in 1,027 tests: (i) test any Preferred Term with at least 50 events and (ii) test any term higher up the MedDRA hierarchical tree which reports 50% extra information†

Simulated scenario*	Multiple testing approach (α=5%)	True Positives	Number of False Positives; mean (95% CI)‡
100% H_0, 0% H_1	No MHT correction (Type-I)	0	55.3 (40 - 72)
	Hommel (FWER)		0.4 (0 - 2)
	Mehrotra & Adewale (FDR)		0.5 (0 - 3)
99% H_0, 1% H_1	No MHT correction (Type-I)	10	55.2 (42 - 73)
	Hommel (FWER)		0.4 (0 - 1)
	Mehrotra & Adewale (FDR)		1.5 (0 - 4)
95% H_0, 5% H_1	No MHT correction (Type-I)	51	52.9 (39 - 70)
	Hommel (FWER)		0.4 (0 - 1)
	Mehrotra & Adewale (FDR)		3.7 (0 - 9)
90% H_0, 10% H_1	No MHT correction (Type-I)	102	50.2 (37 - 67)
	Hommel (FWER)		0.4 (0 - 1)
	Mehrotra & Adewale (FDR)		6.1 (1 - 13)

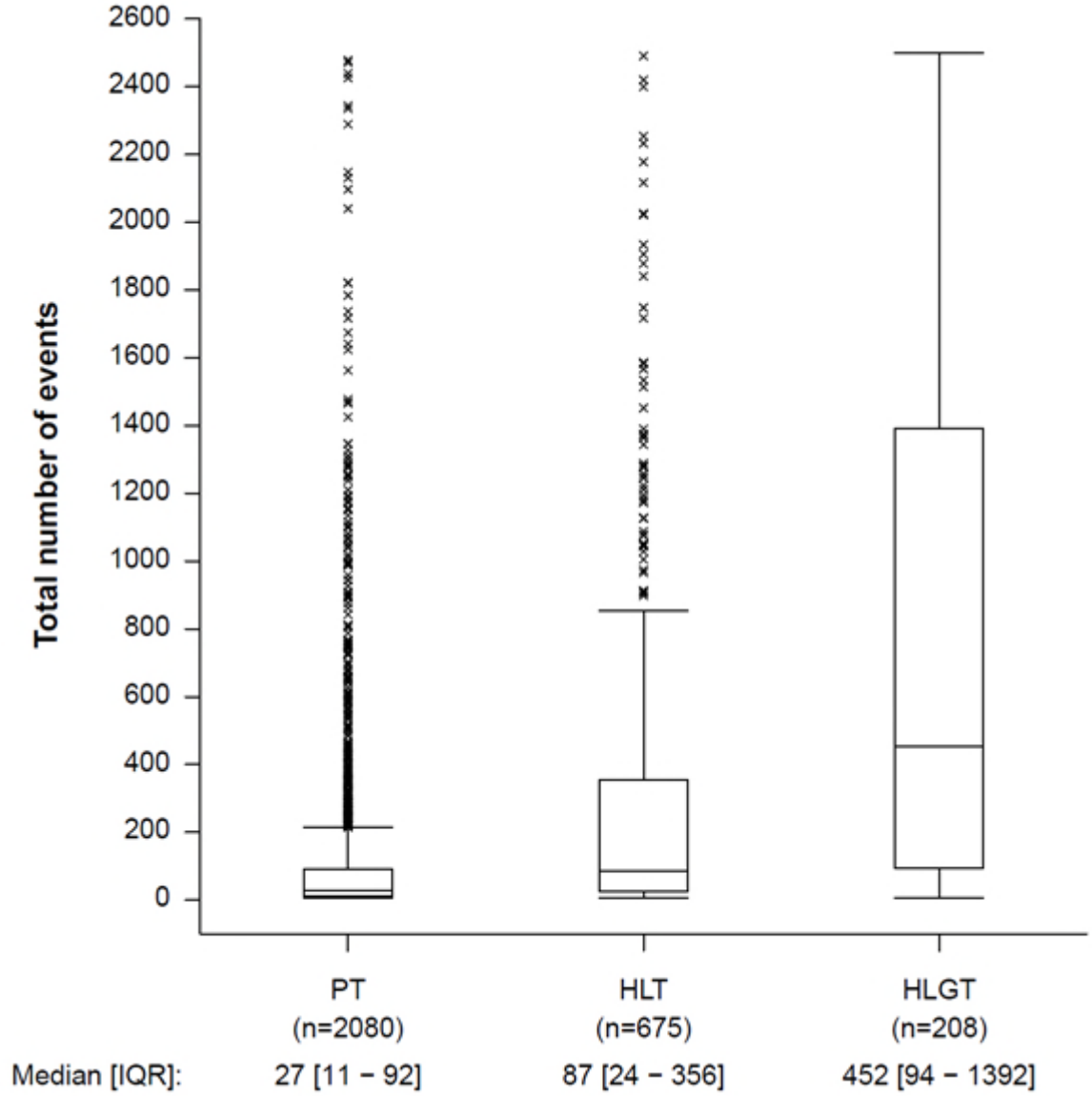
† Detailed description of the multiple testing strategy in three steps: (i) test a PT if it reports at least 50 events; (ii) Test a HLT with at least 50 events if no associated PT was being tested (ie, all PTs underneath the HLT reported less than 50 events). Otherwise, test a HLT if it reports at least 50% more events than any of the PTs underneath it; and (iii) Test a HLGT with at least 50 events if no HLT or PT underneath it is being tested. Otherwise, test a HLGT if it reports at least 50% more events than any of the HLTs or PTs underneath it.

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR=1.0 for H_0 and RR=10.00 for H_1 (a value of 10 is used to nearly ensure that such effects are currently declared as significant).

‡The Confidence Interval (CI) is taken as the 2.5th and 97.5th percentiles from the simulated results.

FDR=False Discovery Rate (FDR) and FWER=FamilyWise Error Rate.

Figure 4.1: Trials of statin vs placebo: Distribution of the total number of events across the MedDRA hierarchical levels



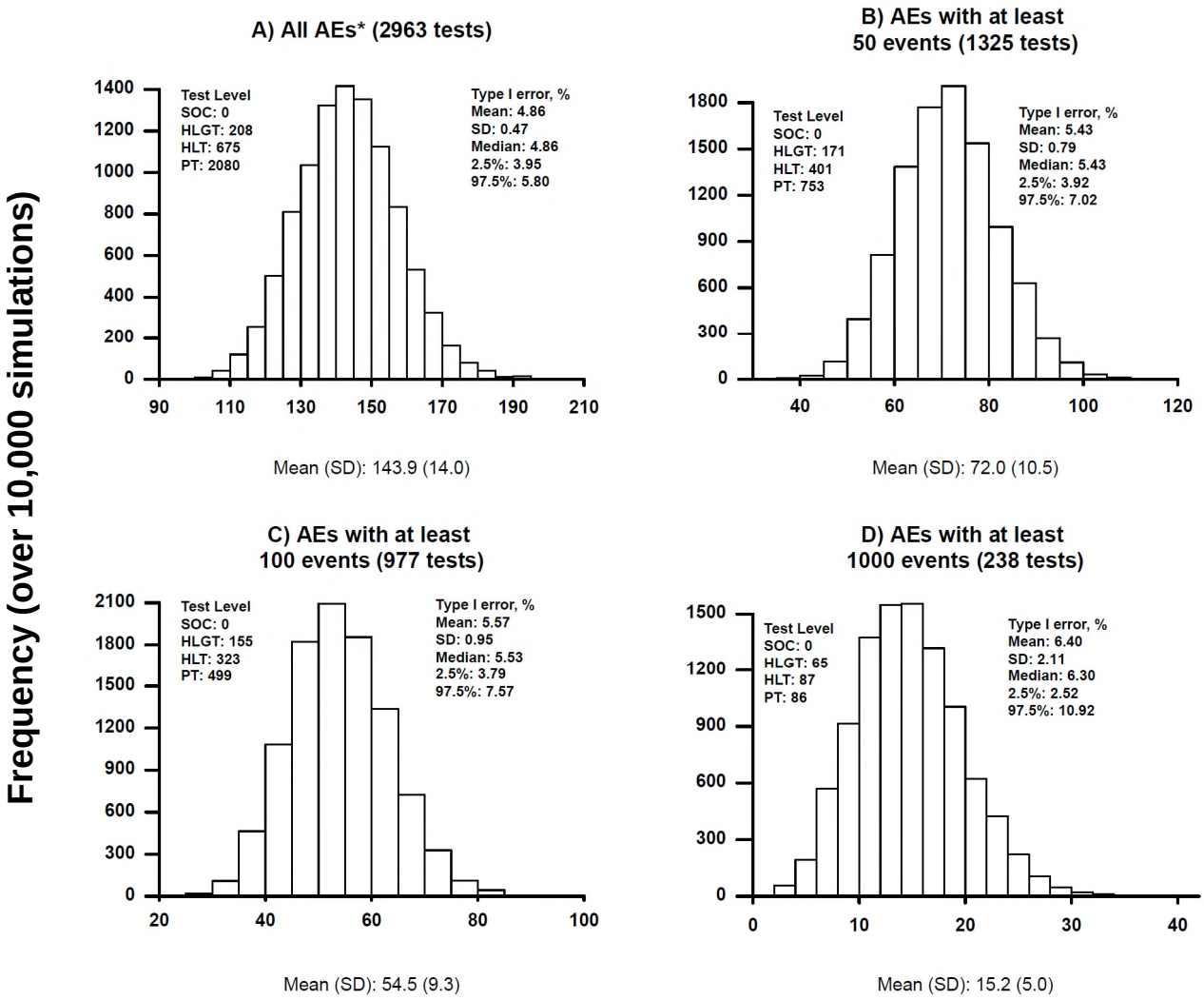
PT= Preferred Term, HLT= Higher Level Term and HLGT= Higher Level Group Term.

28 PTs, 37 HLTs and 29 HLGTs reported a total number of events greater than 2500 (data not shown in figure but incorporated when producing the boxplots).

Only MedDRA terms with potentially significant hypotheses are included in simulations (ie, AEs which involve a total of at least 6 events) as hypotheses involving fewer than 6 events can yield a minimum two-sided p-value of 0.0625.

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

Figure 4.2: Trials of statin vs placebo: under the null hypothesis, simulation of the distribution of false positives when no correction for multiple testing is applied (ie, p-value<0.05);



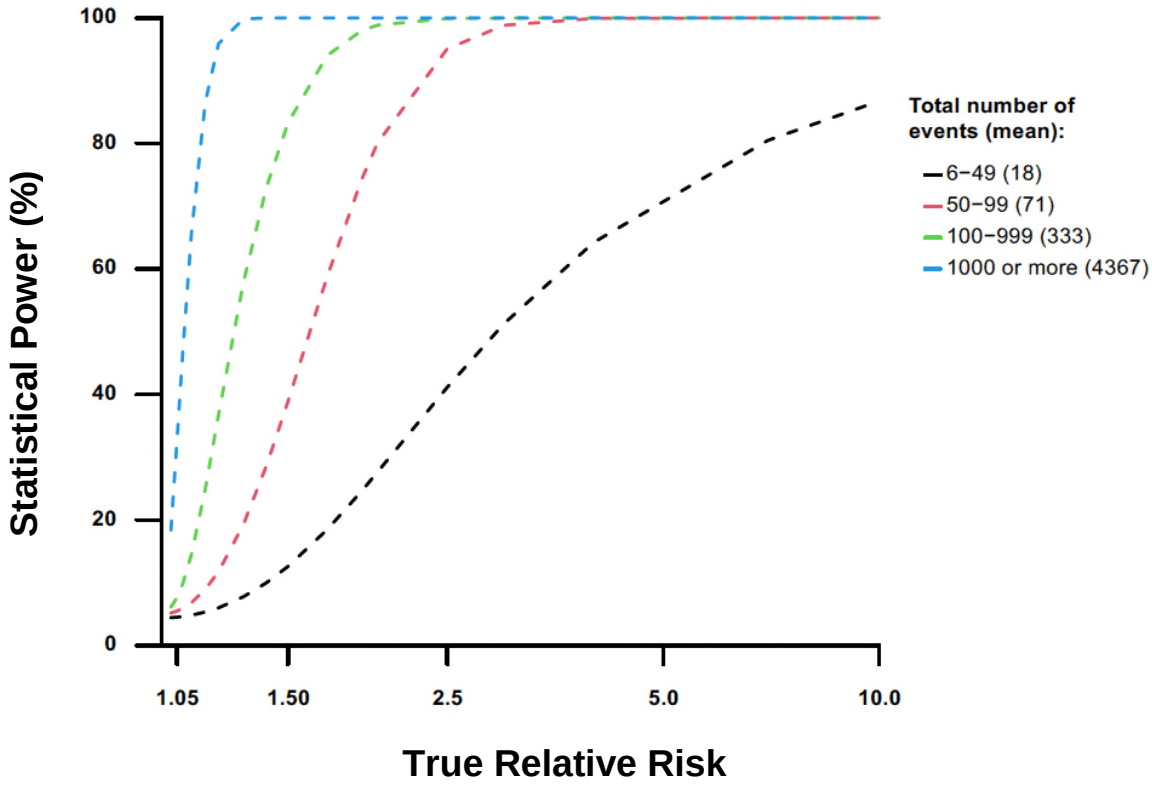
Results are produced based on 10,000 simulations.

Null hypotheses were generated assuming no treatment effect (ie, 'true' relative risk of 1.00).

*Only MedDRA terms with potentially significant hypotheses are included in simulations (ie, AEs which involve a total of at least 6 events) as hypotheses involving fewer than 6 events can yield a minimum two-sided p-value of 0.0625.

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

Figure 4.3: Expected power to reject the null hypothesis (at the 5% level) when statin therapy is assumed to increase the risk of adverse events by between 5% and 900% (ie, RR of between 1.05 and 10), and if no correction for multiple testing is applied



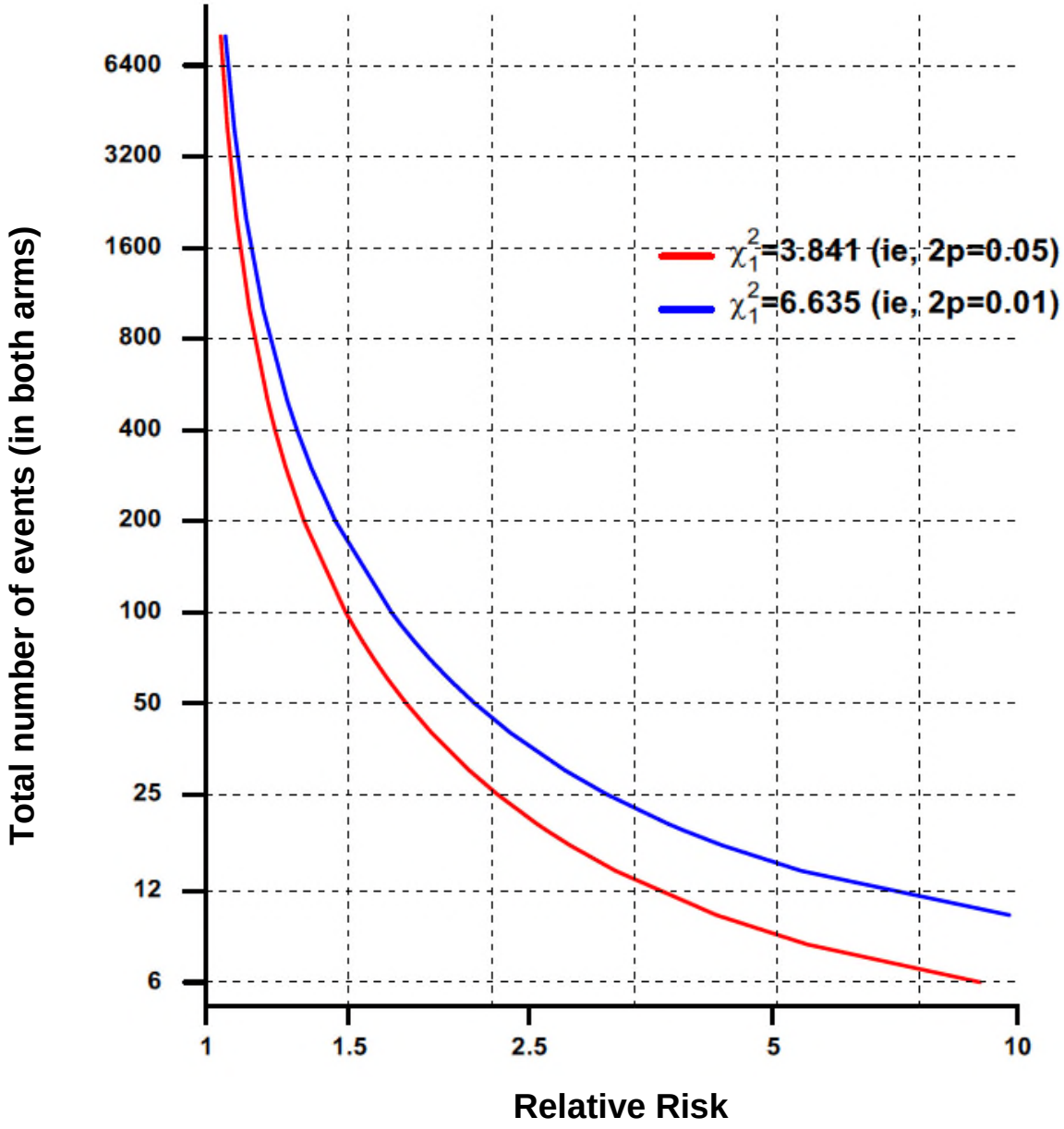
Results are produced based on 10,000 simulations.

Alternative hypotheses assuming real treatment effects were generated for 'true' relative risks of 1.03, 1.05, 1.07, 1.10, 1.15, 1.20, 1.25, 1.30, 1.40, 1.50, 1.70, 1.90, 2.00, 3.00, 4.00, 5.00, 7.00 and 10.0

Only MedDRA terms with potentially significant hypotheses are included in simulations (ie, AEs which involve a total of at least 6 events) as hypotheses involving fewer than 6 events can yield a minimum two-sided p-value of 0.0625.

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

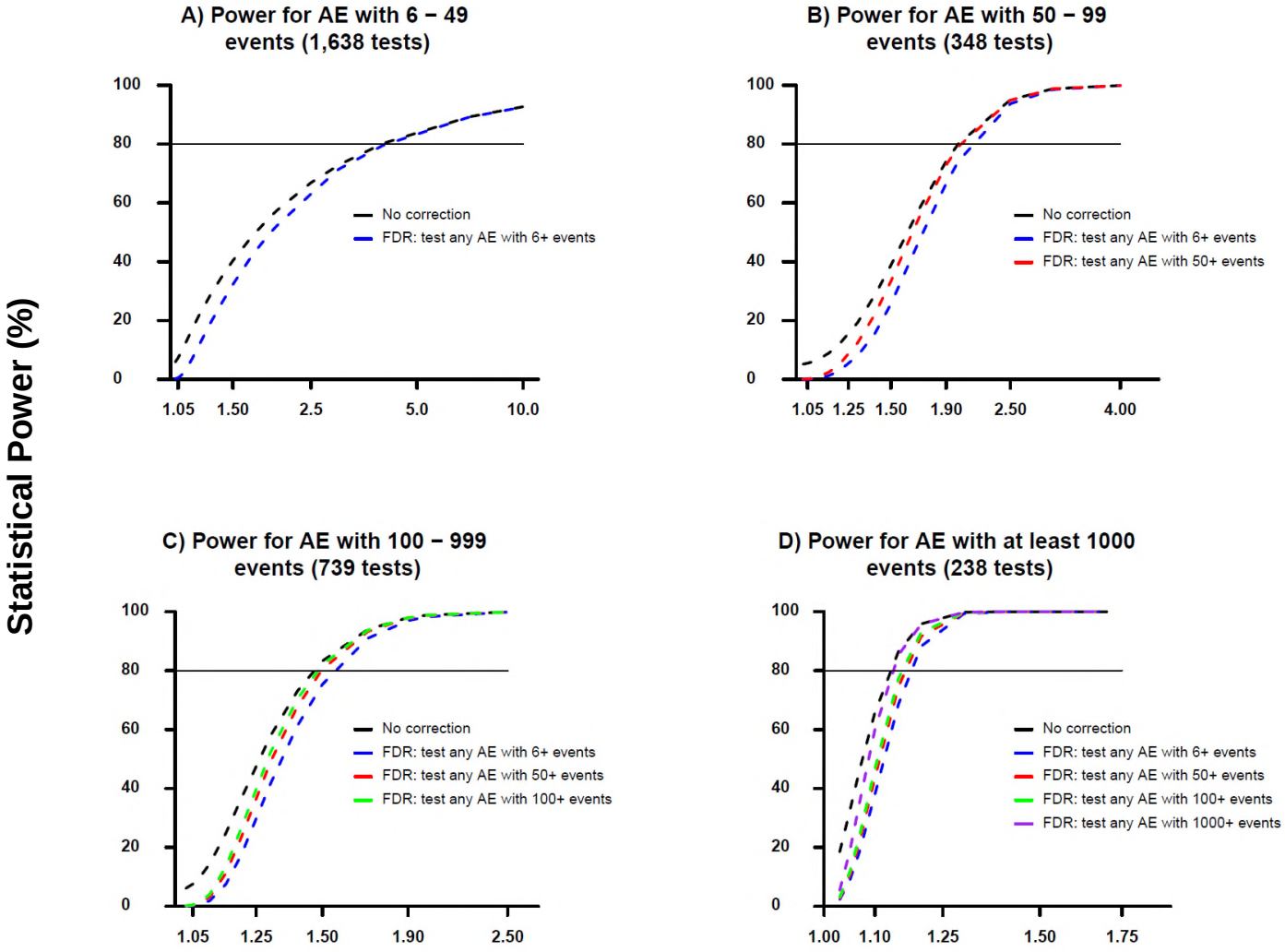
Figure 4.4: Relationship between the relative risk and the smallest number of events needed for such relative risk to be declared 'significant' at the 5% or 1% level



For 2x2 contingency tables, the observed (O) – expected (E) statistic is equal to $\mathrm{OE} = \mathrm{E}_1 - \frac{\mathrm{N}_1\mathrm{C}}{\mathrm{Z}}$ and its' associated variance (V) equals $\mathrm{V} = \frac{\frac{\mathrm{N}_1\mathrm{C}}{\mathrm{Z}} * (1 - \frac{\mathrm{N}_1}{\mathrm{Z}}) * (\mathrm{Z} - \mathrm{C})}{\mathrm{Z} - 1}$, where E_1 = number of events in statin arm, N_1 = number of participants in statin arm, Z = number of participants in both arms for a 1:1 randomisation (ie, 120,000 participants) and C = number of events in both arms. The chi-square test statistic is calculated as $\mathrm{x}_1^2 = \frac{\mathrm{OE}^2}{\mathrm{V}}$. Then, the Relative Risk (RR) (for example, of statin therapy) can be expressed, after appropriate mathematical calculations (see appendix I), using the following formula: $\mathrm{RR} = (\frac{1}{\frac{1}{2} + \frac{\sqrt{\mathrm{x}_1^2}}{2\mathrm{C}} \sqrt{\frac{\mathrm{C}(\mathrm{Z}-\mathrm{C})}{\mathrm{Z}-1}}} - 1)^{-1}$. For example, for Z=120,000 (ie, N_1=N_2=60,000) C=100 and x_1^2=3.841 (ie, 2p=0.05), the minimum RR detected as significant is equal to 1.49 (see **Appendix B** for details).

When performing time-to-event analysis, the minimum RR detected as significant at 2p=0.05 is expected to be equal or lower to the RR detected as significant using the OE statistic for 2x2 contingency tables

Figure 4.5: Trials of statin vs placebo: breakdown of expected power by number of events in each test, in the extreme situation whereby all adverse events are increased between 5% to 900% (ie, relative risk of 1.05 to 10.00); two statistical methods are applied for handling multiple testing: no correction and FDR correction



Results are produced based on 10,000 simulations and true relative risk of: 1.03, 1.05, 1.10, 1.20, 1.30, 1.40, 1.50, 1.70, 1.90, 2.00, 2.50, 3.00, 4.00, 5.00, 7.00 and 10.00.

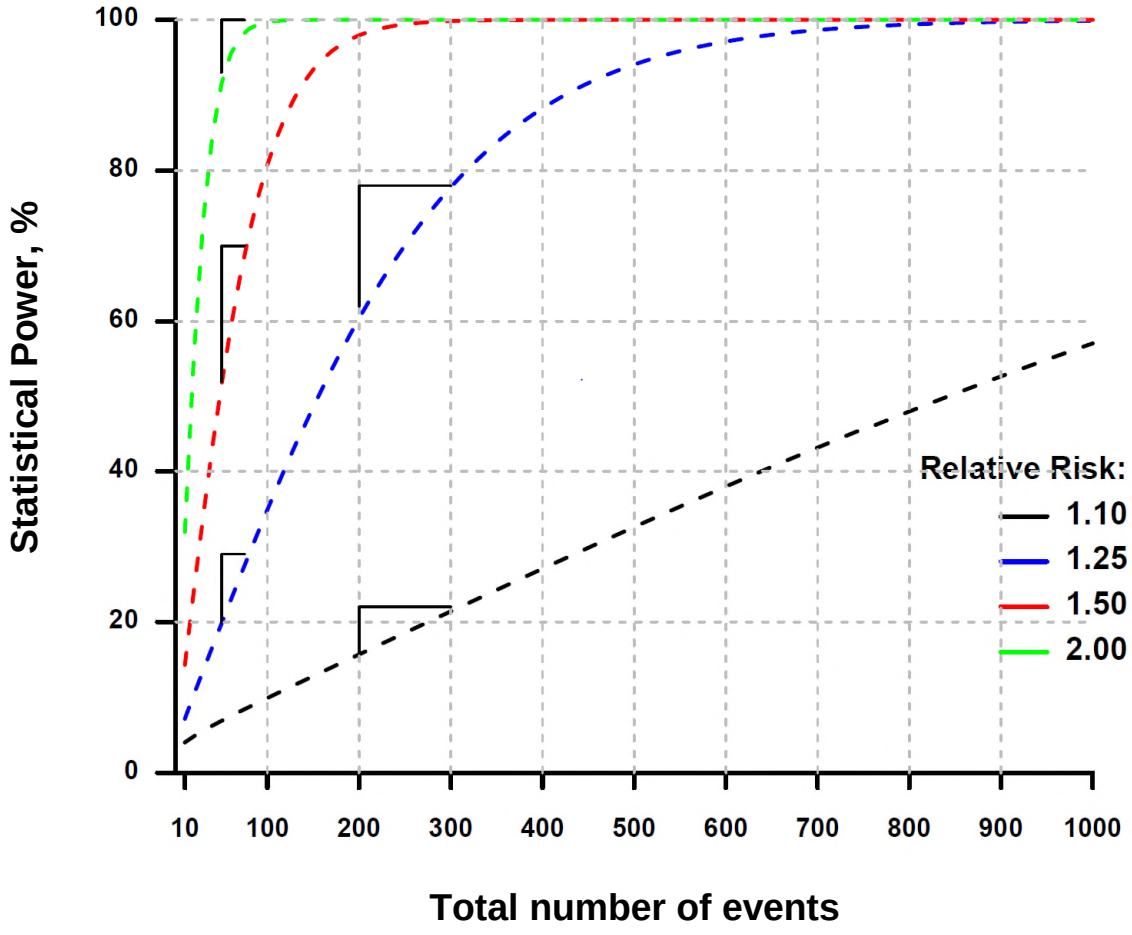
The Mehrotra & Adewale method was used to correct for multiple testing and control of the False Discovery Rate (FDR).

Only MedDRA terms with potentially significant hypotheses are included in simulations (ie, AEs which involve a total of at least 6 events) as hypotheses involving fewer than 6 events can yield a minimum two-sided p-value of 0.0625.

System Organ Class terms were not analysed, while terms across the two levels of MedDRA higher than the preferred term were tested only if they reported at least one extra event.

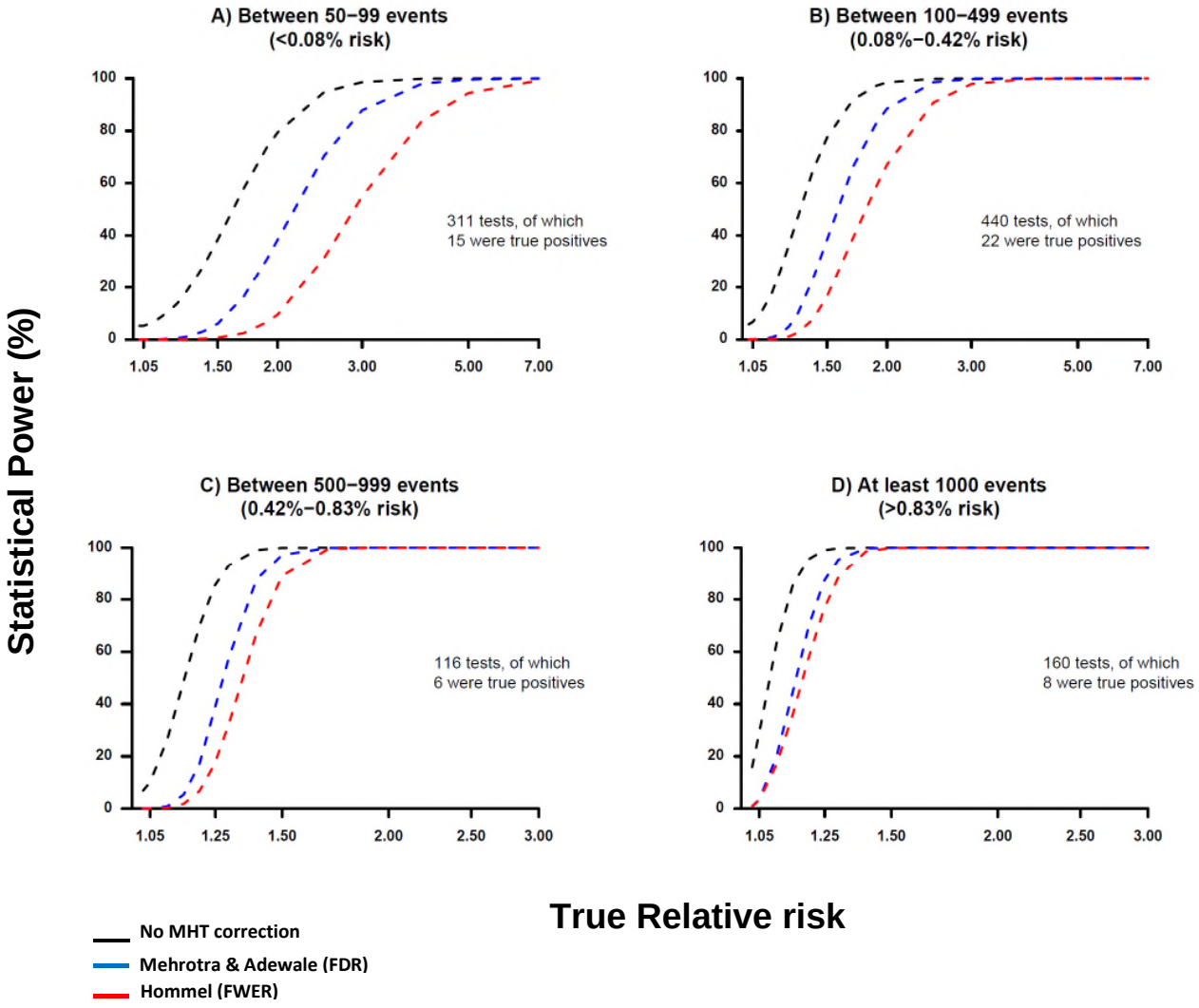
Methods correcting for multiple testing are designed to account for the number of hypotheses which need to be tested jointly, known as a 'family' of hypotheses, because this number is directly associated with the appropriate threshold used to determine 'statistical significance' (ie, a lower significance threshold is needed for a 'family' with a large number of hypotheses).

Figure 4.6: Trials of statin vs placebo: Mathematical relationship between the total number of events and the statistical power for detecting an increase in risk of 10%, 25%, 50% and 100% (ie, RR of 1.10, 1.25, 1.50 and 2.0)



Statistical power was calculated for test between proportions at 2p=0.05 and when total number of participants (in both arms for a 1:1 randomisation) was equal to 120,000.

Figure 4.7: Trials of statin vs placebo: breakdown of expected power by number of events in each test when applying various MHT methods and the following multiple testing strategy (resulting in 1,027 tests): (i) test any Preferred Term with at least 50 events and (ii) test any term higher up the MedDRA hierarchical tree which reported 50% extra information†; simulated scenario: null hypothesis (ie, no effect) assumed randomly in 95% of tests and alternative hypothesis (ie, true effect exist) assumed in 5% of tests



† Detailed description of the multiple testing strategy in three steps: (i) test a PT if it reports at least 50 events; (ii) Test a HLT with at least 50 events if no associated PT was being tested (ie, all PTs underneath the HLT reported less than 50 events). Otherwise, test a HLT if it reports at least 50% extra events as in any of the PTs underneath it; and (iii) Test a HLGT with at least 50 events if no HLT or PT underneath it is being tested. Otherwise, test a HLGT if it reports at least 50% extra events as in any of the HLTs or PTs underneath it.

*Simulations for the combination of tests from the null hypothesis (H_0) and alternative hypothesis (H_1) were using RR of 1.0 for H_0 and RR of 1.03, 1.05, 1.10, 1.20, 1.30, 1.40, 1.50, 1.70, 1.90, 2.00, 2.50, 3.00, 4.00, 5.00, 7.00 and 10.00 for H_1.

Results are produced based on 10,000 simulations.

FDR=False Discovery Rate (FDR) and FWER=FamilyWise Error Rate.

5 Meta-analyses of the double-blind trials of statin vs placebo

5.1 Summary

This chapter presents the results of the meta-analyses of all AEs reported in 19 double-blind trials and 123,940 patients comparing statin therapy with placebo. The results from blinded (ie, scrambled) analyses reveal the importance of using statistical methods to correct for multiple hypotheses testing to avoid generating false positive findings. Results from the unblinded analyses (ie, true treatment allocation) then confirm that the beneficial effects of statin therapy on major cardiovascular disease outcomes remain, confirming that the new methodologies developed for the harmonisation of outcomes into a common medical dictionary are reliable. Allocation to a statin is also shown to reduce the risk of other vascular outcomes such as peripheral embolism and thrombosis, and the rarely reported outcome (in these trials) of carotid endarterectomy. By contrast, after accounting for multiple testing, the results indicate that statin therapy increases the risk of hepatobiliary investigations (ie, increases biochemistry biomarkers related to liver function) and, extremely rarely, parosmia. The recently-investigated and documented adverse effects of statin therapy on muscle symptoms are also identified as a large component of the AE of musculoskeletal and connective tissue disorders. These results present the very few generally mild adverse effects of statin therapy which are outweighed by the established beneficial effects of statin for preventing life-threatening cardiovascular disease.

5.2 Introduction

Cardiovascular disease is the leading cause of adult mortality and morbidity worldwide with an average of about 18 million deaths each year (3). Studies have shown a log-linear positive association between low-density lipoprotein (LDL) cholesterol and the risk of atherosclerotic coronary heart disease exist (5, 6). Statins are designed to reduce the production of LDL

cholesterol in the liver by about 30% with standard dose (eg, 20-40 mg simvastatin daily), with such reduction in LDL cholesterol leading to reduction in risk of major vascular disease.

In particular, standard statins were shown to reduce the risk in major vascular events (MVE) by about 21% when compared with placebo for every 1.0 mmol/L reduction in LDL cholesterol (18, 19). The proportional reduction in risk of MVE per 1.0 mmol/L reduction in LDL cholesterol was similar irrespective of baseline cholesterol levels, prior disease or other patient characteristics (18, 22-24).

Non-randomised observational studies have shown that statins are associated with an increased risk of various AEs (eg, myalgia, hepatic dysfunction cataracts, and depression). Such claims have led many people at high risk of vascular disease to stop or not start statin therapy, putting their health in danger.

However, non-randomised studies may be subject to moderate systematic biases and thereby produce very misleading results. Randomisation is the only way to guarantee the avoidance of moderate biases while studying large number of outcomes is the best was to avoid moderate random errors. Therefore, the most robust way to assess any adverse effects of statin is through a meta-analysis of all AEs reported in all large-scale randomised trials of statin therapy.

For this reason, the CTT Collaboration initiated a new programme of work, with a protocol describing this published in 2016 (35), to collect and analyse patient level data on all AEs from CTT-eligible statin trials. These AE outcomes can be broadly classified into 2 main classes: AEs probably or possibly related to statin therapy (ie, muscle pain and diabetes); and AEs with no prior knowledge (from randomised evidence) of an association with statin therapy.

This project is focused on the second category of AEs (ie, with previously unknown association with statin therapy) which contains the vast majority of AEs. The analyses of those AEs face substantial challenges related to multiple testing and the aim of this thesis is to develop appropriate statistical methods to address the problems associated with multiple testing and

identify any real, previously unknown, effects of statin (ie, both harmful and beneficial). This chapter reports the meta-analyses of all AEs reported in the main (analysis) population, namely the 19 double-blind trials of statin vs placebo involving 123,940 participants.

5.3 Methods

5.3.1 Data

CDISC global standards and a streamlined subset of the SDTM (version 3.2) were selected as the preferred system to harmonise the IPD received from the 19 trials into a single dataset with a common format ready for analyses (see chapter 3.2.2). In addition, MedDRA version 20.0 was employed as the reference medical terminology dictionary to harmonise the AEs from the multiple trials using numerous event coding formats, blind to treatment allocation, into a single common language (see chapter 3.2.3).

5.3.2 Analysis

The primary analysis population was the 19 double-blind trials comparing statin therapy with placebo. Analyses were to include all randomised participants irrespective of whether they received their allocated treatment (ie, intention-to-treat). Adverse events reporting at least one event were included for analyses with a time to event analysis being performed. Such survival analysis (ie, accounting for the time of the event) is more powerful than the analyses of count data used in the simulation analyses and it may require less than 6 events to detect a difference as statistically significant at 5% level.

5.3.2.1 Statistical analysis

For each participant, the time to the first occurrence of an AE of a specific type was considered for analysis. For every AE reported by at least one participant in a particular trial, the log-rank 'observed minus expected' statistic (o – e) and its variance (v) for the first occurrence of each outcome among patients randomised into each trial (chapter 3.6.1).

In a meta-analysis of trials, the log of the rate ratio of was calculated as S/V with variance 1/V where S is the sum over all trials of (o–e) and V is the sum over all trials of v and hence the 95% confidence interval calculated as S/V±1·96/√V (chapter 3.6.2). Then, the p-value (uncorrected for MHT) was calculated from the comparison of the test statistic (ie, S/V) against a chi-square distribution (ie, χ^2) with 1 degree of freedom.

5.3.2.2 Multiple testing

The Mehrotra & Adewale MHT method controlling for the FDR at 5% (ie, 0.05) level of significance (see chapter 2.5.2.4) was used to test the multiple null hypotheses of all AEs examined within a single analyses (ie, family) and reject those null hypotheses (through their p-values) in order to identify the effects of statin therapy (beneficial or harmful) which were statistically significant.

Mehrotra & Adewale method requires a 2-step approach. In the first step, for each SOC, the minimum nominal p-value from all tests performed within that group were identified and the Benjamini & Hochberg MHT method at 5% FDR level to identify the significant p-values applied. The Benjamini & Hochberg method requires the p-values to be sorted in descending order moving from higher to lower p-values, and it rejects the null hypothesis as soon as 0.05*j/m, where j is the order and m is total number of tests, after which all remaining hypotheses are rejected. Then, in the second step, for the SOCs identified as being significant from step 1 above, all AEs underneath are tested using again the Benjamini & Hochberg MHT method at 5% FDR.

5.3.2.3 Analyses of AEs

To illustrate the importance and impact of the various stages of the MHT strategy, five different sets of analyses were performed:

1. Analyses of all AEs blinded to (true) treatment allocation

   Such analyses were performed to estimate the number of false discoveries generated when there was no correction for MHT (ie, allowing a type I error equal to 5% for each

test performed) and when the Mehrotra & Adewale MHT method are applied to all AEs when using shuffled allocation (121). The true treatment allocations were shuffled to ensure that any 'significant' results derived from the shuffled dataset were, by definition, type I errors.

2. Unblinded analyses of all non-lipid related AEs

   The aim of such analyses is to assess that the data harmonisation methodologies used for this project (ie, SDTM for standardisation of IPD and MedDRA for harmonisation of AEs to a common code) were performing as expected and resulting in identification of previously known effects of statin therapy (ie, myocardial infarction, death, etc.). The list of lipid-related AE (MedDRA) terms excluded from the analyses is presented in **Appendix Table 2** with such events being enormously significant after MHT testing if they are allowed to contribute to the analysis.

3. Unblinded analyses of all AEs with prior unknown association with statins

   These analyses aimed to identify any new beneficial or harmful effects of statin therapy. To achieve this, any AE with an already known association with statin therapy from previous CTT work was excluded from analyses. These included: (i) lipid related AEs; (ii) muscle AEs pre-defined (122); (iii) diabetes AEs pre-defined; and (iv) AEs related to the known cardiovascular benefits of statin (**Appendix Table 2**).

4. Unblinded analyses of all AEs with prior unknown association with statins after excluding very rare outcomes and avoiding irrelevant overlap in tests

   These analyses were performed to identify any additional (previously unknown) effects of statin therapy by applying the chosen multiple testing strategy (Chapter 4). Specifically, analyses exclude very rare outcomes for which the power of detecting any true effect of statin therapy, even of large magnitude, is low and avoid irrelevant overall in tests across the MedDRA tree (chapter 4.4.4.1). To achieve this, after excluding all terms listed in **Appendix Table 2**, the following hierarchical strategy was applied to select the appropriate hypotheses:

I. Test a PT if it was reported for at least 50 participants (across all trials combined).

II. Test all HLTs where the number of events is at least 50 and where the number of events is at least 50% higher than for any PT underneath it.

III. Test all HLGTs where the number of events is at least 50 and where the number of events is at least 50% higher than for any HLT or PT underneath it.

IV. Test all SOCs as these terms contain disorders by body system (ie, eye disorders), relate to aetiology (ie, infections and infestations), or serve a classification purpose (ie, surgical and medical procedures) (109).

5. Further exploratory analyses

   For any AEs with a prior unknown association of effect with statin therapy identified as FDR significant (ie, from analyses 3 and 4), further exploratory analyses were performed to examine the effect of statin on all sublevels of each AE as reported at the lower levels of MedDRA.

5.4 Results

5.4.1 Data

Individual participant data were available from 19 randomised double-blind trials of any statin regimen versus placebo resulting in a total of 62,028 patients assigned to receive a statin and 61,912 to receive placebo (**Table 5.1**). Adverse event data were redacted by the data provider in approximately 5% of patients in the AURORA, CORONA and JUPITER trials (because of data privacy concerns). Overall, the mean age of the 123,940 participants was 63 (SD 8) years, 28% of them were women, 48% had a history of vascular disease, and 18% had diabetes mellitus (**Table 5.1**). During a weighted-average median trial follow-up of 4.3 years, a total of 890,273 AEs were reported, of which 40,221 were unique types of AEs (**Table 3.1**).

The distribution of the total number of (participants with) events across the three MedDRA hierarchical levels (ie, PT, HLT and HLGT) is shown in **Figure 5.1** with additional details organised alphabetically by the 27 System Organ Classes presented in **Table 5.2**. In total, there were 7,639 AE terms reported by at least one participant (including the 27 SOCs), 5,652 (74.0%) of those reported less than 50 events, 578 (7.6%) had a count of between 50 and 99 events, 1,027 (13.4%) reported between 100 and 999 events and 382 (5.0%) reported at least 1000 events. From the total of 7,639 terms (reported across the MedDRA hierarchical tree), 6,063 (79.4%) were PTs with median (IQR) number of events equal to 5 (2 – 25), 1,233 (16.1%) were HLTs with median (IQR) number of events equal to 47 (7 – 245), 316 (4.1%) were HLGTs with median (IQR) number of events equal to 288 (61 – 1,462) and 27 (0.4%) were SOCs with median (IQR) number of events equal to 10,896 (3,580 – 24,662) (**Figure 5.1**).

5.4.2 Analyses of AEs blinded to treatment allocation

Figure 5.2 presents the distribution of nominal p-values for all AEs with at least 6 events, organised alphabetically by SOC and ordered by direction of the effect of statins (ie, higher or lower), and when masking the true treatment assignment. In total, 4,094 tests were performed, 27 SOCs, 250 HLGTs, 892 HLTs and 2,925 PTs. The approach of not correcting for MHT (ie, control the type I error at 5% level of significance) resulted in a total of 179 (4.4%) false positive discoveries while the use of the Mehrotra & Adewale method at 5% FDR reduced the number of FP findings to zero.

5.4.3 Unblinded analyses of all non-lipid related AEs

Figure 5.3 presents the (nominal) p-values for all non-lipid related AEs with at least 6 events organised alphabetically by SOC and ordered by direction of the effect of statins. In total, 4,068 tests were performed, 27 SOCs, 248 HLGTs, 887 HLTs and 2,906 PTs. The Mehotra & Adewale method identified 4 SOCs ('Pregnancy, puerperium and perinatal conditions', 'Endocrine disorders', 'Hepatobiliary disorders', 'Product issues') which did not contribute

further to the second step since the representative (ie, minimum) nominal p-values (0.414, 0.052, 0.069 and 0.156 respectively) were not found to be FDR significant in the first step of this method. In total, 57 AEs were found significant at the 5% FDR level shown with closed circles and enumerated in **Figure 5.3,** with further details for the event rates and treatment effects shown in **Figure 5.4**.

Most of those 57 AEs are known cardiovascular benefits of statin therapy, with the most known effects being the PT 'Myocardial infarction' (2,154 [3.5%] in statin arm vs 3,046 [4.9%] in placebo arm, RR 0.70, 95% CI: 0.66 – 0.74), the HLT 'Death and sudden death' (4,996 [8.1%] vs 5,440 [8.8%], RR 0.91, 95% CI: 0.88 – 0.95) and PT 'Ischaemic stroke' (901 [1.5%] vs 1184 [1.9%], RR 0.76, 95% CI 0.69 – 0.82) (**Figure 5.4**). Due to the hierarchical structure of MedDRA, more generic cardiovascular terms (ie, beyond those described in previous CTT work) were identified as FDR significant, including the SOC 'Vascular disorder' (8,359 [13.5%] vs 8,864 [14.3%], RR 0.93, 95% CI 0.91 – 0.96), the HLGT 'Cardiac and vascular investigations (excl enzyme tests)' (3,560 [5.7%] vs 3,999 [6.5%], RR 0.87, 95% CI 0.83 – 0.91) and the SOC 'Cardiac disorders' (14,878 [24.0%] vs 16,400 [26.5%], RR 0.89, 95% CI 0.87 – 0.91) (**Figure 5.4**).

Other important cardiovascular benefits found to be FDR significant were the HLGT 'Coronary artery disorders' (9,833 ([15.9%] vs 11,338 [18.3%], RR 0.84, 95% CI 0.82 – 0.87), the HLGT 'Heart failures' (3,075 [5.0%] vs 3,372 [5.4%], RR 0.91, 95% CI 0.86 – 0.95) and the HGLT 'Central nervous system vascular disorders' (3,112 [5.0%] vs 3,701 [6.0%], RR 0.83, 95% CI 0.79 – 0.87) (**Figure 5.4**). Also, allocation to a statin resulted in a risk reduction of the very rare AE PT 'Nodule' (6 [<0.1%] vs 23 [<0.1%], RR 0.31, 95% CI 0.15 – 0.64), the PT 'Angiogram' (367 [0.6%] vs 535 [0.9%], RR 0.67, 95% CI 0.58 – 0.76) and the PT 'Arteriogram coronary' (1,377 [2.2%] vs 1,636 [2.6%], RR 0.82, 95% CI 0.77 – 0.89) (**Figure 5.4**).

Seven AE terms were found to be significantly increased with statin therapy which could be classified into three types: the SOC of 'Musculoskeletal and connective tissue disorder'

(22,012 [35.5%] vs 21,417 [34.6%], RR 1.04, 95% CI 1.02 – 1.06), the HGLT 'Hepatobiliary investigations' (1,349 [2.2%] vs 1,025 [1.7%], RR 1.32, 95% CI 1.22 – 1.43) including further sub-classifications (ie, liver function analyses, hepatic enzyme increased, alanine and aspartate aminotransferase increased) and the very rare PT 'Parosmia' (18 [<0.1%] vs 2 [<0.1%], RR 4.93, 95% CI 2.05 – 11.86) (**Figure 5.4**).[k] One of the two participants with parosmia on placebo arm was on a non-study statin at the time of the event.

5.4.4 Unblinded analyses of all AEs with prior unknown association with statins

Figure 5.5 presents the (nominal) p-values for AEs with at least 6 events except those related to muscle, diabetes and the known cardiovascular benefits of statins, organised alphabetically by SOC and presented by direction of the effect of statins. In total, 3,970 tests were performed, 25 SOCs, 243 HLGTs, 877 HLTs and 2,825 PTs. The Mehotra & Adewale method identified the same 4 SOCs (ie, 'Pregnancy, puerperium and perinatal conditions', 'Endocrine disorders', 'Hepatobiliary disorders', 'Product issues') as in section 5.4.3 which were not FDR significant at the first step, and consequently none of the terms reported within those SOCs were examined in the second step. 25 AEs were found significant at the 5% FDR level shown in **Figure 5.5** with further details for the event rates and treatment effects shown in **Figure 5.6**.

Compared with the previous analyses (ie, all non-lipid related AEs, section 5.4.3), no new beneficial or harmful effects of statin were identified (**Figures 5.4** and **5.6**).

[k] Note that Peto's 'one-step' method provides a good approximation to the RR when the RR is moderate (ie, between 0.5 and 2), and is especially useful in the current setting where event counts in a given trial may be zero in one of the treatment arms, but it becomes increasingly biased for RRs more extreme than this. For the PT of parosmia, therefore, the 'one-step' estimate of the RR would be a substantial underestimate, and the crude risk ratio (ie, calculated from a single 2x2 contingency table across all trials pooled together) of 9.0 (95% CI 2.1 – 81.6) likely be a better estimate.

5.4.5 Unblinded analyses of all AEs with prior unknown association with statins after excluding very rare outcomes and avoiding irrelevant overlap in tests

Figure 5.7 presents the distribution of the (nominal) p-values for all AEs examined in analyses just above (section 5.4.4) after applying a testing strategy which excluded very rare AEs and avoided irrelevant overlap in tests performed across the higher levels of MedDRA. In total, 1,486 tests were performed, 25 SOCs, 118 HLGTs, 295 HLTs and 1,049 PTs. The Mehotra & Adewale method identified the same 4 SOCs as in section 5.4.3 which did not contribute any AE terms to the second step of the MHT analyses. 22 AEs were found significant at the 5% FDR level shown in **Figure 5.7** with further details for the event rates and treatment effects shown in **Figure 5.8**.

Compared with analyses in section 5.4.4, two additional beneficial effects of statin therapy were identified: the HLT 'Peripheral embolism and thrombosis' (361 [0.6%] vs 451 [0.7%], RR 0.80, 95% CI 0.69 – 0.91) and the PT 'Carotid endarterectomy' (47 [<0.1%] vs 83 [0.1%], RR 0.57, 95% CI 0.40 – 0.80) (**Figure 5.8**). No new harmful effects of statin therapy were found to be FDR significant. (Note that these analyses were tested on AEs with at least 50 events, or 50% extra information compared with the MedDRA levels underneath and, among other AEs, the PTs 'Nodule', 'Hemiplegia' [ie, a condition caused by a stroke event] and 'Parosmia' were not being tested.)

5.4.6 Further exploratory analyses

Further exploratory analyses examining the effects of statin therapy on all sublevels of each AE declared as FDR significant in the analyses above identified that the beneficial effect of statin on the HLGT 'Heart failures NEC' was driven by the PT 'Cardiac failure' (1,874 [3.0%] vs 2,068 [3.3%], RR 0.90, 95% CI 0.84 – 0.96, P=0.0008). Also, within the HLGT 'Hepatobiliary investigation', statin therapy was associated with a 77% increase in the risk of the PT 'Blood bilirubin increased' (56 [<0.1%] vs 31 [<0.1%], RR 1.77, 95% CI 1.16 – 2.69, P=0.0078).

5.5 Discussion

This chapter aimed to analyse all AEs reported in the 19 trials of statin vs placebo (involving 123,940 participants) and identify any real, previously unknown, beneficial or harmful effects of statin.

The analyses of all AEs in trials of statin vs placebo were expected to examine a very large number hypotheses and the use of statistical methods which account for the multiple hypotheses testing problem was required to avoid the high likelihood of identifying FP findings. As expected, results from the analyses of all AEs blinded to treatment allocation (ie, replacing true treatment with a shuffled allocation) led to a large number of FPs (ie, 179 out of a total of 4,094 AEs) when no correction for multiple testing was applied (ie, each hypothesis was tested against a type I error equal to 5%), but this was reduced to zero when the analyses accounted for MHT correction by controlling the FDR error using the Mehrotra & Adewale method. This highlights: (i) the importance of using statistical techniques which account for the problem of MHT, and (ii) any significant results after application of the Mehrotra & Adewale correction method are likely to be real.

For the purpose of this project, a new methodology was employed to harmonise all AEs reported across the 28 trials (contributing to the various meta-analyses) into a single medical dictionary (ie, MedDRA V20.0) (see chapter 3.2.3). As with any new tool, a set of analyses is required which can be used as a method of validation by confirming that the results produced using this new harmonisation method are in agreement with respect to the already known (from previous CTT work) effects of statin therapy. To achieve this, the initial set of analyses unblinded to treatment allocation (ie, showing true effects of statins) included all non-lipid AEs. Results showed that statins, as expected, reduce the risk of various types of cardiovascular disease such as myocardial infarction by 30%, death by 9%, angina pectoris by 11% and ischaemic stroke by 24% over an average follow up of 4.3 years of treatment. The treatment effect estimates from these analyses are similar to those reported in previous CTT papers (chapter 1.3.1) (albeit those previous analyses were reported per 1 mmol/L reduction in LDL

cholesterol). This confirmed that the methodology developed for the harmonisation of outcomes using the MedDRA dictionary performs well for known effects on AEs and consequently can be safely used for detecting any new (ie previously undetected) effects of statin therapy.

The analyses also confirmed the cardiovascular benefits on heart failure (often reported as cardiac failure) (ie, RR 0.91, 95% CI 0.86 − 0.95). Also, allocation to statin therapy was shown to reduce the risk of the AEs 'arteriogram coronary' by 18% and 'angiogram' by 33%. A possible reason that the reduction on the number of those procedures (ie, arteriogram coronary and angiogram) when taking statins has not been identified in previous CTT work is that such type of AEs have not been adjudicated and are reported in raw AE datasets.

Two further analyses were performed to identify any additional new effects of statin therapy when reducing the number of hypotheses being tested (ie, since this number is negatively associated with the statistical power of detecting true effects of statin therapy (chapter 4.4.3.3) by (i) removing already known effects of statin (ie, cardiovascular, muscle and diabetes) (**Appendix Table 2**) and (ii) in addition to those AEs, removing very rare outcomes and avoiding irrelevant overlap in tests (section 5.3.2.3 and chapter 4.4.4). These analyses identified two new benefits of statin therapy which was shown to reduce the risk of peripheral embolism and thrombosis by 20% and the very rare AE of carotid endarterectomy (ie, reported in 1 per 1,000 participants) by 43%.

Statins were shown to increase the risk of a few adverse events. In particular, allocation to a statin was associated with a 4% increased risk of musculoskeletal and connective tissue disorders, with a large proportion of outcomes for such generic AE coming from muscle-related symptoms. The effects of statin therapy on muscle symptoms (ie, myalgia, muscle cramps or spasm, limb pain, and other musculoskeletal pain) has been recently investigated and documented showing similar treatment effects as above, with statin regimens causing a small (ie, 3%) relative increase in the number of first reports of muscle symptoms, with the excess

in risk due to statin therapy largely limited to the first year of treatment (ie, absolute excess rate of 11 (95% CI 6 - 16) events per 1000 person-years) (122).

Also, allocation to a statin was associated with a 32% increased risk of hepatobiliary investigations with such an effect driven by the AE of liver function analyses which increased by 37% (RR 1.37, 95% CI 1.26 − 1.49) when receiving statin therapy (ie, which comprised of the AEs of hepatic enzyme increase, blood bilirubin, alanine and aspartate aminotransferease increased). In addition, the very rare occurrence of parosmia (reported by less than 2 in 10,000 participants) was shown to be about 9 times higher when taking statins (with such an AE reported by at least one participant in 8 out of 19 trials).

These analyses have several advantages over other studies reporting adverse effects of statin therapy. Unlike non-randomised observational studies which may be subject to moderate biases (chapter 1.3.3), these analyses minimise such biases by including data from large-scale, randomised, double-blind trials of statin therapy in which there was systematic and unbiased event reporting. In addition, patient level data including all adverse events reported in the individual trials were collected to overcome any potential bias arising from the selected publication of results. Also, all adverse events were coded blinded to treatment allocation to a common standard (ie, MedDRA) to reduce any biases coming from the various methodologies used by the individual trials.

A limitation of these analyses is that the effects of statin therapy on hepatobiliary investigations and liver enzymes is estimated by data reported within the AE datasets of each trial without incorporating any further biochemistry abnormalities (ie, high values of bilirubin, transaminases, alanine and aspartate aminotransferease) directly detected from the laboratory datasets. Further analyses are required (ie, based on hypotheses generated from these analyses) to estimate the effects of statins on more detailed endpoints for hepatobiliary disease and whether these effects may be different in various type of participants.

In summary, the results presented in this chapter confirm the wide range of cardiovascular benefits of statin therapy and identify very few, previously unknown, adverse effects of statin therapy (ie, hepatobiliary investigationsand parosmia). These results can be used to highlight the importance of taking statin therapy by comparing the few adverse effects caused while receiving this treatment with its' beneficial effect on lowering LDL cholesterol and reducing the risk of cardiovascular disease. The absolute excess in risk (when taking statins) for the three types of adverse effects identified (ie, musculoskeletal and connective tissue disorders, hepatobiliary investigations and parosmia) ranges from 2 to 100 per 10,000 treated for 5 years For example, in 10,000 people, lowering LDL cholesterol by about 1 mmol/L for 5 years typically caused around 110 musculoskeletal and connective tissue disorders, 60 hepatobiliary disorders and 2 cases of parosmia, which compares to the prevention of around 250 heart attacks, strokes, or coronary procedures in primary prevention. Thus, although the results in this chapter have identified some novel potential side effects of statin therapy, the identified AEs are not serious and are greatly outweighed by reductions in major cardiovascular diseases and in the overall risk of death.

Table 5.1: Characteristics of the participating trials

Study	Year of publication of primary results	Number of patients	Treatment comparison (mg/day)	Median follow-up (years)	Baseline LDL-C (mmol/L)	Baseline Age (years)	Women (%)	White/Caucasian (%)	History of vascular disease (%)	History of Diabetes (%)	Baseline Body Mass Index (kg/m2)
SSSS	1994	4444	S20-40 vs placebo	5.4	4.9 (0.7)	59 (7)	827 (19%)	NA	4444 (100%)	202 (5%)	26.0 (3.3)
WOSCOPS	1995	6595	P40 vs placebo	4.8	5.0 (0.5)	55 (6)	0 (0%)	NA	1066 (16%)	77 (1%)	26.0 (3.2)
CARE	1996	4159	P40 vs placebo	4.9	3.6 (0.4)	59 (9)	576 (14%)	3851 (93%)	4159 (100%)	586 (14%)	27.6 (4.4)
AFCAPS/TexCAPS	1998	6605	L20-40 vs placebo	5.0	3.9 (0.4)	58 (7)	997 (15%)	5860 (89%)	0	155 (2%)	26.9 (3.1)
LIPID	1998	9014	P40 vs placebo	5.9	3.9 (0.8)	61 (8)	1516 (17%)	NA	9014 (100%)	782 (9%)	26.8 (3.8)
LIPS	2002	1677	F80 vs placebo	4.0	3.4 (0.8)	60 (10)	271 (16%)	1650 (98%)	1677 (100%)	202 (12%)	26.5 (3.3)
HPS	2002	20536	S40 vs placebo	5.2	3.4 (0.8)	64 (8)	5082 (25%)	19901 (97%)	17386 (85%)	5963 (29%)	27.6 (4.4)
PROSPER	2002	5804	P40 vs placebo	3.3	3.8 (0.8)	75 (3)	3000 (52%)	NA	2565 (44%)	623 (11%)	26.8 (4.2)
ASCOT-LLA	2003	10240	A10 vs placebo	3.3	3.4 (0.7)	63 (9)	1919 (19%)	9687 (95%)	1684 (16%)	2540 (25%)	28.6 (4.6)
ALERT	2003	2102	F40 vs placebo	5.5	4.1 (1.0)	50 (11)	715 (34%)	2039 (97%)	409 (19%)	396 (19%)	25.8 (4.5)
CARDS	2004	2838	A10 vs placebo	4.2	2.9 (0.8)	61 (8)	909 (32%)	2676 (94%)	106 (4%)	2838 (100%)	29.0 (5.1)
DDDD	2005	1255	A20 vs placebo	2.7	3.3 (0.8)	66 (8)	578 (46%)	924 (74%)	1041 (83%)	1255 (100%)	27.6 (4.8)
ASPEN	2006	2410	A10 vs placebo	4.0	2.9 (0.7)	60 (8)	811 (34%)	2029 (84%)	747 (31%)	2410 (100%)	28.9 (3.8)
SPARCL	2006	4731	A80 vs placebo	4.9	3.5 (0.6)	63 (11)	1908 (40%)	4415 (93%)	4731 (100%)	794 (17%)	27.9 (5.2)
CORONA	2007	4982	R10 vs placebo	2.7	3.6 (0.9)	72 (7)	1175 (24%)	NA	4982 (100%)	1473 (30%)	26.4 (3.6)
GISSI-HF	2008	4574	R10 vs placebo	3.9	3.1 (0.9)	68 (11)	1032 (23%)	4574 (100%)	4574 (100%)	1196 (26%)	27.1 (4.5)
JUPITER	2008	16714	R20 vs placebo	1.9	2.7 (0.5)	65 (8)	6374 (38%)	NA	0 (0%)	44 (0%)	27.5 (3.6)
AURORA	2009	2555	R10 vs placebo	3.9	2.6 (0.9)	64 (9)	969 (38%)	NA	1025 (40%)	658 (26%)	24.8 (3.9)
HOPE-3	2016	12705	R10 vs placebo	5.5	3.3 (0.9)	66 (6)	5874 (46%)	2546 (20%)	0 (0%)	731 (6%)	27.1 (4.7)
Subtotal 19 trials	**-**	**123940**	**-**	**4.3**	**3.5 (0.7)**	**63 (8)**	**34533 (28%)**	**60152 (81%)**	**59610 (48%)**	**22925 (18%)**	**27.2 (4.1)**

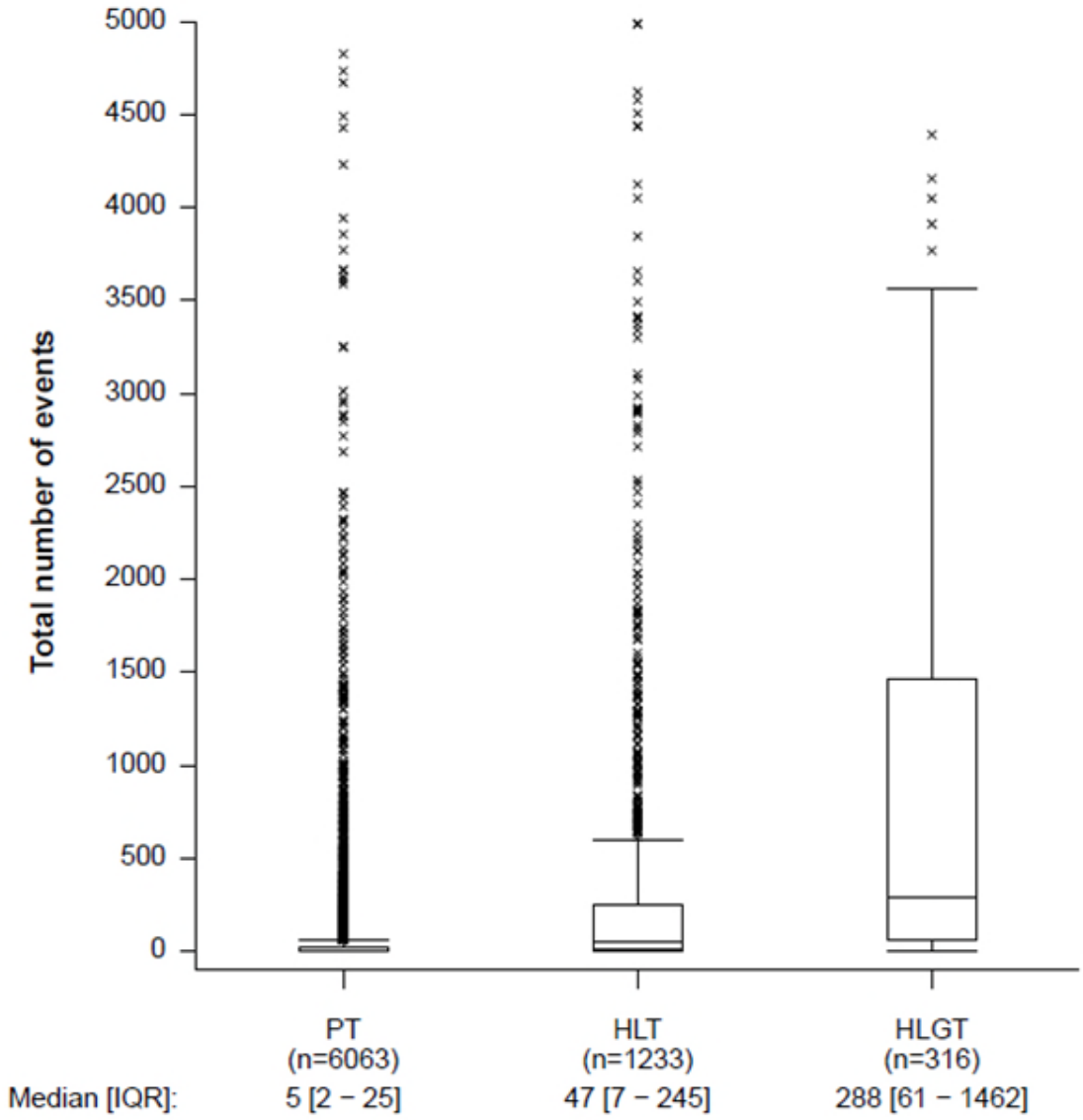
S=Simvastatin; P=Pravastatin; L= Lovastatin; F=Fluvastatin; A=Atorvastatin; R=Rosuvastatin; Age, LDL-C and BMI presented as mean(SD); NA=Not available; SSSS=Scandinavian Simvastatin Survival Study; WOSCOPS=West of Scotland Coronary Prevention Study; CARE=Cholesterol And Recurrent Events; AFCAPS/TexCAPS=Air Force/Texas Coronary Atherosclerosis Prevention Study; LIPID=Long-term Intervention with Pravastatin in Ischaemic Disease; LIPS=Lescol Intervention Prevention Study; HPS=Heart Protection Study; PROSPER=PROspective Study of Pravastatin in the Elderly at Risk; ASCOT-LLA=Anglo-Scandinavian Cardiac Outcomes Trial-Lipid Lowering Arm; ALERT=Assessment of Lescol in Renal Transplantation; CARDS=Collaborative Atorvastatin Diabetes Study; 4D=Die Deutsche Diabetes Dialyse Studie; ASPEN=Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus; SPARCL=Stroke Prevention by Aggressive Reduction in Cholesterol Levels; CORONA=Controlled Rosuvastatin Multinational Trial in Heart Failure; JUPITER=Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin study group; GISSI-HF=Gruppo Italiano per lo Studio della Sopravvivenza nell'Insufficienza cardiaca; AURORA=A Study to Evaluate the Use of Rosuvastatin in Subjects on Regular Hemodialysis: An Assessment of Survival and Cardiovascular Events; HOPE-3=Heart Outcomes Prevention Evaluation-3 trial; AURORA, CORONA and JUPITER supplied age in categorical bands - and so the midpoint of each categorical band was used as a surrogate for baseline age in all analyses. A small number of participants in the AURORA(n=218), CORONA(n=27) and JUPITER(n=1088) trials withdrew consent for use of their data post-trial, and hence data from these participants is excluded. The ASCOT-LLA trial excludes 65 patients for whom data was not available due to protocol violations. If data was NA then denominator did not contribute to weighted subtotal.

Table 5.2: Distribution of the total number of events across the three analysable MedDRA hierarchical levels, organized alphabetically by System Organ Class (19 trials, 123,940 randomised participants)

	Number of tests at each sublevel with total number of events between <50| 50-99|100-999|>1000		
System Organ Class (Events)	**Higher Level Group Term**	**Higher Level Term**	**Preferred Term**
Blood and lymphatic system disorders (3,649)	3|2|4|1	14|6|3|1	76|4|3|1
Cardiac disorders (31,278)	2|0|3|4	10|0|11|6	103|20|28|9
Congenital, familial and genetic disorders (535)	12|2|1|0	40|1|1|0	88|1|1|0
Ear and labyrinth disorders (5,768)	0|0|2|3	6|1|4|2	37|4|11|1
Endocrine disorders (1,362)	4|1|1|1	16|1|4|0	44|4|4|0
Eye disorders (10,118)	0|2|6|3	16|11|14|3	189|15|22|3
Gastrointestinal disorders (29,982)	1|0|8|10	35|10|28|13	311|28|45|12
General disorders and administration site conditions (48,326)	0|0|3|4	13|5|6|8	131|13|16|14
Hepatobiliary disorders (2,811)	0|0|1|2	2|2|8|1	64|8|7|0
Immune system disorders (1,706)	2|0|1|1	8|0|4|0	35|1|5|0
Infections and infestations (39,991)	7|1|0|4	70|5|11|15	476|26|53|17
Injury, poisoning and procedural complications (18,664)	1|1|2|4	32|3|12|9	286|36|49|4
Investigations (23,157)	2|2|8|10	40|11|27|10	588|31|53|8
Metabolism and nutrition disorders (10,896)	3|2|4|4	18|4|15|3	77|14|18|2
Musculoskeletal and connective tissue disorders (43,429)	0|2|1|7	6|5|16|9	166|12|37|10
Neoplasms benign, malignant and unspecified (incl cysts and polyps) (12,450)	11|10|12|6	107|19|21|5	400|22|27|3
Nervous system disorders (26,167)	3|4|4|7	31|13|17|9	270|25|35|8
Pregnancy, puerperium and perinatal conditions (9)	4|0|0|0	5|0|0|0	5|0|0|0
Product issues (1,058)	0|0|2|0	5|2|2|0	15|1|2|0
Psychiatric disorders (10,501)	6|7|5|3	32|8|11|4	140|13|9|4
Renal and urinary disorders (10,314)	0|1|5|2	9|5|10|4	123|14|23|2
Reproductive system and breast disorders (7,445)	2|1|9|2	15|9|11|3	164|11|13|2
Respiratory, thoracic and mediastinal disorders (21,133)	3|0|2|5	8|4|12|6	172|15|29|7
Skin and subcutaneous tissue disorders (14,373)	0|1|3|4	22|4|16|5	190|13|25|5
Social circumstances (3,511)	4|1|1|1	14|3|0|1	50|1|3|1
Surgical and medical procedures (26,460)	1|0|8|10	48|19|37|11	632|42|68|6
Vascular disorders (17,233)	1|1|4|5	10|3|13|5	115|9|26|2
Total number of tests	**72|41|100|103**	**632|154|314|133**	**4947|383|612|121**

MedDRA=Medical Dictionary for Regulatory Activities. MedDRA V20.0 was used to code and analyse the adverse event data. System organ class is shown in alphabetical order.

Figure 5.1: Distribution of total number of events across the MedDRA hierarchical levels (19 trials, 123,940 randomised patients)



PT= Preferred Term, HLT= Higher Level Term, HLGT= Higher Level Group Term.

There were 18 PTs, 26 HLTs and 25 HLGTs which reported a total number of events greater than 5,000.

For the 27 SOCs, the median (IQR) for the total number of events was equal to 10,896 (3,580 – 24,662).

Figure 5.2: All adverse events; BLINDED to treatment allocation (19 trials, 123,940 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among statin−allocated patients versus event rates among placebo−allocated patients (with values above the horizontal axis indicating that the event was more common among statin−allocated patients and values below the axis indicating that the event was less common among statin−allocated patients). 4094 such p−values are shown, corresponding to, 27 tests at the SOC level, 250 tests at the HLGT level, 892 tests at the HLT level and 2925 tests at the PT level. Of these tests 179 (4.4%) would be considered significant at the 5% level if no correction for multiple testing is applied while 0 (0%) would be considered significant after correction for the FDR at 5% level. SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term, PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested.

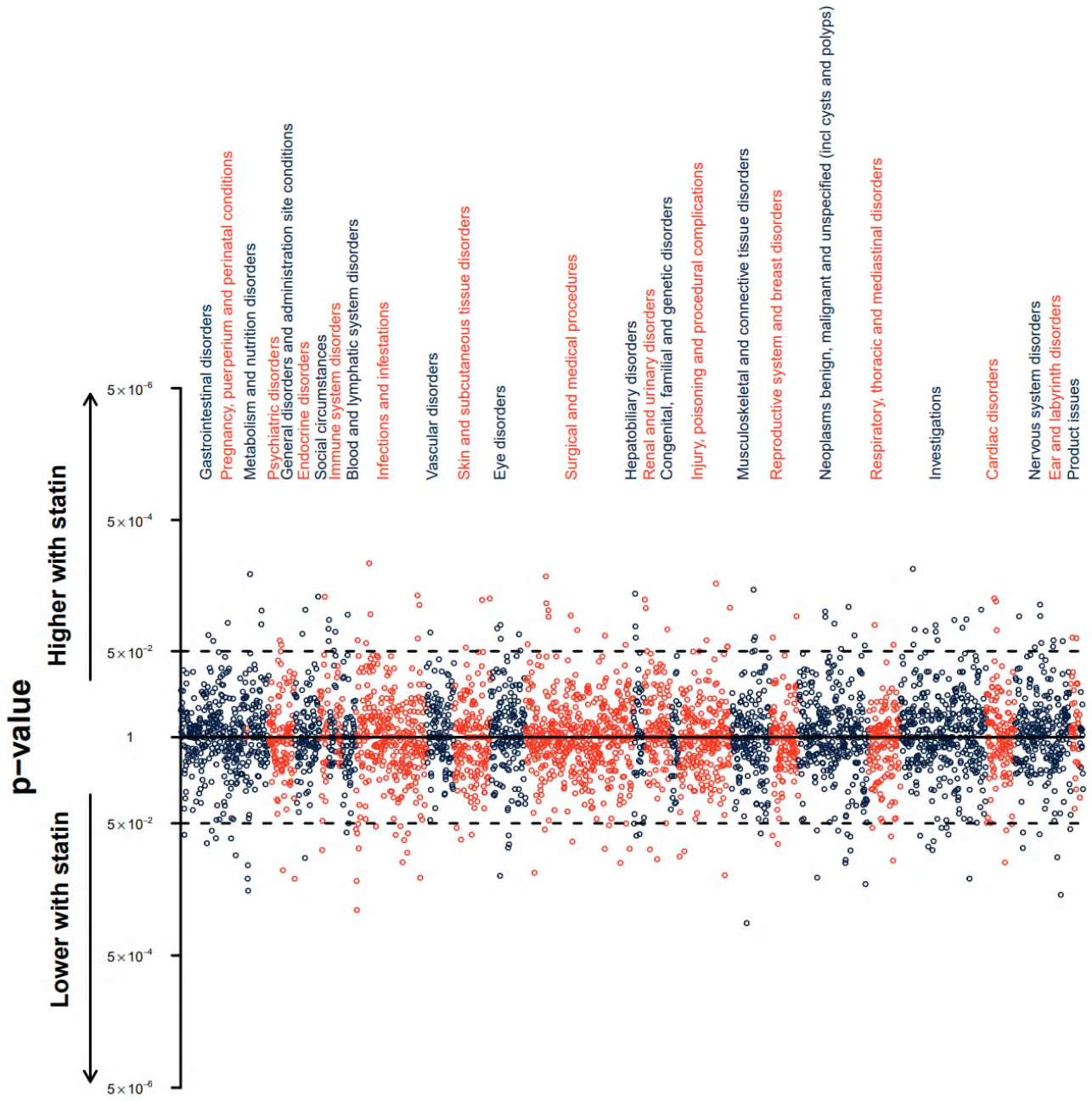


Figure 5.3: UNBLINDED analyses of all non lipid−related adverse events (19 trials, 123,940 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among statin−allocated patients versus event rates among placebo−allocated patients (with values above the horizontal axis indicating that the event was more common among statin−allocated patients and values below the axis indicating that the event was less common among statin−allocated patients). 4068 such p−values are shown, corresponding to, 27 tests at the SOC level, 248 tests at the HLGT level, 887 tests at the HLT level and 2906 tests at the PT level. Of these tests 57 (1.4%) are significant after correction of the FDR at 5% level. These are indicated with closed circles and are enumerated (see Figure 5.4 for the identification and results of each of these 57 event categories). SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term. PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested.

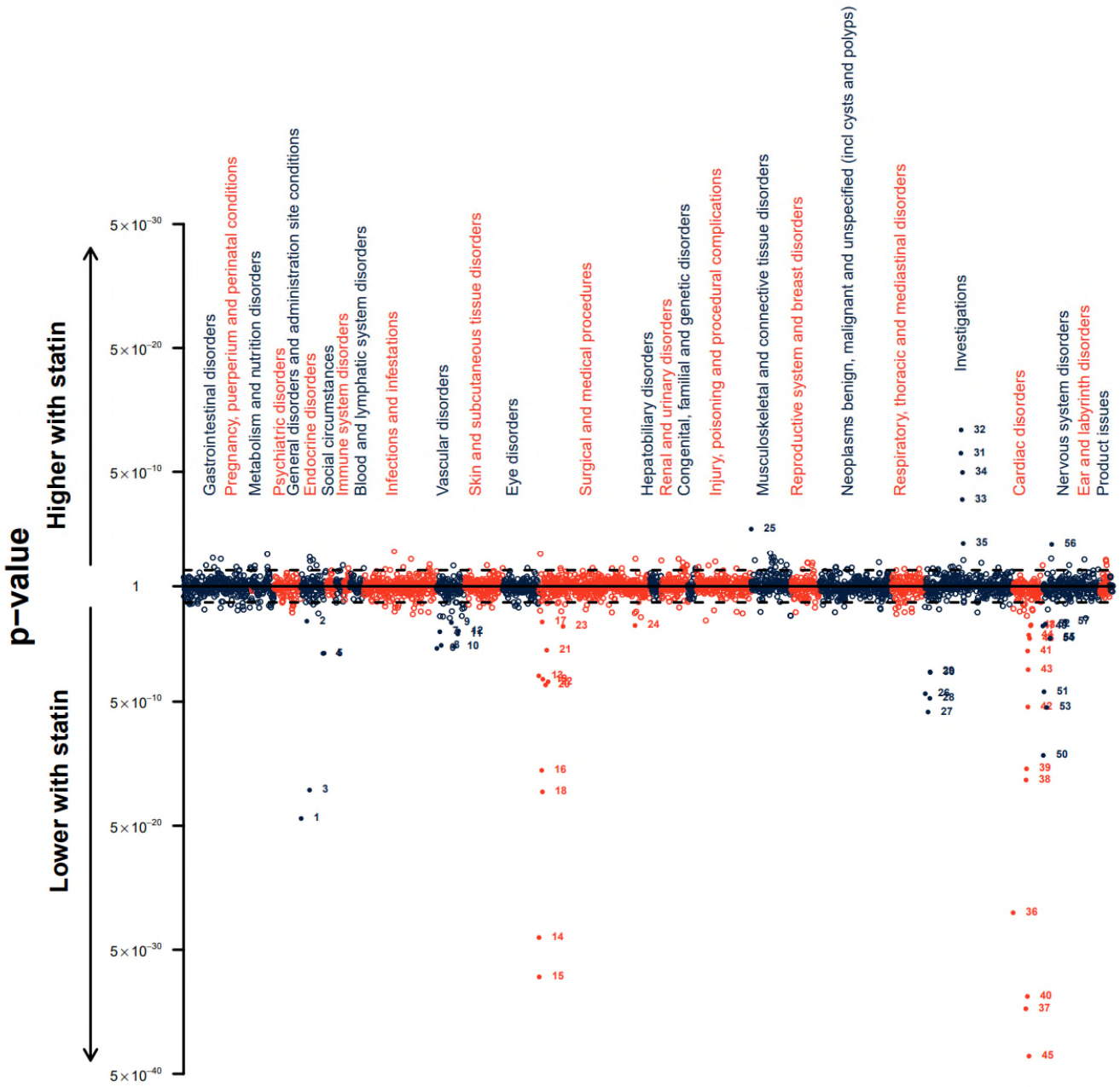
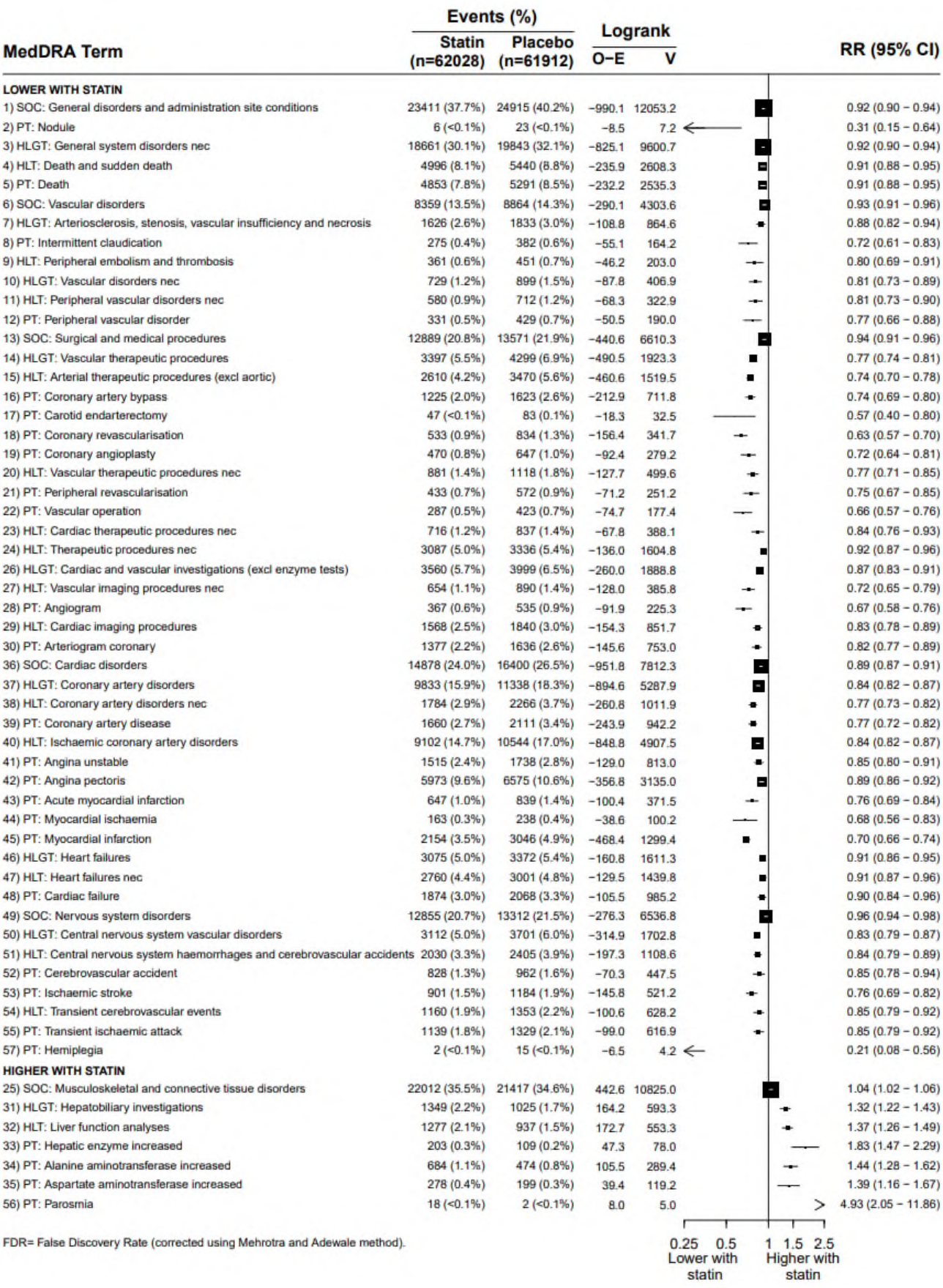


Figure 5.4: FDR significant effects of statin therapy on non- lipid-related adverse events (19 trials, 123,940 randomised patients)



MedDRA Term	Events (%) Statin (n=62028)	Events (%) Placebo (n=61912)	Logrank O−E	Logrank V	RR (95% CI)
LOWER WITH STATIN					
1) SOC: General disorders and administration site conditions	23411 (37.7%)	24915 (40.2%)	−990.1	12053.2	0.92 (0.90 − 0.94)
2) PT: Nodule	6 (<0.1%)	23 (<0.1%)	−8.5	7.2	0.31 (0.15 − 0.64)
3) HLGT: General system disorders nec	18661 (30.1%)	19843 (32.1%)	−825.1	9600.7	0.92 (0.90 − 0.94)
4) HLT: Death and sudden death	4996 (8.1%)	5440 (8.8%)	−235.9	2608.3	0.91 (0.88 − 0.95)
5) PT: Death	4853 (7.8%)	5291 (8.5%)	−232.2	2535.3	0.91 (0.88 − 0.95)
6) SOC: Vascular disorders	8359 (13.5%)	8864 (14.3%)	−290.1	4303.6	0.93 (0.91 − 0.96)
7) HLGT: Arteriosclerosis, stenosis, vascular insufficiency and necrosis	1626 (2.6%)	1833 (3.0%)	−108.8	864.6	0.88 (0.82 − 0.94)
8) PT: Intermittent claudication	275 (0.4%)	382 (0.6%)	−55.1	164.2	0.72 (0.61 − 0.83)
9) HLT: Peripheral embolism and thrombosis	361 (0.6%)	451 (0.7%)	−46.2	203.0	0.80 (0.69 − 0.91)
10) HLGT: Vascular disorders nec	729 (1.2%)	899 (1.5%)	−87.8	406.9	0.81 (0.73 − 0.89)
11) HLT: Peripheral vascular disorders nec	580 (0.9%)	712 (1.2%)	−68.3	322.9	0.81 (0.73 − 0.90)
12) PT: Peripheral vascular disorder	331 (0.5%)	429 (0.7%)	−50.5	190.0	0.77 (0.66 − 0.88)
13) SOC: Surgical and medical procedures	12889 (20.8%)	13571 (21.9%)	−440.6	6610.3	0.94 (0.91 − 0.96)
14) HLGT: Vascular therapeutic procedures	3397 (5.5%)	4299 (6.9%)	−490.5	1923.3	0.77 (0.74 − 0.81)
15) HLT: Arterial therapeutic procedures (excl aortic)	2610 (4.2%)	3470 (5.6%)	−460.6	1519.5	0.74 (0.70 − 0.78)
16) PT: Coronary artery bypass	1225 (2.0%)	1623 (2.6%)	−212.9	711.8	0.74 (0.69 − 0.80)
17) PT: Carotid endarterectomy	47 (<0.1%)	83 (0.1%)	−18.3	32.5	0.57 (0.40 − 0.80)
18) PT: Coronary revascularisation	533 (0.9%)	834 (1.3%)	−156.4	341.7	0.63 (0.57 − 0.70)
19) PT: Coronary angioplasty	470 (0.8%)	647 (1.0%)	−92.4	279.2	0.72 (0.64 − 0.81)
20) HLT: Vascular therapeutic procedures nec	881 (1.4%)	1118 (1.8%)	−127.7	499.6	0.77 (0.71 − 0.85)
21) PT: Peripheral revascularisation	433 (0.7%)	572 (0.9%)	−71.2	251.2	0.75 (0.67 − 0.85)
22) PT: Vascular operation	287 (0.5%)	423 (0.7%)	−74.7	177.4	0.66 (0.57 − 0.76)
23) HLT: Cardiac therapeutic procedures nec	716 (1.2%)	837 (1.4%)	−67.8	388.1	0.84 (0.76 − 0.93)
24) HLT: Therapeutic procedures nec	3087 (5.0%)	3336 (5.4%)	−136.0	1604.8	0.92 (0.87 − 0.96)
26) HLGT: Cardiac and vascular investigations (excl enzyme tests)	3560 (5.7%)	3999 (6.5%)	−260.0	1888.8	0.87 (0.83 − 0.91)
27) HLT: Vascular imaging procedures nec	654 (1.1%)	890 (1.4%)	−128.0	385.8	0.72 (0.65 − 0.79)
28) PT: Angiogram	367 (0.6%)	535 (0.9%)	−91.9	225.3	0.67 (0.58 − 0.76)
29) HLT: Cardiac imaging procedures	1568 (2.5%)	1840 (3.0%)	−154.3	851.7	0.83 (0.78 − 0.89)
30) PT: Arteriogram coronary	1377 (2.2%)	1636 (2.6%)	−145.6	753.0	0.82 (0.77 − 0.89)
36) SOC: Cardiac disorders	14878 (24.0%)	16400 (26.5%)	−951.8	7812.3	0.89 (0.87 − 0.91)
37) HLGT: Coronary artery disorders	9833 (15.9%)	11338 (18.3%)	−894.6	5287.9	0.84 (0.82 − 0.87)
38) HLT: Coronary artery disorders nec	1784 (2.9%)	2266 (3.7%)	−260.8	1011.9	0.77 (0.73 − 0.82)
39) PT: Coronary artery disease	1660 (2.7%)	2111 (3.4%)	−243.9	942.2	0.77 (0.72 − 0.82)
40) HLT: Ischaemic coronary artery disorders	9102 (14.7%)	10544 (17.0%)	−848.8	4907.5	0.84 (0.82 − 0.87)
41) PT: Angina unstable	1515 (2.4%)	1738 (2.8%)	−129.0	813.0	0.85 (0.80 − 0.91)
42) PT: Angina pectoris	5973 (9.6%)	6575 (10.6%)	−356.8	3135.0	0.89 (0.86 − 0.92)
43) PT: Acute myocardial infarction	647 (1.0%)	839 (1.4%)	−100.4	371.5	0.76 (0.69 − 0.84)
44) PT: Myocardial ischaemia	163 (0.3%)	238 (0.4%)	−38.6	100.2	0.68 (0.56 − 0.83)
45) PT: Myocardial infarction	2154 (3.5%)	3046 (4.9%)	−468.4	1299.4	0.70 (0.66 − 0.74)
46) HLGT: Heart failures	3075 (5.0%)	3372 (5.4%)	−160.8	1611.3	0.91 (0.86 − 0.95)
47) HLT: Heart failures nec	2760 (4.4%)	3001 (4.8%)	−129.5	1439.8	0.91 (0.87 − 0.96)
48) PT: Cardiac failure	1874 (3.0%)	2068 (3.3%)	−105.5	985.2	0.90 (0.84 − 0.96)
49) SOC: Nervous system disorders	12855 (20.7%)	13312 (21.5%)	−276.3	6536.8	0.96 (0.94 − 0.98)
50) HLGT: Central nervous system vascular disorders	3112 (5.0%)	3701 (6.0%)	−314.9	1702.8	0.83 (0.79 − 0.87)
51) HLT: Central nervous system haemorrhages and cerebrovascular accidents	2030 (3.3%)	2405 (3.9%)	−197.3	1108.6	0.84 (0.79 − 0.89)
52) PT: Cerebrovascular accident	828 (1.3%)	962 (1.6%)	−70.3	447.5	0.85 (0.78 − 0.94)
53) PT: Ischaemic stroke	901 (1.5%)	1184 (1.9%)	−145.8	521.2	0.76 (0.69 − 0.82)
54) HLT: Transient cerebrovascular events	1160 (1.9%)	1353 (2.2%)	−100.6	628.2	0.85 (0.79 − 0.92)
55) PT: Transient ischaemic attack	1139 (1.8%)	1329 (2.1%)	−99.0	616.9	0.85 (0.79 − 0.92)
57) PT: Hemiplegia	2 (<0.1%)	15 (<0.1%)	−6.5	4.2	0.21 (0.08 − 0.56)
HIGHER WITH STATIN					
25) SOC: Musculoskeletal and connective tissue disorders	22012 (35.5%)	21417 (34.6%)	442.6	10825.0	1.04 (1.02 − 1.06)
31) HLGT: Hepatobiliary investigations	1349 (2.2%)	1025 (1.7%)	164.2	593.3	1.32 (1.22 − 1.43)
32) HLT: Liver function analyses	1277 (2.1%)	937 (1.5%)	172.7	553.3	1.37 (1.26 − 1.49)
33) PT: Hepatic enzyme increased	203 (0.3%)	109 (0.2%)	47.3	78.0	1.83 (1.47 − 2.29)
34) PT: Alanine aminotransferase increased	684 (1.1%)	474 (0.8%)	105.5	289.4	1.44 (1.28 − 1.62)
35) PT: Aspartate aminotransferase increased	278 (0.4%)	199 (0.3%)	39.4	119.2	1.39 (1.16 − 1.67)
56) PT: Parosmia	18 (<0.1%)	2 (<0.1%)	8.0	5.0	4.93 (2.05 − 11.86)

FDR= False Discovery Rate (corrected using Mehrotra and Adewale method).

Figure 5.5: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (19 trials, 123,940 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among statin−allocated patients versus event rates among placebo−allocated patients (with values above the horizontal axis indicating that the event was more common among statin−allocated patients and values below the axis indicating that the event was less common among statin−allocated patients). 3970 such p−values are shown, corresponding to, 25 tests at the SOC level, 243 tests at the HLGT level, 877 tests at the HLT level and 2825 tests at the PT level. Of these tests 25 (0.6%) are significant after correction of the FDR at 5% level. These are indicated with closed circles and are enumerated (see Figure 5.6 for the identification and results of each of these 25 event categories). SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term. PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested.

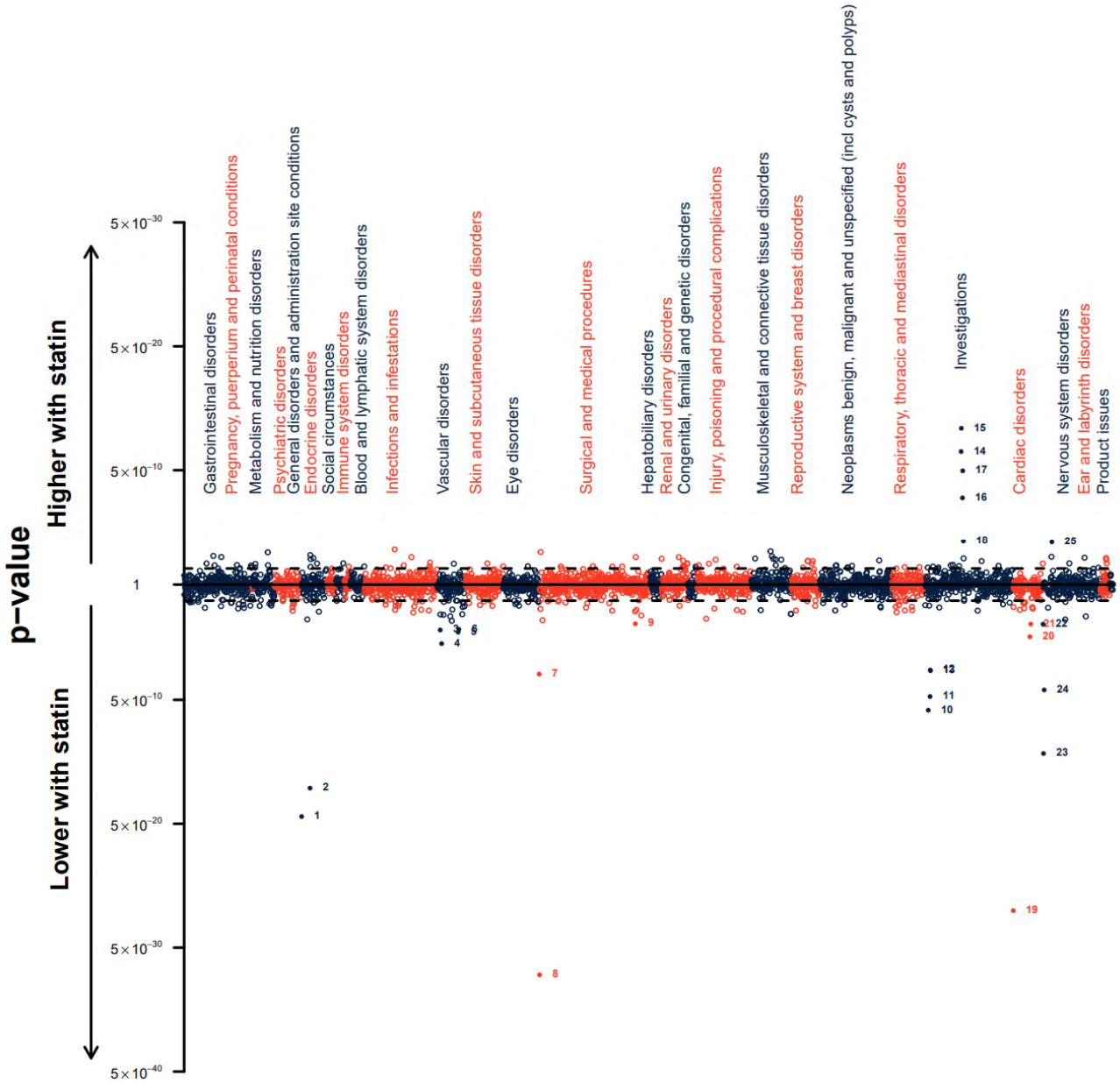
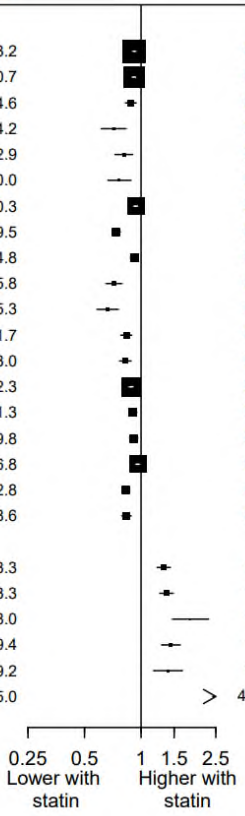


Figure 5.6: FDR significant effects of statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (19 trials, 123,940 randomised patients)

MedDRA Term	Events (%) Statin (n=62028)	Events (%) Placebo (n=61912)	Logrank O−E	Logrank V	RR (95% CI)
LOWER WITH STATIN					
1) SOC: General disorders and administration site conditions	23411 (37.7%)	24915 (40.2%)	−990.1	12053.2	0.92 (0.90 − 0.94)
2) HLGT: General system disorders nec	18661 (30.1%)	19843 (32.1%)	−825.1	9600.7	0.92 (0.90 − 0.94)
3) HLGT: Arteriosclerosis, stenosis, vascular insufficiency and necrosis	1626 (2.6%)	1833 (3.0%)	−108.8	864.6	0.88 (0.82 − 0.94)
4) PT: Intermittent claudication	275 (0.4%)	382 (0.6%)	−55.1	164.2	0.72 (0.61 − 0.83)
5) HLT: Peripheral vascular disorders nec	580 (0.9%)	712 (1.2%)	−68.3	322.9	0.81 (0.73 − 0.90)
6) PT: Peripheral vascular disorder	331 (0.5%)	429 (0.7%)	−50.5	190.0	0.77 (0.66 − 0.88)
7) SOC: Surgical and medical procedures	12889 (20.8%)	13571 (21.9%)	−440.6	6610.3	0.94 (0.91 − 0.96)
8) HLT: Arterial therapeutic procedures (excl aortic)	2610 (4.2%)	3470 (5.6%)	−460.6	1519.5	0.74 (0.70 − 0.78)
9) HLT: Therapeutic procedures nec	3087 (5.0%)	3336 (5.4%)	−136.0	1604.8	0.92 (0.87 − 0.96)
10) HLT: Vascular imaging procedures nec	654 (1.1%)	890 (1.4%)	−128.0	385.8	0.72 (0.65 − 0.79)
11) PT: Angiogram	367 (0.6%)	535 (0.9%)	−91.9	225.3	0.67 (0.58 − 0.76)
12) HLT: Cardiac imaging procedures	1568 (2.5%)	1840 (3.0%)	−154.3	851.7	0.83 (0.78 − 0.89)
13) PT: Arteriogram coronary	1377 (2.2%)	1636 (2.6%)	−145.6	753.0	0.82 (0.77 − 0.89)
19) SOC: Cardiac disorders	14878 (24.0%)	16400 (26.5%)	−951.8	7812.3	0.89 (0.87 − 0.91)
20) HLGT: Heart failures	3075 (5.0%)	3372 (5.4%)	−160.8	1611.3	0.91 (0.86 − 0.95)
21) HLT: Heart failures nec	2760 (4.4%)	3001 (4.8%)	−129.5	1439.8	0.91 (0.87 − 0.96)
22) SOC: Nervous system disorders	12855 (20.7%)	13312 (21.5%)	−276.3	6536.8	0.96 (0.94 − 0.98)
23) HLGT: Central nervous system vascular disorders	3112 (5.0%)	3701 (6.0%)	−314.9	1702.8	0.83 (0.79 − 0.87)
24) HLT: Central nervous system haemorrhages and cerebrovascular accidents	2030 (3.3%)	2405 (3.9%)	−197.3	1108.6	0.84 (0.79 − 0.89)
HIGHER WITH STATIN					
14) HLGT: Hepatobiliary investigations	1349 (2.2%)	1025 (1.7%)	164.2	593.3	1.32 (1.22 − 1.43)
15) HLT: Liver function analyses	1277 (2.1%)	937 (1.5%)	172.7	553.3	1.37 (1.26 − 1.49)
16) PT: Hepatic enzyme increased	203 (0.3%)	109 (0.2%)	47.3	78.0	1.83 (1.47 − 2.29)
17) PT: Alanine aminotransferase increased	684 (1.1%)	474 (0.8%)	105.5	289.4	1.44 (1.28 − 1.62)
18) PT: Aspartate aminotransferase increased	278 (0.4%)	199 (0.3%)	39.4	119.2	1.39 (1.16 − 1.67)
25) PT: Parosmia	18 (<0.1%)	2 (<0.1%)	8.0	5.0	4.93 (2.05 − 11.86)



FDR= False Discovery Rate (corrected using Mehrotra and Adewale method).

Figure 5.7: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (19 trials, 123,940 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among statin−allocated patients versus event rates among placebo−allocated patients (with values above the horizontal axis indicating that the event was more common among statin−allocated patients and values below the axis indicating that the event was less common among statin−allocated patients). 1486 such p−values are shown, corresponding to, 25 tests at the SOC level, 118 tests at the HLGT level, 294 tests at the HLT level and 1049 tests at the PT level. Of these tests 22 (1.5%) are significant after correction of the FDR at 5% level. These are indicated with closed circles and are enumerated (see Figure 5.8 for the identification and results of each of these 18 event categories). SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term. PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested. *PTs with 50 or more events and HLTs or HLGTs which report 50% or more extra events to level(s) below.

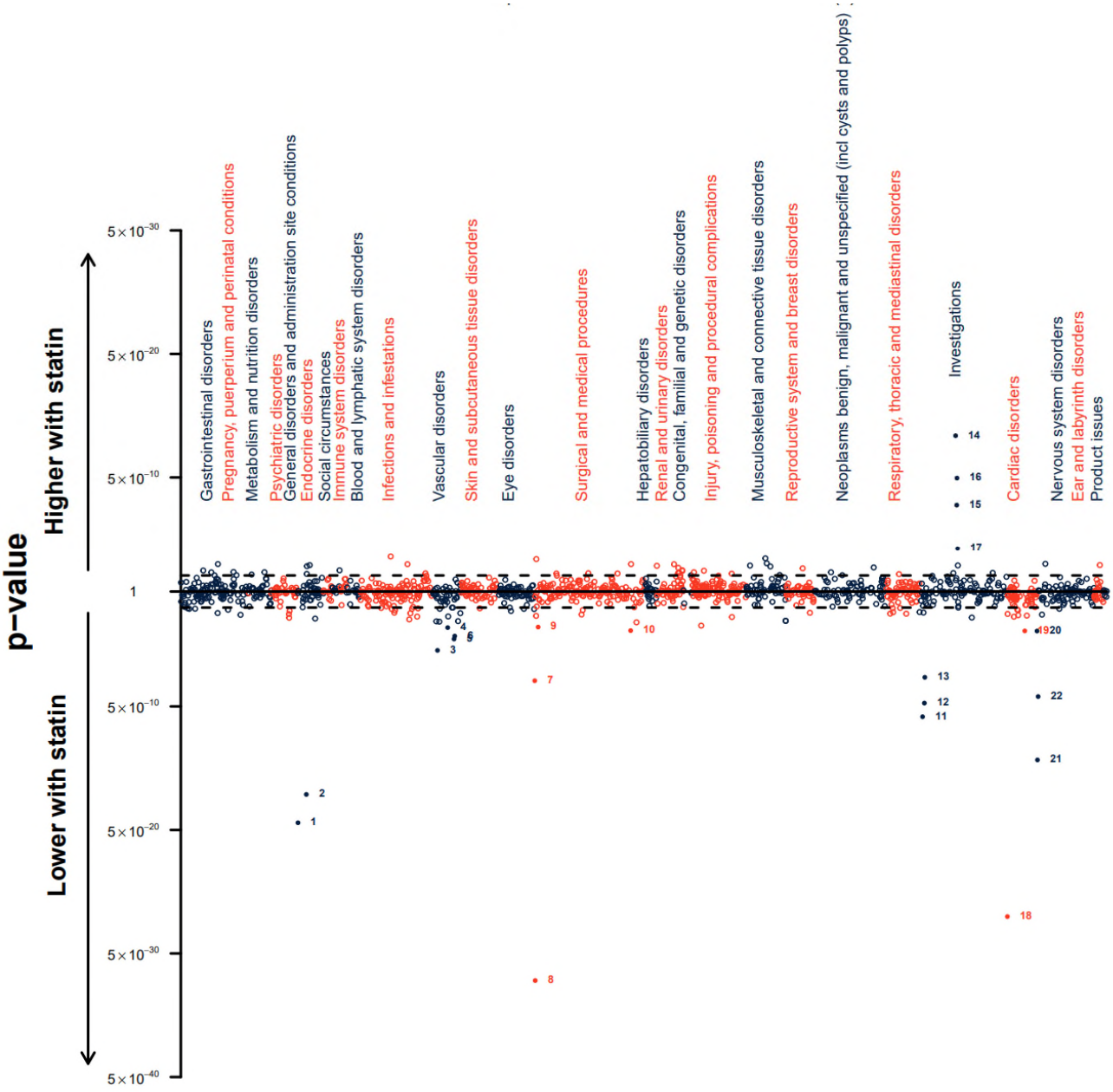
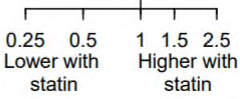


Figure 5.8: FDR significant effects of statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (19 trials, 123,940 randomised patients)

MedDRA Term	Events (%) Statin (n=62028)	Events (%) Placebo (n=61912)	Logrank O−E	Logrank V	RR (95% CI)
LOWER WITH STATIN					
1) SOC: General disorders and administration site conditions	23411 (37.7%)	24915 (40.2%)	−990.1	12053.2	0.92 (0.90 − 0.94)
2) HLGT: General system disorders nec	18661 (30.1%)	19843 (32.1%)	−825.1	9600.7	0.92 (0.90 − 0.94)
3) PT: Intermittent claudication	275 (0.4%)	382 (0.6%)	−55.1	164.2	0.72 (0.61 − 0.83)
4) HLT: Peripheral embolism and thrombosis	361 (0.6%)	451 (0.7%)	−46.2	203.0	0.80 (0.69 − 0.91)
5) HLT: Peripheral vascular disorders nec	580 (0.9%)	712 (1.2%)	−68.3	322.9	0.81 (0.73 − 0.90)
6) PT: Peripheral vascular disorder	331 (0.5%)	429 (0.7%)	−50.5	190.0	0.77 (0.66 − 0.88)
7) SOC: Surgical and medical procedures	12889 (20.8%)	13571 (21.9%)	−440.6	6610.3	0.94 (0.91 − 0.96)
8) HLT: Arterial therapeutic procedures (excl aortic)	2610 (4.2%)	3470 (5.6%)	−460.6	1519.5	0.74 (0.70 − 0.78)
9) PT: Carotid endarterectomy	47 (<0.1%)	83 (0.1%)	−18.3	32.5	0.57 (0.40 − 0.80)
10) HLT: Therapeutic procedures nec	3087 (5.0%)	3336 (5.4%)	−136.0	1604.8	0.92 (0.87 − 0.96)
11) HLT: Vascular imaging procedures nec	654 (1.1%)	890 (1.4%)	−128.0	385.8	0.72 (0.65 − 0.79)
12) PT: Angiogram	367 (0.6%)	535 (0.9%)	−91.9	225.3	0.67 (0.58 − 0.76)
13) PT: Arteriogram coronary	1377 (2.2%)	1636 (2.6%)	−145.6	753.0	0.82 (0.77 − 0.89)
18) SOC: Cardiac disorders	14878 (24.0%)	16400 (26.5%)	−951.8	7812.3	0.89 (0.87 − 0.91)
19) HLT: Heart failures nec	2760 (4.4%)	3001 (4.8%)	−129.5	1439.8	0.91 (0.87 − 0.96)
20) SOC: Nervous system disorders	12855 (20.7%)	13312 (21.5%)	−276.3	6536.8	0.96 (0.94 − 0.98)
21) HLGT: Central nervous system vascular disorders	3112 (5.0%)	3701 (6.0%)	−314.9	1702.8	0.83 (0.79 − 0.87)
22) HLT: Central nervous system haemorrhages and cerebrovascular accidents	2030 (3.3%)	2405 (3.9%)	−197.3	1108.6	0.84 (0.79 − 0.89)
HIGHER WITH STATIN					
14) HLT: Liver function analyses	1277 (2.1%)	937 (1.5%)	172.7	553.3	1.37 (1.26 − 1.49)
15) PT: Hepatic enzyme increased	203 (0.3%)	109 (0.2%)	47.3	78.0	1.83 (1.47 − 2.29)
16) PT: Alanine aminotransferase increased	684 (1.1%)	474 (0.8%)	105.5	289.4	1.44 (1.28 − 1.62)
17) PT: Aspartate aminotransferase increased	278 (0.4%)	199 (0.3%)	39.4	119.2	1.39 (1.16 − 1.67)



FDR= False Discovery Rate (corrected using Mehrotra and Adewale method).
*PTs with 50 or more events and HLTs or HLGTs which report 50% or more extra events to level(s) below.

6 Meta-analyses of the double-blind trials of more vs less intensive statin therapy

6.1 Summary

This chapter presents results of the meta-analyses of all AEs reported in 4 double-blind trials and 30,724 patients comparing more vs less intensive statin therapy. Results show that the use of methods correcting multiple hypotheses testing are required to overcome the problem of false positive findings. As anticipated, more intensive statin therapy was associated with a reduction in major cardiovascular events but also a reduction in other outcomes such as 'general system disorders' (including AEs of, for example, oedema, pain and discomfort, and chest pain), surgical or medical procedures, and renal therapeutic procedures. However, more intensive statin therapy was also associated with an increased risk of hepatobiliary investigations (with no clear evidence of such an effect being dose-related), myopathy and elevated blood creatine phosphokinase. The findings highlight the substantial cardiovascular and other medical benefits of high-intensity statin therapy for participants at high risk of such disease which greatly outweigh any identified risks.

6.2 Introduction

Studies have shown that a continuous positive relation between coronary disease risk and blood cholesterol exist with larger reductions in LDL cholesterol produce larger reductions in risk of cardiovascular disease (2, 18). High dose or newer more effective statins (eg, 40-80 mg atorvastatin or 10-20 mg rosuvastatin), which are typically used for the secondary prevention of major vascular events, are shown to reduce LDL cholesterol by about 50% with such reductions producing further reductions in the risk of major vascular disease (2).

In particular, more intensive statin therapy was shown to further reduce LDL cholesterol by 0.5 mmol/L which led to a further proportional reduction in major vascular events (MVE) by about 15% when compared with less intensive statin therpay (18). This proportional reduction in risk of MVE per 0.5 mmol/L reduction in LDL cholesterol was similar in various types of participant irrespective of baseline cholesterol levels, prior medical history or other patient characteristics (2, 18).

Some observational studies have shown that statins are associated with an increased risk of various AEs (eg, myalgia, and depression) leading people at high risk of vascular disease to stop or not start statin therapy (chapter 1.3.3). This chapter presents the meta-analyses of all the AEs with no prior knowledge (from randomised evidence) of an association with statin therapy in the 4 double-blind trials of more vs less statin therapy involving 30,724 participants.

6.3 Methods

6.3.1 Data

A data format based on the CDISC and SDTM version 3.2 was used to harmonise the IPD received from the 4 trials into a single dataset with a common format ready for analyses (see chapter 3.2.2). MedDRA version 20.0 was used to harmonise the AEs from the multiple trials and the various event coding formats into a single common language (see chapter 3.2.3).

6.3.2 Analysis

To identify any new (ie, previously undetected) effects of more intensive statin therapy and estimate whether any apparent effects of statin therapy may be dose-related, 4 double-blind trials of a more intensive vs a less intensive statin regimen involving 30,724 participants were considered as the analysis population. Analyses were to include all randomised participants irrespective of whether they received their allocated treatment (ie, intention-to-treat). Adverse

events reporting at least one event were included for analyses. Time to event analyses were performed.

6.3.2.1 Statistical analysis

Similar to the statistical analysis for the trials of statin vs placebo, the log-rank observed minus expected statistic (O–E) and its variance were calculated for the first occurrence of each outcome among participants randomly assigned into each trial (chapter 3.6.1).

In a meta-analysis of trials, the inverse-variance-weighted average of the log of the rate ratio (log RR) across all trials was calculated as S/V (with variance 1/V, and hence with a 95% CI of S/V ±1·96/√V), where S is the sum of (O−E) and V is the sum of variance over all trials (chapter 3.6.2). The nominal (ie, uncorrected for MHT) p-value was calculated from the comparison of the test statistic (ie, S/V) against a chi-square distribution (ie, χ^2) with 1 degree of freedom.

6.3.2.2 Multiple testing

The Mehrotra & Adewale MHT method was used to test the effect of more statin (when compared with less statin) therapy on all AEs examined within a single family and identify (through p-values) those effects (beneficial or harmful) which were statistically significant at the 5% (ie, 2p=0.05) FDR level (chapter 3.6.3).

6.3.2.3 Analyses of AEs

Five different analyses were performed, identical to those done for the trials of statin vs placebo in chapter 5: (i) analyses of all AEs blinded to (true) treatment allocation; (ii) unblinded analyses of all non-lipid related AEs (see **Appendix Table 2** for lipid related AE terms); (iii) unblinded analyses of all AEs with prior unknown association with statins (ie, AEs related to muscle, diabetes and known cardiovascular benefits, **Appendix Table 2**); (iv) unblinded analyses of all AEs with prior unknown association with statins after excluding very rare outcomes and avoiding irrelevant overlap in tests (with details for the hierarchical strategy

applied presented in chapter 4.4.4.1); (v) exploratory analyses examining the effects on all sublevels for AEs identified as FDR (in the more vs less statin trials).

Additional exploratory meta-analyses of the adverse events identified to be increased with statin or more intensive statin therapy at the 5% FDR level was performed to estimate if those harmful effects were dose related with statin therapy. The standard inverse-variance-weighted average log RR was calculated across the two meta-analyses of trials (ie, statin vs placebo and more vs less statin) as S/V (with variance 1/V, and hence with a 95% CI of S/V ±1·96/√V), where S is the sum of (O−E) and V is the sum of variance of the two separate meta-analyses.

6.4 Results

6.4.1 Data

Individual participant data were available from 4 randomised double-blind trials of a more versus a less intensive statin regimen resulting in a total of 15,390 patients assigned to receive more statin and 15,334 to receive less statin (**Table 6.1**). Overall, the mean age of the 30,724 participants was 62 (SD 9) years, mean BMI was 28.4 (SD 5.1), 19% of them were women and 15% had diabetes mellitus (**Table 6.1**). All had a previous history of coronary heart disease. During a weighted-average median trial follow-up of 4.9 years, a total of 185,177 AEs were reported, of which 6,210 were unique types of AEs (**Table 3.1**).

The distribution of the total number of (participants with) events across the three MedDRA hierarchical levels (ie, PT, HLT and HLGT) is presented in **Figure 6.1,** with additional details organised alphabetically by the 27 System Organ Classes being presented in **Table 6.2**. In total, there were 4,436 AE terms reported for at least one participant (including the 27 SOCs), 3,491 (78.7%) reported for at least 50 participants, 293 (6.6%) reported for between 50 and 99 participants, 535 (12.1%) reported for between 100 and 999 participants and 117 (2.6%) reported for 1000 or more participants. From the total of 4,436 (MedDRA) terms, 3,163 (71.3%) were PTs with median (IQR) number of events equal to 4 (1 – 18), 955 (21.5%) were

HLTs with median (IQR) number of events equal to 16 (4 – 86), 291 (6.6%) were HLGTs with median (IQR) number of events equal to 77 (14 – 434) and 27 (0.6%) were SOCs with median (IQR) number of events equal to 3,325 (817 – 7,010) (**Figure 6.1**).

6.4.2 Analyses of AEs blinded to treatment allocation

Figure 6.2 presents the nominal p-values for all AEs with at least 6 events, organised alphabetically by SOC and ordered by direction of the effect of statins (ie, higher or lower), when masking the true treatment received (ie more vs less statin). A total of 2,169 tests were performed with 27 SOCs, 215 HLGTs, 551 HLTs and 1,376 PTs being tested. When not correcting for MHT (ie, controlling for a 5% type I error), this approach resulted in a total of 115 (5.3%) false discoveries, while the use of the Mehrotra & Adewale method at 5% FDR level resulted in zero FPs.

6.4.3 Unblinded analyses of all non-lipid related AEs

Figure 6.3 presents the distribution of (nominal) p-values for all non-lipid related AEs with at least 6 events. A total of 2,151 tests were performed with 27 SOCs, 213 HLGTs, 546 HLTs and 1,365 PTs being analysed. The Mehrotra & Adewale method identified 6 SOCs ('Pregnancy, puerperium and perinatal conditions', 'Endocrine disorders', 'Social circumstances', 'Immune system disorders' 'Congenital, familial and genetic disorders', 'Reproductive system and breast disorder') which did not contribute further to the second step since the representative (ie, minimum) nominal p-values (0.105, 0.112, 0.413, 0.060, 0.096 and 0.054 respectively) were not FDR significant in the first step of this method. In total, 22 AEs were found significant at the 5% FDR level shown with closed circles and enumerated in **Figure 6.3,** with further details for the event rates and treatment effects shown in **Figure 6.4**.

Of those 22 AEs, 9 were shown to be reduced when receiving more intensive statin therapy, with most being dose related cardiovascular benefits of statin therapy. In particular, more intensive statins were associated with reduction in the HLT 'Arterial therapeutic procedures (excl aortic)' (1,479 [9.6%] vs. 1,752 [11.4%], RR 0.83, 95% CI 0.78 – 0.89), a reduction in the

HLT ‘Ischaemic coronary artery disorders’ (3,479 [22.6%] vs 3,740 [24.4%], RR 0.92, 95% CI 0.88 – 0.95) and a reduction in the PT ‘Myocardial infarction’ (922 [6.0%] vs 1,079 [7.0%], RR 0.85, 95% CI 0.78 – 0.92) (**Figure 6.4**). In addition, other identified benefits of more intensive statin therapy were a reduction in the HLGT ‘General system disorders NEC’ (5,098 [33.1%] vs 5,682 [37.1%], RR 0.87, 95% CI 0.84 – 0.91), and the SOC ‘Surgical and medical procedures’ (4,067 [26.4%] vs 4,333 [28.3%], RR 0.93, 95% CI 0.77 – 0.97) (**Figure 6.4**).

13 AE terms were found to be FDR-significantly higher among participants allocated more intensive statin therapy. These included an increased risk of the HLT ‘Myopathies’ (63 [0.4%] vs 13 [<0.1%], RR 3.74, 95% CI 2.38 – 5.86) and the PT of ‘Blood creatine phosphokinase increased’ (223 [1.4%] vs 146 [1.0%], RR 1.51, 95% CI 1.24 – 1.86) (**Figure 6.4**). In addition, similar to the results in the trials of statin vs placebo, more intensive statin regimens were associated with an increased risk of the HGLT ‘Hepatobiliary investigations’ (502 [3.3%] vs 253 [1.6%], RR 1.95, 95% CI 1.69 – 2.25) including further sub-classifications (ie, liver function analyses, transaminases increased, hepatic enzyme increased, alanine and aspartate aminotransferase increased) (**Figure 6.4**).

6.4.4 Unblinded analyses of all AEs with prior unknown association with statins

Figure 6.5 presents the (nominal) p-values for AEs with at least 6 events except those related to muscle, diabetes and the known cardiovascular benefits of statins. A total of 2,084 tests were performed with 25 SOCs, 208 HLGTs, 539 HLTs and 1,312 PTs being tested. Seven SOCs were not declared FDR significant in the first step of the Mehrotra & Adewale method, the same 6 as in section 6.4.3 and the SOC ‘Hepatobiliary disorders’ (minimum P=0.049), with none of the terms reported within those SOCs being analysed further. 14 AEs were found to be FDR significant (**Figures 6.5** and **6.6**) with all such AEs already detected in previous analyses in section 6.4.3.

6.4.5 Unblinded analyses of all AEs with prior unknown association with statins after excluding very rare outcomes and avoiding irrelevant overlap in tests

Figure 6.7 presents the (nominal) p-values for all AEs examined in the analyses above (section 6.4.4) after applying a testing strategy which excluded very rare AEs and avoided irrelevant overlap in tests performed across the higher levels of MedDRA. A total of 645 tests were performed with 25 SOCs, 77 HLGTs, 132 HLTs and 411 PTs being tested. The Mehotra & Adewale method identified the same 7 SOCs as in section 6.4.4 which did not contribute any AE terms to the second step of the MHT analyses. 14 AEs were found to be FDR significant as shown in **Figure 6.7** with further details for the event rates and treatment effects shown in **Figure 6.8**.

Compared with the previous analyses (sections 6.4.3 and 6.4.4), two additional beneficial effects of more intensive statin therapy were found FDR significant at 5% level, the effect on the HLT 'Renal therapeutic procedures' (30 [0.2%] vs 57 [0.4%], RR 0.54, 95% CI 0.35 – 0.82) and the PT 'Blood urea increased' (15 [0.1%] vs 35 [0.2%], RR 0.45, 95% CI 0.26 – 0.78) (**Figure 6.8**). No new harmful effects of more intensive statin therapy were found to be FDR significant.

6.4.6 Further exploratory analyses

Further analyses for the effects of more intensive statin on all sublevels of each identified FDR significant AE reported in unblinded analyses above showed that the beneficial effect of statin on the HLGT 'Coronary artery disorders' except for myocardial infarction was driven mostly by the PT 'Angina pectoris' (2,362 [15.3%] vs 2,489 [16.2%], RR 0.94, 95% CI 0.89 – 0.99, P=0.026). The reduction in the HLT of 'Renal therapeutic procedures' included reductions in the PT 'Nephrostomy' (5 vs 17), the PT 'Renal stone removal' (16 vs 24) and the PT 'Nephrectomy' (11 vs 15).

In a meta-analyses of the AEs identified to be increased with statin or more statin therapy at the 5% FDR level (ie, exploring if such effects were dose-related), for AEs identified as FDR significant in both groups of studies, the RRs were higher in trials of more vs less statin therapy than in trials of statin vs placebo (**Figure 6.9**). For example, the RR (95% CI) for the HLGT 'Hepatobiliary investigations' was 1.32 (1.22 − 1.43) in trials of statin vs placebo and 1.95 (1.69 − 2.25) in trials of more vs less statin with similar differences reported in sub-classifications (ie, liver function analyses, hepatic enzyme increased, alanine and aspartate aminotransferase increased). For the two AEs identified as FDR significant only in trials of statin vs placebo, no events of PT 'Parosmia' were reported by participants in trials of more vs less statin, while the RRs (95% CI) for SOC 'Musculoskeletal and connective tissue disorders' were similar in both groups of trials (ie, 1.04 (1.02 − 1.06) in statin vs placebo and 1.03 (1.00 − 1.07) in more vs less statin) (**Figure 6.9**). For AEs identified as FDR significant only in trials of more vs less statin therapy, the RRs were higher in those trials compared to trials of statin vs placebo (**Figure 6.9**). For example, the RR (95% CI) for the HLGT 'Enzyme investigations NEC' was 1.09 (0.99 − 1.20) in trials of statin vs placebo and 1.45 (1.21 − 1.73) in trials of more vs less statin. Also, the RR (95% CI) for the PT 'Blood creatine phosphokinase increased' was 1.08 (0.97 − 1.20) in trials of statin vs placebo and 1.51 (1.24 − 1.86) in trials of more vs less statin (**Figure 6.9**).

6.5 Discussion

This chapter aimed to analyse all AEs reported in the 4 double-blind trials of more vs less intensive statin therapy (involving 30,724 participants) to identify any real, previously unknown, effects of more intensive statin therapy and explore if such effects may be dose-related.

In total, these analyses examined the effects of more vs less intensive statin regimens for 2,169 AEs reported across the various levels of MedDRA (ie, PT, HLT, HLGT and SOC). As

shown from the blinded analyses of all AEs (ie, participants were randomly allocated to more or less statin arm), a very large number of FPs is expected (ie, 115 FPs) when not accounting (in the analyses) for the problem of multiple hypotheses testing, with this number being close zero when using the Mehrotra & Adewale method at the 5% FDR level.

The main aim of the first set of the unblinded analyses (ie, reporting true treatment effects) on all non-lipid AEs was to confirm (ie, as a method of validation) that the results produced using the MedDRA medical dictionary and the CTT harmonisation tool are in agreement with the already known (from previous CTT work) effects of more intensive statin therapy. Results showed that more intensive statin therapy further reduces the risk of numerous types of AEs related to cardiovascular disease including the (already) known effects of coronary arterial procedures by 17%, ischaemic coronary artery disorders by 8%, and myocardial infarction by 15% when taken systematically for a follow-up period of 5 years. Similar to the results from the trials of statin vs placebo, the above findings (ie, identifying all known effects of more intensive statin therapy) suggest that the methodology used for harmonisation of outcomes can be applied to detect any new effects of more intensive statin therapy.

Analyses showed that allocation to more intensive statin therapy (when compared with less intensive regimens) reduced the risk of the coronary artery disorders (ie, including angina pectoris and unstable angina) by about 8% (ie, RR 0.92, 95% CI 0.88 − 0.96), the general system disorders NEC by 13% (ie, RR 0.87, 95% CI 0. .84 − 0.91) with such a beneficial effect found across various sub-classification of such AE within the MedDRA tree (ie, oedema, general signs and symptoms, pain and discomfort and chest pain). Also, more intensive statin therapy was found to prevent about 2 out of 100 treated patients over 5 years from any surgical and medical procedures (ie, RR 0.93, 95% CI 0.89 − 0.97).

Further analyses removing already known effects or very rare outcomes plus any irrelevant overlap in tests (sections 6.4.4 and 6.4.5) showed that more intensive statin therapy reduced the risk of the rarely reported AE of ‘renal therapeutic procedures’ (ie, nephrostomy,

nephrectomy and stone removal) by about half (ie, RR 0.54, 95% CI 0.35 − 0.82). In addition, allocation to more intensive statin therapy reduce the (very low) risk of increased blood urea by more than 50% (ie, RR 0.45, 95% CI 0.26− 0.78).

More intensive statin regimens were shown to cause very few additional adverse effects compared with less intensive regimens. In particular, allocation to more intensive statin therapy nearly doubled the risk of hepatobiliary investigations (ie, RR 1.95, 95% CI 1.69 − 2.25) which comprised of AEs related to liver function (ie, hepatic enzyme increase, transaminases, alanine and aspartate aminotransferease increased). The risk of the very rare adverse event of myopathy (ie, reported by about 2 in 1,000 participants) was shown to be about 4 times higher when taking more intensive statin therapy, which was very similar to the result reported in the recently published analyses for effects of statin on all muscle symptoms (122). In addition, allocation to more intensive statin doubled the risk of increasing the levels blood creatine phosphokinase (ie, RR 1.51, 95% CI 1.24 − 1.86). No events of parosmia (or similar AE types) were reported in any of the 4 more vs less statin trials, probably due to the very low reporting rate.

To explore if the effects of statin therapy were dose-related on the AEs identified as being FDR significant from the two analyses (ie, statin vs placebo and more vs less intensive statin regimens), a pooled meta-analysis of those AEs was performed (section 6.4.6). Analyses showed no clear evidence that the increase in risk for hepatobiliary investigations is dose related (ie, RRs of 1.32% vs 1.95% with standard and more intensive statin respectively) with similar findings reported in its sub-classifications. None of the other FDR significant AEs were shown to be dose-related (except for the outcome of myopathy and musculoskeletal and connective tissue disorders for which more detailed findings are reported elsewhere (122)). Additional analyses are required to estimate the treatment effect for particular type of statins (ie, low, moderate and high intensive statin) and detect if the effects on hepatobiliary investigations are dose-related.

As highlighted in chapter 5.5, these analyses have important advantages over previous studies reporting adverse effects of statin or more statin therapy, such as the avoidance of biases from non-systematic reporting of adverse events (ie, only large-scale, randomised, double-blind trials included in analyses) and any potential bias arising from the selected publication of results or various medical coding methodologies used in the individual trials (ie, patient-level data with all AEs being coded to a common medical dictionary).

As for the trials of statin vs placebo, a limitation of these analyses is the impact (if any) on the absolute risk and treatment effect from the laboratory abnormalities (ie, high values of transaminases, alanine and aspartate aminotransferease) reported in the laboratory datasets when combined with the current AE dataset. In addition, more detailed analyses might be needed to estimate the effects of statins on a set of pre-defined (blinded to treatment allocation) endpoints for hepatobiliary disease and to determine if such effects differ across various sub-populations.

In summary, high-intensity statin regimens are recommended for patients at elevated risk of vascular events. The results presented in this chapter confirm the wide range of cardiovascular benefits as well as identifying some new beneficial effects of more intensive statin therapy. However, high-intensity statin therapy was also associated with an increased risk of several outcomes including hepatobiliary investigations (ie, liver function analyses including laboratory tests) and myopathy. Such adverse effects are minimal compared with the beneficial effects of high intensity statin for preventing cardiovascular diseases however.

Table 6.1: Characteristics of the participating trials

Study	Year of publication of primary results	Number of patients	Treatment comparison (mg/day)	Median follow-up (years)	Baseline LDL-C (mmol/L)	Baseline Age (years)	Women (%)	White/Caucasian (%)	History of Diabetes (%)	Baseline Body Mass Index (kg/m2)
PROVE-IT	2004	4162	A80 vs P40	2.1	2.6 (0.7)†	58 (11)	911 (22%)	3776 (91%)	762 (18%)	29.5 (5.7)
A to Z	2004	4497	S40 then S80 vs placebo then S20	2.0	2.1 (0.7) †	60 (11)	1100 (24%)	3825 (85%)	1059 (24%)	27.6 (4.8)
TNT	2005	10001	A80 vs A10	5.0	2.5 (0.5)	61 (9)	1902 (19%)	9410 (94%)	1501 (15%)	28.8 (6.1)
SEARCH	2010	12064	S80 vs S20	7.0	2.5 (0.6)	64 (9)	2052 (17%)	11854 (98%)	1267 (11%)	28.1 (4.1)
Subtotal 4 trials	**-**	**30724**	**-**	**4.9**	**2.5 (0.6)**	**62 (9)**	**5965 (19%)**	**28865 (94%)**	**4589 (15%)**	**28.4 (5.1)**

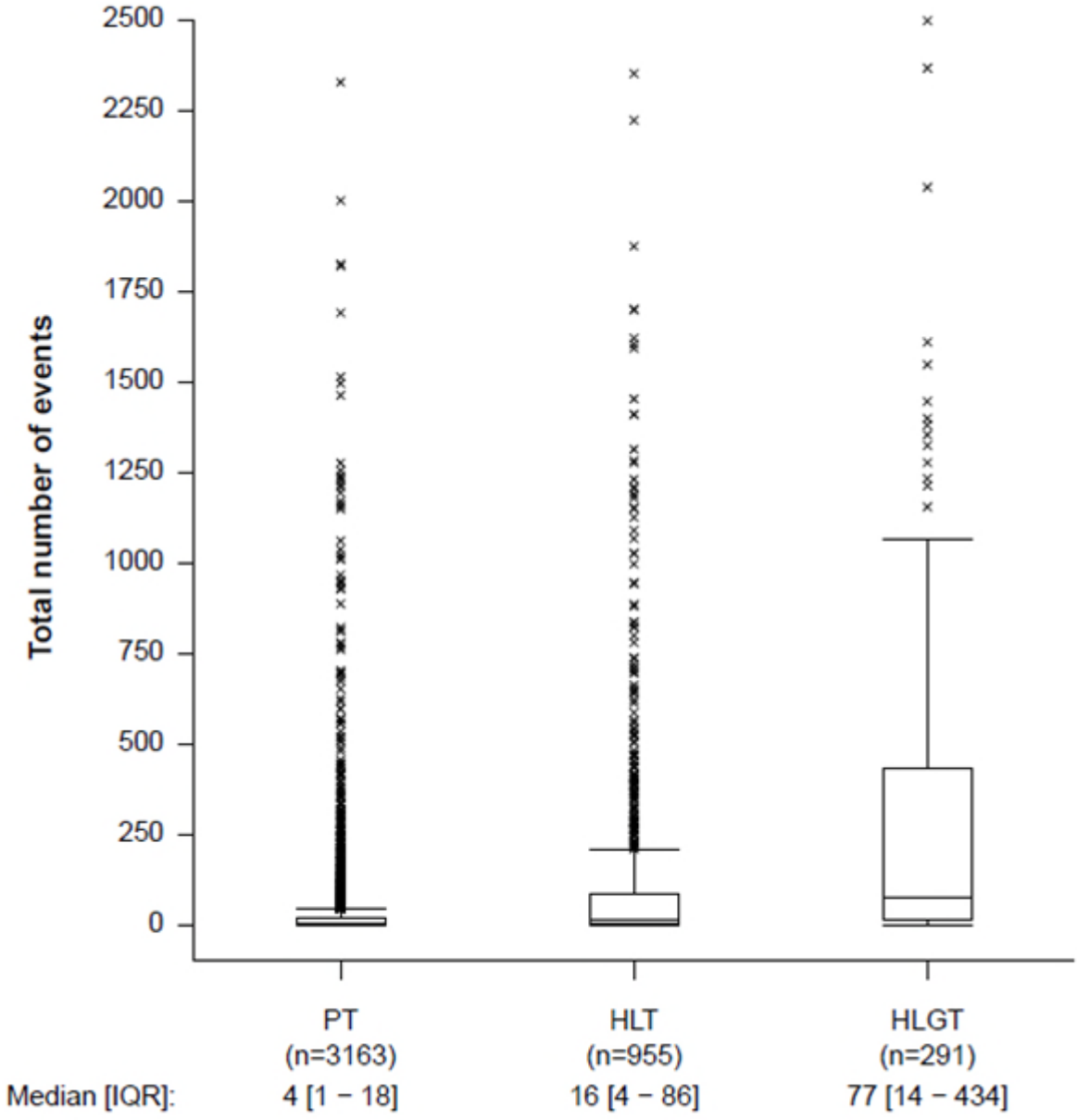
S=Simvastatin; P=Pravastatin; A=Atorvastatin; Age, LDL-C and BMI presented as mean (SD); NA=Not available; PROVE-IT=Pravastatin or Atorvastatin Evaluation and Infection Therapy; A to Z=Aggrastat to Zocor; TNT=Treating to New Targets; SEARCH=Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine; If data was NA then denominator did not contribute to weighted subtotal. †These two trials did not have active run-in periods; the values shown are the estimated on-treatment LDL cholesterol concentrations in the standard statin group.

Table 6.2: Distribution of the total number of events across the three analysable MedDRA hierarchical levels, organized alphabetically by System Organ Class (4 trials, 30,724 randomised patients)

	Number of tests at each sublevel with total number of events between <50| 50-99|100-999|>1000		
System Organ Class (Events)	**Higher Level Group Term**	**Higher Level Term**	**Preferred Term**
Blood and lymphatic system disorders (1,600)	4|4|0|1	16|3|1|1	41|3|0|1
Cardiac disorders (10,085)	0|1|3|3	12|1|8|3	95|8|15|2
Congenital, familial and genetic disorders (100)	12|0|0|0	22|0|0|0	32|0|0|0
Ear and labyrinth disorders (882)	2|0|3|0	5|0|3|0	24|3|2|0
Endocrine disorders (466)	5|0|1|0	11|2|1|0	24|1|1|0
Eye disorders (1,746)	3|1|7|0	26|5|5|0	109|4|4|0
Gastrointestinal disorders (7,989)	3|0|14|2	44|9|19|5	208|16|25|2
General disorders and administration site conditions (12,413)	2|0|3|2	17|1|5|5	79|3|13|5
Hepatobiliary disorders (751)	1|0|2|0	10|2|1|0	45|1|2|0
Immune system disorders (602)	3|0|1|0	7|1|2|0	23|1|2|0
Infections and infestations (8,844)	5|0|2|2	36|6|12|4	255|16|25|3
Injury, poisoning and procedural complications (4,336)	3|1|1|2	30|4|8|1	176|17|12|0
Investigations (7,635)	6|2|11|3	41|9|18|3	252|16|24|1
Metabolism and nutrition disorders (3,831)	8|1|3|1	25|1|7|1	57|1|8|1
Musculoskeletal and connective tissue disorders (13,082)	2|1|4|3	10|7|11|5	106|12|14|7
Neoplasms benign, malignant and unspecified (incl cysts and polyps) (3,467)	25|5|8|0	100|8|7|0	238|10|8|0
Nervous system disorders (6,385)	9|1|5|3	40|3|13|2	189|10|15|2
Pregnancy, puerperium and perinatal conditions (6)	2|0|0|0	3|0|0|0	3|0|0|0
Product issues (25)	2|0|0|0	4|0|0|0	7|0|0|0
Psychiatric disorders (2,767)	11|3|3|1	30|2|3|1	73|3|3|1
Renal and urinary disorders (2,305)	3|1|3|1	14|2|7|0	85|9|9|0
Reproductive system and breast disorders (1,772)	6|3|4|0	26|2|5|0	98|1|6|0
Respiratory, thoracic and mediastinal disorders (5,365)	1|0|5|2	13|2|11|2	116|9|13|2
Skin and subcutaneous tissue disorders (3,335)	1|0|6|1	26|6|8|1	108|7|11|0
Social circumstances (30)	4|0|0|0	7|0|0|0	13|0|0|0
Surgical and medical procedures (8,400)	3|2|9|4	56|10|21|1	171|19|24|2
Vascular disorders (4,751)	1|2|6|1	15|3|8|1	88|6|6|1
Total number of tests	**127|28|104|32**	**646|89|184|36**	**2715|176|242|30**

MedDRA=Medical Dictionary for Regulatory Activities. MedDRA V20.0 was used to code and analyse the adverse event data. System organ class is shown in alphabetical order.

Figure 6.1: Distribution of total number of events across the MedDRA hierarchical levels (4 trials, 30,724 randomised patients)



PT= Preferred Term, HLT= Higher Level Term, HLGT= Higher Level Group Term. There were 5 PTs, 10 HLTs and 15 HLGTs which reported a total number of events greater than 2,500.

For the 27 SOCs, the median (IQR) for the total number of events was equal to 3,325 (817 – 7,010).

Figure 6.2: All adverse events; BLINDED to treatment allocation (4 trials, 30,724 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among patients allocated more statin versus event rates among patients allocated less statin (with values above the horizontal axis indicating that the event was more common among patients allocated more statin and values below the axis indicating that the event was less common among patients allocated more statin). 2169 such p−values are shown, corresponding to, 27 tests at the SOC level, 215 tests at the HLGT level, 551 tests at the HLT level and 1376 tests at the PT level. Of these tests 115 (5.3%) would be considered significant at the 5% level if no correction for multiple testing is applied while 0 (0%) would be considered significant after correction for the FDR at 5% level. SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term, PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested.

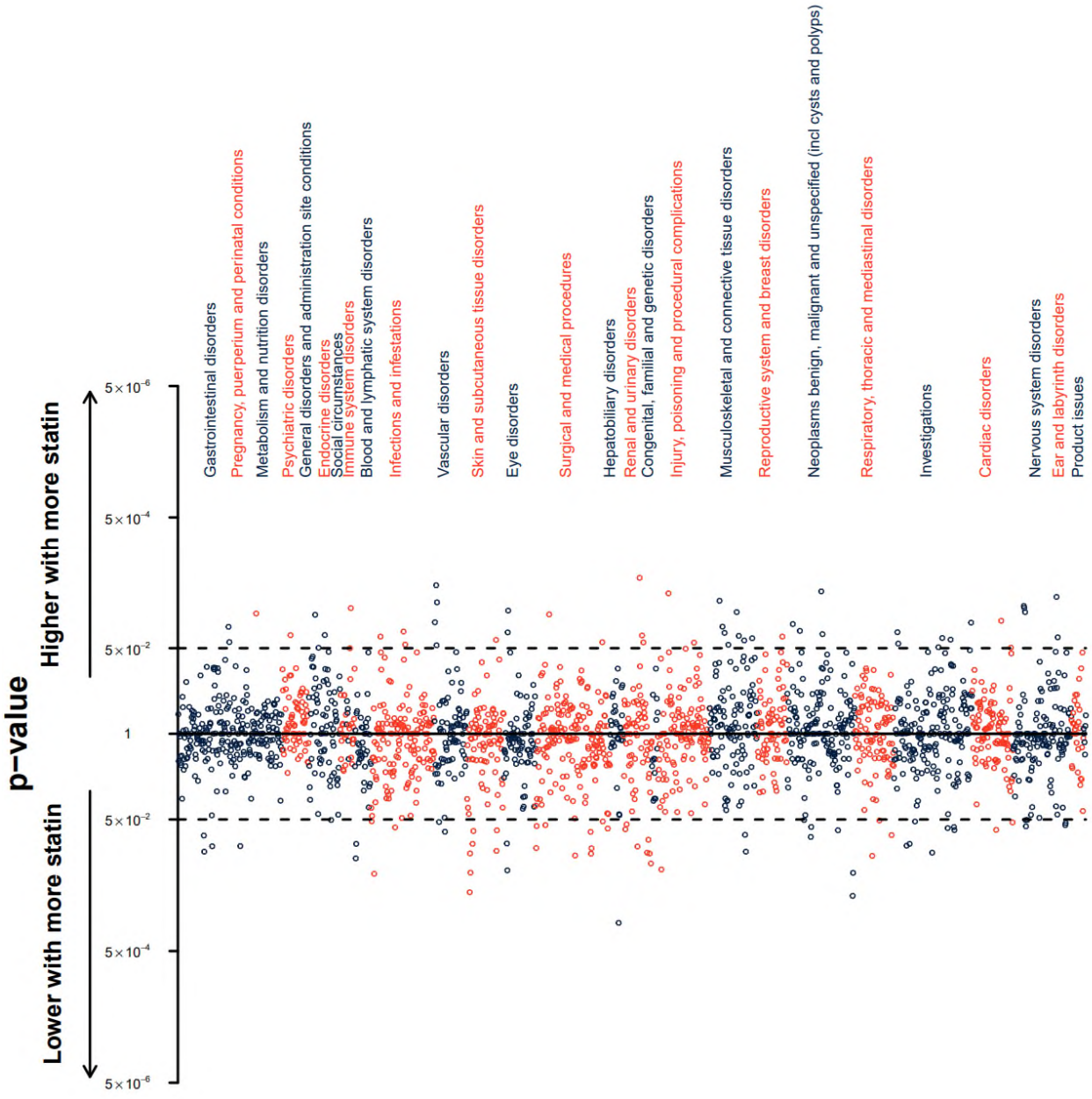


Figure 6.3: UNBLINDED analyses of all non lipid−related adverse events (4 trials, 30,724 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among patients allocated more statin versus event rates among patients allocated less statin (with values above the horizontal axis indicating that the event was more common among patients allocated more statin and values below the axis indicating that the event was less common among patients allocated more statin). 2151 such p−values are shown, corresponding to, 27 tests at the SOC level, 213 tests at the HLGT level, 546 tests at the HLT level and 1365 tests at the PT level. Of these tests 22 (1%) are significant after correction of the FDR at 5% level. These are indicated with closed circles and are enumerated (see Figure 6.4 for the identification and results of each of these 22 event categories). SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term. PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested.

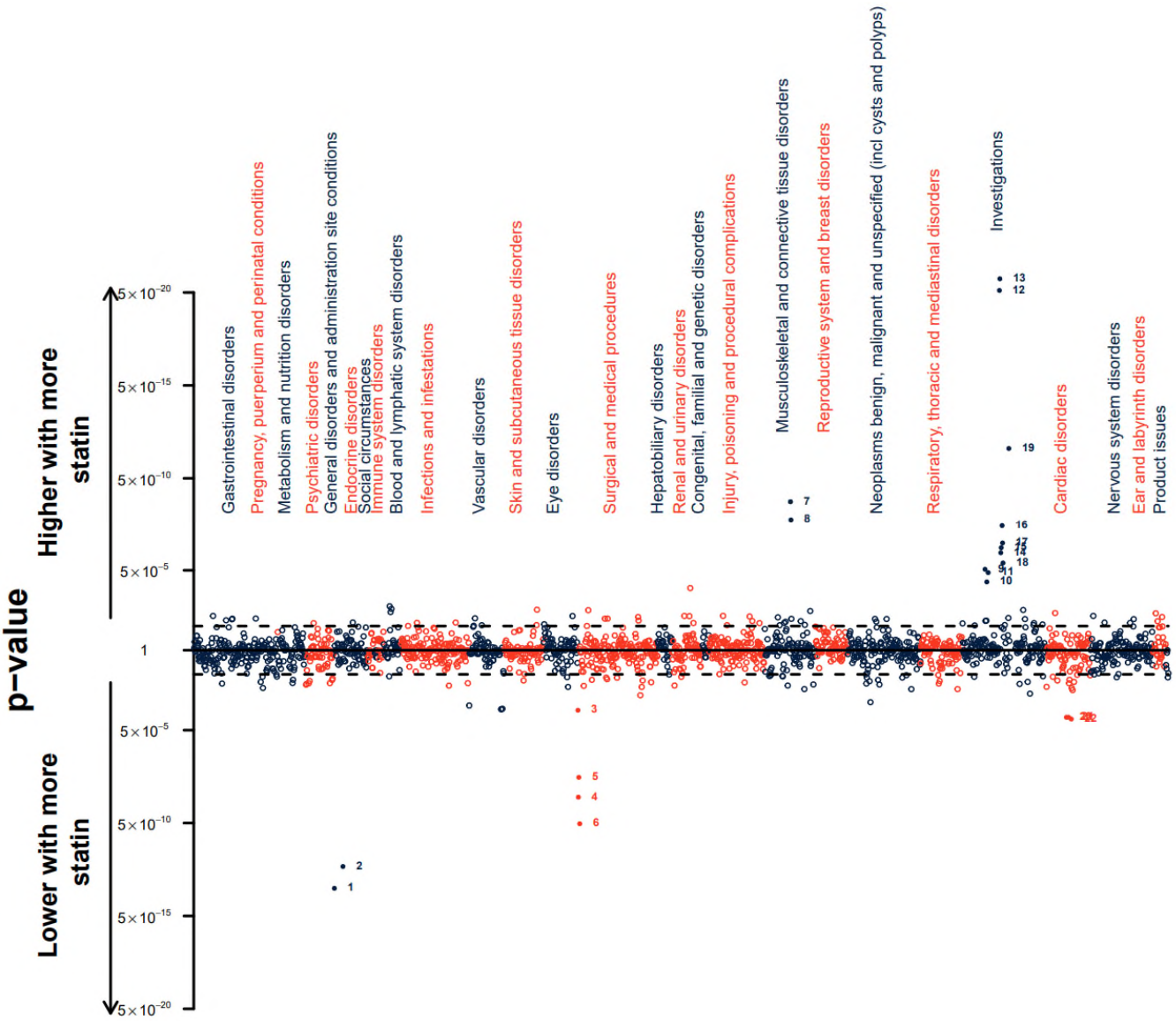
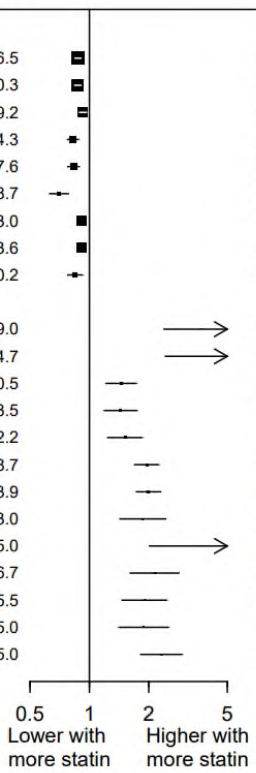


Figure 6.4: FDR significant effects of more statin therapy on non- lipid-related adverse events (4 trials, 30,724 randomised patients)

MedDRA Term	Events (%) More Statin (n=15390)	Events (%) Less Statin (n=15334)	Logrank O−E	Logrank V	RR (95% CI)
LOWER WITH MORE STATIN					
1) SOC: General disorders and administration site conditions	5897 (38.3%)	6516 (42.5%)	−410.9	3096.5	0.88 (0.85 − 0.91)
2) HLGT: General system disorders nec	5098 (33.1%)	5682 (37.1%)	−363.9	2690.3	0.87 (0.84 − 0.91)
3) SOC: Surgical and medical procedures	4067 (26.4%)	4333 (28.3%)	−157.5	2099.2	0.93 (0.89 − 0.97)
4) HLGT: Vascular therapeutic procedures	1614 (10.5%)	1924 (12.5%)	−169.4	884.3	0.83 (0.77 − 0.88)
5) HLT: Arterial therapeutic procedures (excl aortic)	1479 (9.6%)	1752 (11.4%)	−149.5	807.6	0.83 (0.78 − 0.89)
6) PT: Percutaneous coronary intervention	515 (3.3%)	720 (4.7%)	−109.5	308.7	0.70 (0.63 − 0.78)
20) HLGT: Coronary artery disorders	3518 (22.9%)	3779 (24.6%)	−156.6	1823.0	0.92 (0.88 − 0.96)
21) HLT: Ischaemic coronary artery disorders	3479 (22.6%)	3740 (24.4%)	−155.8	1803.6	0.92 (0.88 − 0.96)
22) PT: Myocardial infarction	922 (6.0%)	1079 (7.0%)	−83.2	500.2	0.85 (0.78 − 0.92)
HIGHER WITH MORE STATIN					
7) HLT: Myopathies	63 (0.4%)	13 (<0.1%)	25.0	19.0	3.74 (2.38 − 5.86)
8) PT: Myopathy	50 (0.3%)	9 (<0.1%)	20.5	14.7	4.03 (2.42 − 6.71)
9) HLGT: Enzyme investigations nec	286 (1.9%)	196 (1.3%)	44.8	120.5	1.45 (1.21 − 1.73)
10) HLT: Skeletal and cardiac muscle analyses	245 (1.6%)	169 (1.1%)	37.7	103.5	1.44 (1.19 − 1.75)
11) PT: Blood creatine phosphokinase increased	223 (1.4%)	146 (1.0%)	38.3	92.2	1.51 (1.24 − 1.86)
12) HLGT: Hepatobiliary investigations	502 (3.3%)	253 (1.6%)	126.2	188.7	1.95 (1.69 − 2.25)
13) HLT: Liver function analyses	493 (3.2%)	243 (1.6%)	126.7	183.9	1.99 (1.72 − 2.30)
14) PT: Liver function test abnormal	139 (0.9%)	73 (0.5%)	33.0	53.0	1.87 (1.42 − 2.44)
15) PT: Transaminases increased	48 (0.3%)	12 (<0.1%)	18.1	15.0	3.34 (2.01 − 5.54)
16) PT: Hepatic enzyme increased	129 (0.8%)	58 (0.4%)	35.6	46.7	2.14 (1.61 − 2.85)
17) PT: Alanine aminotransferase increased	147 (1.0%)	75 (0.5%)	35.7	55.5	1.90 (1.46 − 2.47)
18) PT: Aspartate aminotransferase increased	119 (0.8%)	61 (0.4%)	28.6	45.0	1.89 (1.41 − 2.53)
19) PT: Laboratory test abnormal	184 (1.2%)	76 (0.5%)	54.6	65.0	2.32 (1.82 − 2.95)



FDR= False Discovery Rate (corrected using Mehrotra and Adewale method).

Figure 6.5: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (4 trials, 30,724 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among patients allocated more statin versus event rates among patients allocated less statin (with values above the horizontal axis indicating that the event was more common among patients allocated more statin and values below the axis indicating that the event was less common among patients allocated more statin). 2084 such p−values are shown, corresponding to, 25 tests at the SOC level, 208 tests at the HLGT level, 539 tests at the HLT level and 1312 tests at the PT level. Of these tests 14 (0.7%) are significant after correction of the FDR at 5% level. These are indicated with closed circles and are enumerated (see Figure 6.6 for the identification and results of each of these 14 event categories). SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term. PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested.

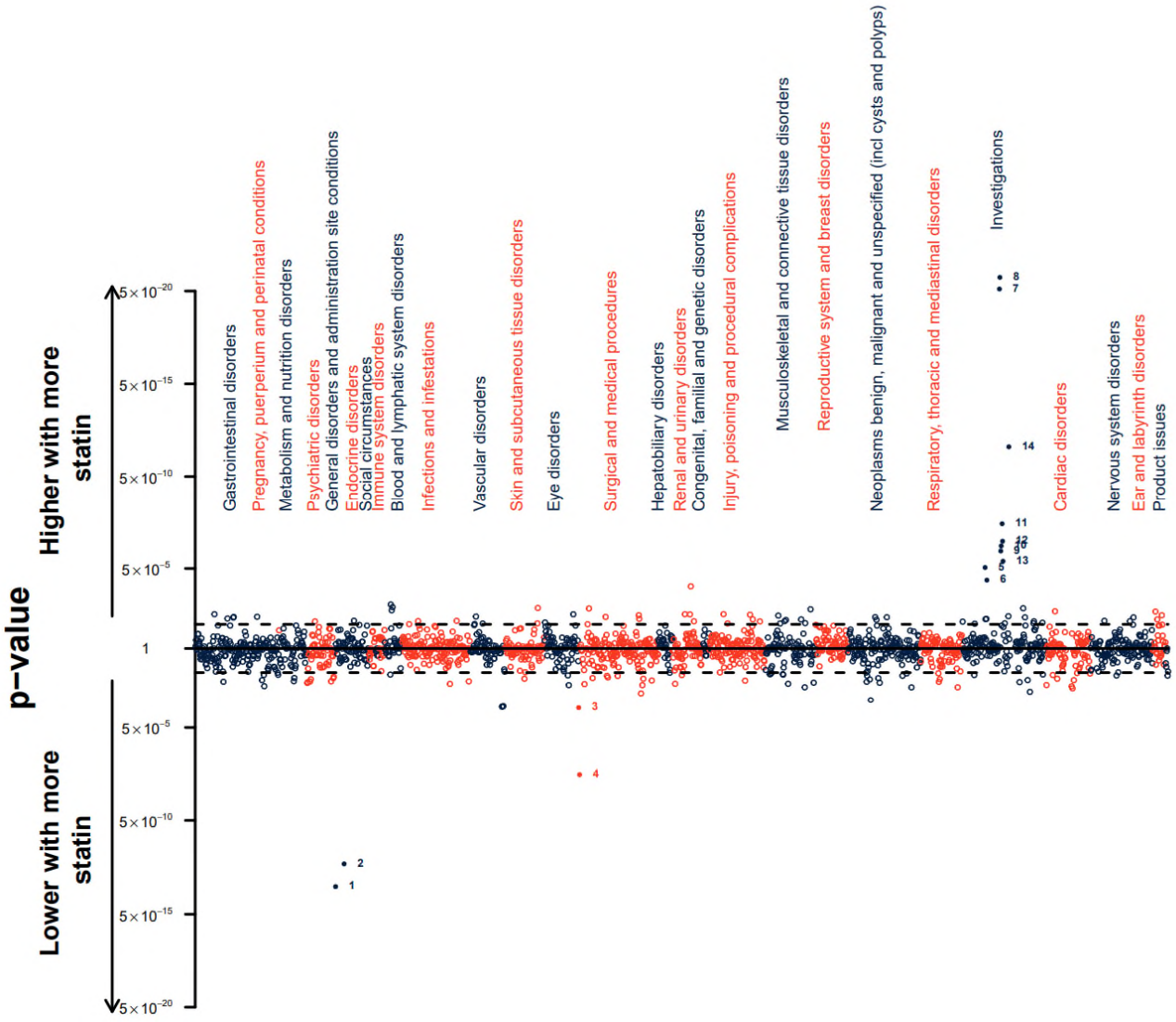
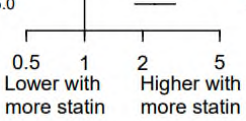


Figure 6.6: FDR significant effects of more statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins (4 trials, 30,724 randomised patients)

MedDRA Term	Events (%) More Statin (n=15390)	Events (%) Less Statin (n=15334)	Logrank O−E	Logrank V	RR (95% CI)
LOWER WITH MORE STATIN					
1) SOC: General disorders and administration site conditions	5897 (38.3%)	6516 (42.5%)	−410.9	3096.5	0.88 (0.85 − 0.91)
2) HLGT: General system disorders nec	5098 (33.1%)	5682 (37.1%)	−363.9	2690.3	0.87 (0.84 − 0.91)
3) SOC: Surgical and medical procedures	4067 (26.4%)	4333 (28.3%)	−157.5	2099.2	0.93 (0.89 − 0.97)
4) HLT: Arterial therapeutic procedures (excl aortic)	1479 (9.6%)	1752 (11.4%)	−149.5	807.6	0.83 (0.78 − 0.89)
HIGHER WITH MORE STATIN					
5) HLGT: Enzyme investigations nec	286 (1.9%)	196 (1.3%)	44.8	120.5	1.45 (1.21 − 1.73)
6) HLT: Skeletal and cardiac muscle analyses	245 (1.6%)	169 (1.1%)	37.7	103.5	1.44 (1.19 − 1.75)
7) HLGT: Hepatobiliary investigations	502 (3.3%)	253 (1.6%)	126.2	188.7	1.95 (1.69 − 2.25)
8) HLT: Liver function analyses	493 (3.2%)	243 (1.6%)	126.7	183.9	1.99 (1.72 − 2.30)
9) PT: Liver function test abnormal	139 (0.9%)	73 (0.5%)	33.0	53.0	1.87 (1.42 − 2.44)
10) PT: Transaminases increased	48 (0.3%)	12 (<0.1%)	18.1	15.0	3.34 (2.01 − 5.54)
11) PT: Hepatic enzyme increased	129 (0.8%)	58 (0.4%)	35.6	46.7	2.14 (1.61 − 2.85)
12) PT: Alanine aminotransferase increased	147 (1.0%)	75 (0.5%)	35.7	55.5	1.90 (1.46 − 2.47)
13) PT: Aspartate aminotransferase increased	119 (0.8%)	61 (0.4%)	28.6	45.0	1.89 (1.41 − 2.53)
14) PT: Laboratory test abnormal	184 (1.2%)	76 (0.5%)	54.6	65.0	2.32 (1.82 − 2.95)



FDR= False Discovery Rate (corrected using Mehrotra and Adewale method).

Figure 6.7: UNBLINDED analyses of all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (4 trials, 30,724 randomised patients)

For each of the 27 System Organ Classes shown, the plotted points show the nominal p−values for tests of the event rates among patients allocated more statin versus event rates among patients allocated less statin (with values above the horizontal axis indicating that the event was more common among patients allocated more statin and values below the axis indicating that the event was less common among patients allocated more statin). 645 such p−values are shown, corresponding to, 25 tests at the SOC level, 77 tests at the HLGT level, 132 tests at the HLT level and 411 tests at the PT level. Of these tests 14 (2.2%) are significant after correction of the FDR at 5% level. These are indicated with closed circles and are enumerated (see Figure 6.8 for the identification and results of each of these 14 event categories). SOC= System Organ Class. HLGT= Higher Level Group Term. HLT= Higher Level Term. PT=Preferred Term. FDR= False Discovery Rate (corrected using the Mehrotra and Adewale method). MedDRA V20.0 was used to report the adverse events, organised by SOC. MedDRA terms with fewer than 6 events or with identical counts at the above levels were not being tested. *PTs with 50 or more events and HLTs or HLGTs which report 50% or more extra events to level(s) below.

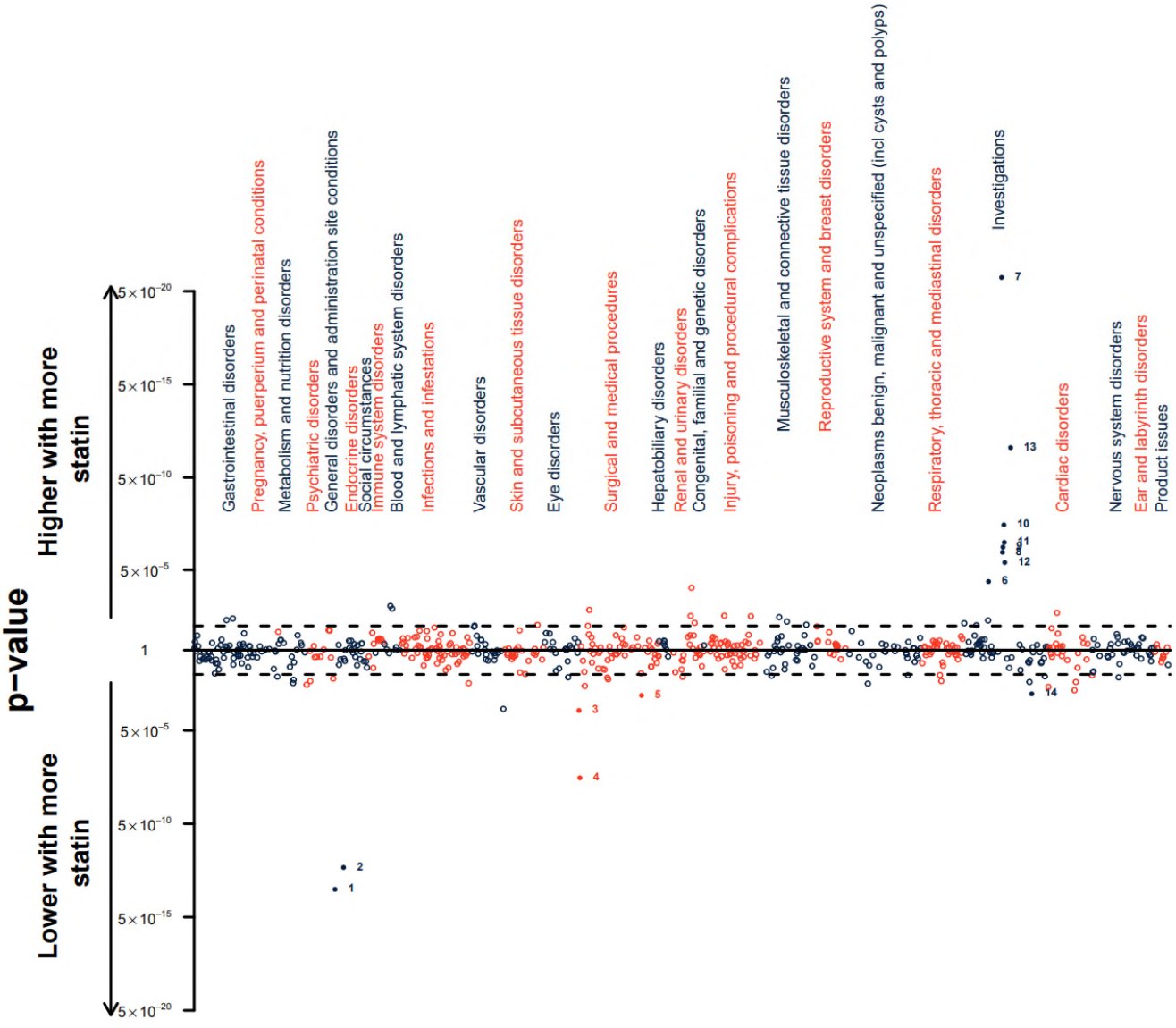
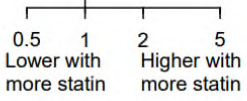


Figure 6.8: FDR significant effects of more statin therapy on all adverse events EXCEPT those related to muscle, diabetes and the known cardiovascular benefits of statins, and AFTER applying a testing strategy which excludes very rare outcomes and avoids irrelevant overlap in tests* (4 trials, 30,724 randomised patients)

MedDRA Term	Events (%) More Statin (n=15390)	Events (%) Less Statin (n=15334)	Logrank O−E	Logrank V	RR (95% CI)
LOWER WITH MORE STATIN					
1) SOC: General disorders and administration site conditions	5897 (38.3%)	6516 (42.5%)	−410.9	3096.5	0.88 (0.85 − 0.91)
2) HLGT: General system disorders nec	5098 (33.1%)	5682 (37.1%)	−363.9	2690.3	0.87 (0.84 − 0.91)
3) SOC: Surgical and medical procedures	4067 (26.4%)	4333 (28.3%)	−157.5	2099.2	0.93 (0.89 − 0.97)
4) HLT: Arterial therapeutic procedures (excl aortic)	1479 (9.6%)	1752 (11.4%)	−149.5	807.6	0.83 (0.78 − 0.89)
5) HLT: Renal therapeutic procedures	30 (0.2%)	57 (0.4%)	−13.5	21.7	0.54 (0.35 − 0.82)
14) PT: Blood urea increased	15 (<0.1%)	35 (0.2%)	−10.0	12.5	0.45 (0.26 − 0.78)
HIGHER WITH MORE STATIN					
6) HLT: Skeletal and cardiac muscle analyses	245 (1.6%)	169 (1.1%)	37.7	103.5	1.44 (1.19 − 1.75)
7) HLT: Liver function analyses	493 (3.2%)	243 (1.6%)	126.7	183.9	1.99 (1.72 − 2.30)
8) PT: Liver function test abnormal	139 (0.9%)	73 (0.5%)	33.0	53.0	1.87 (1.42 − 2.44)
9) PT: Transaminases increased	48 (0.3%)	12 (<0.1%)	18.1	15.0	3.34 (2.01 − 5.54)
10) PT: Hepatic enzyme increased	129 (0.8%)	58 (0.4%)	35.6	46.7	2.14 (1.61 − 2.85)
11) PT: Alanine aminotransferase increased	147 (1.0%)	75 (0.5%)	35.7	55.5	1.90 (1.46 − 2.47)
12) PT: Aspartate aminotransferase increased	119 (0.8%)	61 (0.4%)	28.6	45.0	1.89 (1.41 − 2.53)
13) PT: Laboratory test abnormal	184 (1.2%)	76 (0.5%)	54.6	65.0	2.32 (1.82 − 2.95)

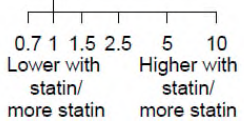


FDR= False Discovery Rate (corrected using Mehrotra and Adewale method).
*PTs with 50 or more events and HLTs or HLGTs which report 50% or more extra events to level(s) below.

Figure 6.9: Meta-analyses of the adverse events identified to be increased with statin or more statin therapy at the 5% FDR level

MedDRA Term	Events (%) Statin/ More statin	Events (%) Placebo/ Less statin	Logrank O−E	Logrank V	RR (95% CI)
HLT: Myopathies					
Statin vs placebo	46 (<0.1%)	30 (<0.1%)	7.9	19.0	1.52 (0.97 − 2.38)
More vs less statin*	63 (0.4%)	13 (<0.1%)	25.0	19.0	3.74 (2.38 − 5.86)
Overall	109 (0.1%)	43 (<0.1%)	33.0	38.0	**2.38 (1.73 − 3.27)**
PT: Myopathy					p<0.001
Statin vs placebo	29 (<0.1%)	11 (<0.1%)	9.0	10.0	2.46 (1.32 − 4.57)
More vs less statin*	50 (0.3%)	9 (<0.1%)	20.5	14.7	4.03 (2.42 − 6.71)
Overall	79 (0.1%)	20 (<0.1%)	29.5	24.7	**3.30 (2.22 − 4.89)**
HLGT: Enzyme investigations nec					p<0.001
Statin vs placebo	873 (1.4%)	798 (1.3%)	35.7	417.6	1.09 (0.99 − 1.20)
More vs less statin*	286 (1.9%)	196 (1.3%)	44.8	120.5	1.45 (1.21 − 1.73)
Overall	1159 (1.5%)	994 (1.3%)	80.6	538.1	**1.16 (1.07 − 1.26)**
HLT: Skeletal and cardiac muscle analyses					p<0.001
Statin vs placebo	767 (1.2%)	710 (1.1%)	26.8	369.1	1.08 (0.97 − 1.19)
More vs less statin*	245 (1.6%)	169 (1.1%)	37.7	103.5	1.44 (1.19 − 1.75)
Overall	1012 (1.3%)	879 (1.1%)	64.5	472.6	**1.15 (1.05 − 1.25)**
PT: Blood creatine phosphokinase increased					p=0.003
Statin vs placebo	704 (1.1%)	651 (1.1%)	24.8	338.6	1.08 (0.97 − 1.20)
More vs less statin*	223 (1.4%)	146 (1.0%)	38.3	92.2	1.51 (1.24 − 1.86)
Overall	927 (1.2%)	797 (1.0%)	63.1	430.9	**1.16 (1.05 − 1.27)**
HLGT: Hepatobiliary investigations					p=0.002
Statin vs placebo*	1349 (2.2%)	1025 (1.7%)	164.2	593.3	1.32 (1.22 − 1.43)
More vs less statin*	502 (3.3%)	253 (1.6%)	126.2	188.7	1.95 (1.69 − 2.25)
Overall	1851 (2.4%)	1278 (1.7%)	290.5	782.0	**1.45 (1.35 − 1.56)**
HLT: Liver function analyses					p<0.001
Statin vs placebo*	1277 (2.1%)	937 (1.5%)	172.7	553.3	1.37 (1.26 − 1.49)
More vs less statin*	493 (3.2%)	243 (1.6%)	126.7	183.9	1.99 (1.72 − 2.30)
Overall	1770 (2.3%)	1180 (1.5%)	299.5	737.2	**1.50 (1.40 − 1.61)**
PT: Liver function test abnormal					p<0.001
Statin vs placebo	114 (0.2%)	107 (0.2%)	3.3	55.2	1.06 (0.81 − 1.38)
More vs less statin*	139 (0.9%)	73 (0.5%)	33.0	53.0	1.87 (1.42 − 2.44)
Overall	253 (0.3%)	180 (0.2%)	36.3	108.2	**1.40 (1.16 − 1.69)**
PT: Transaminases increased					p<0.001
Statin vs placebo	50 (<0.1%)	34 (<0.1%)	8.0	21.0	1.46 (0.95 − 2.24)
More vs less statin*	48 (0.3%)	12 (<0.1%)	18.1	15.0	3.34 (2.01 − 5.54)
Overall	98 (0.1%)	46 (<0.1%)	26.1	36.0	**2.06 (1.49 − 2.86)**
PT: Hepatic enzyme increased					p<0.001
Statin vs placebo*	203 (0.3%)	109 (0.2%)	47.3	78.0	1.83 (1.47 − 2.29)
More vs less statin*	129 (0.8%)	58 (0.4%)	35.6	46.7	2.14 (1.61 − 2.85)
Overall	332 (0.4%)	167 (0.2%)	82.9	124.7	**1.94 (1.63 − 2.32)**
PT: Alanine aminotransferase increased					p<0.001
Statin vs placebo*	684 (1.1%)	474 (0.8%)	105.5	289.4	1.44 (1.28 − 1.62)
More vs less statin*	147 (1.0%)	75 (0.5%)	35.7	55.5	1.90 (1.46 − 2.47)
Overall	831 (1.1%)	549 (0.7%)	141.1	344.9	**1.51 (1.35 − 1.67)**
PT: Aspartate aminotransferase increased					p<0.001
Statin vs placebo*	278 (0.4%)	199 (0.3%)	39.4	119.2	1.39 (1.16 − 1.67)
More vs less statin*	119 (0.8%)	61 (0.4%)	28.6	45.0	1.89 (1.41 − 2.53)
Overall	397 (0.5%)	260 (0.3%)	68.0	164.2	**1.51 (1.30 − 1.76)**
PT: Laboratory test abnormal					p<0.001
Statin vs placebo	337 (0.5%)	297 (0.5%)	19.1	158.5	1.13 (0.97 − 1.32)
More vs less statin*	184 (1.2%)	76 (0.5%)	54.6	65.0	2.32 (1.82 − 2.95)
Overall	521 (0.7%)	373 (0.5%)	73.7	223.5	**1.39 (1.22 − 1.59)**
PT: Parosmia					p<0.001
Statin vs placebo*	18 (<0.1%)	2 (<0.1%)	8.0	5.0	4.93 (2.05 − 11.86)
More vs less statin	0 (0.0%)	0 (0.0%)	0.0	0.0	
Overall	18 (<0.1%)	2 (<0.1%)	8.0	5.0	**4.93 (2.05 − 11.86)**
SOC: Musculoskeletal and connective tissue disorders					p<0.001
Statin vs placebo*	22012 (35.5%)	21417 (34.6%)	442.6	10825.0	1.04 (1.02 − 1.06)
More vs less statin	6618 (43.0%)	6464 (42.2%)	103.0	3265.2	1.03 (1.00 − 1.07)
Overall	28630 (37.0%)	27881 (36.1%)	545.6	14090.2	**1.04 (1.02 − 1.06)**
					p<0.001

*FDR significant result



7. Discussion

7.1. Summary

The aim of this chapter is to summarise the work done for this thesis, report the novel aspects, discuss its strengths and limitations and finally to outline plans for future work. This final chapter starts with a summary of the thesis including its aims, the methods undertaken, the results generated and their interpretation. Then, it reports its various advantages compared to previous studies addressing safety concerns related to statin therapy, as well as the limitations of this work. It continues with proposing various research hypotheses generated to guide future research and, finally, it highlights the project's novel findings and innovative methodologies.

7.2. Overview of findings

Statins reduce the production of LDL cholesterol in the liver leading to a reduction in risk of cardiovascular disease. In particular, previous reports from the CTT Collaboration have shown that standard dose statin therapy produces a 21% reduction in major vascular events per 1.0 mmol/L reduction in LDL cholesterol, while high intensity statin therapy (eg, 40-80 mg atorvastatin) produces a further 0.5 mmol/L reduction in LDL cholesterol corresponding to a further 15% proportional reduction in major vascular events (18, 19).

Some studies, mainly non-randomised observational studies, have linked statins with an increased risk of various adverse events, with such claims leading many people at high risk of vascular disease to stop or not start statin therapy when they could benefit from it. To address this uncertainty, the CTT Collaboration initiated a new program of work to collect and analyse patient level data on all AEs from 23 large-scale randomised double blind trials. This approach was adopted because randomisation is the best way to minimize systematic errors and (double-) blinding is the best method to minimize the risk of reporting bias. This DPhil project

focused on the analyses of AEs with no prior knowledge of an association with statin therapy with the aim of developing appropriate statistical methods to address the problems associated with multiple testing and identifying any real, previously unknown, effects of statin therapy.

Initially, a data harmonization methodology was developed to transform the heterogeneous patient level data received from the trialists (ie, >840 datasets, approximately 30,000 variables and >38 million records) into a single common-format dataset. A 'meta' data-dictionary based on CDISC SDTM was used to convert the data into a single analysable dataset, with 13 of the 46 available SDTM domains (equating to a total of 121 variables) considered sufficient for the aims of this project. In addition, a methodology was established to standardize the approximately 1.2 million AEs (of which about 45,000 were unique AE types) received from the multiple trials, into a common medical terminology based on the internationally recognized MedDRA dictionary. A two-step process was developed based on MedDRA version 20.0, auto-coding the AE terms by starting from the LLT level and progressively moving up to the HLGT level. Approximately 780,000 AEs were auto-coded using this process, whilst approximately 401,000 AEs (of which approximately 32,000 were unique terms) were manually coded by the CTT Secretariat in Oxford.

The next task was to conduct a systematic review of the literature on MHT methods and identify those methods potentially relevant for use in the CTT meta-analysis of AEs with a previously unknown association with statin therapy. A total of 6,569 articles was identified, of which 337 were included for full text review. Of the MHT methods identified, only frequentist approaches designed to allow for two-tailed hypotheses with certain types of correlation between hypotheses, and which made no additional assumptions about the total number of true null hypotheses, were considered suitable for use in this meta-analysis. Nine such MHT methods were identified but only five of them were selected as the most appropriate based on their statistical control of type I and II error rates, and hence taken forward for further exploration through simulation analyses. Of those five methods, three were designed to control

the FWER (ie, Holm, Hochberg and Hommel) and two the FDR (ie, Benjamini & Hochberg and Mehrotra & Adewale).

Following this, simulation analyses were undertaken with a goal of identifying the best of the five suitable MHT methods to use in the CTT AE meta-analyses based on their statistical properties (ie, expected number of FPs and statistical power), as well as the development of an integrated MHT framework for the selection of appropriate AEs (ie, based on the total number of events and/or the correlation structure of hypotheses) for meta-analyses. Five scenarios were implemented with the effects of statin therapy ranging from 0% to 10% of the total hypotheses (ie, AEs) examined. The null hypothesis of no effect (of statin therapy) was implemented by forcing the true RR to be equal to 1.00, while for the alternative hypotheses of a real effect of statin therapy, a range of RRs were considered starting from 1.03 up to 10, with events in the statin arm simulated from a binomial distribution with probability equal to RR/(1+RR). For each MedDRA AE term reported by at least one participant in a trial, the observed minus expected (o-e) test statistic and its associated variance (v) was calculated. In a meta-analysis of trials, the log of the log of the risk ratio was calculated as S/V with variance 1/V where S is the sum over all trials of (o–e) and V is the sum over all trials of v calculated, with the nominal p-value calculated comparing the test statistic (ie, S/V) against a chi-square distribution with 1 degree of freedom. For each MHT method, the number of FPs and true positives were calculated across the various simulated scenarios. The simulation analyses were repeated 10,000 times to produce robust and accurate statistical estimates.

Results from such simulation analyses showed that not correcting for MHT (ie, 5% type I error) would lead to a large number of FPs (with the expected proportion of hypotheses flagged as FPs ranging from 5% to 6%). The three MHT methods controlling the FWER (ie, Holm, Hochberg and Hommel) were shown to produce almost no FPs but maintained very low statistical power to detect any real effects of statin therapy. The methods controlling the FDR (ie, Benjamini & Hochberg and Mehrotra & Adewale) were shown to be the most appropriate to use in these meta-analyses, since these methods resulted in a low number of FPs and

maintained good statistical power to detect any real effects of statin therapy. The Mehrotra & Adewale method was determined to be the most suitable for use in the meta-analyses of the CTT AE database because on various occasions it was marginally more powerful than the Benjamini & Hochberg method. In addition, simulations showed that, for AEs with a very low number of events (ie, <50), the expected power was very low to detect any real effects of statin therapy. In addition, the decision not to test such rare AEs was shown to increase the statistical power of detecting any real effects of statin therapy on all other AEs considered for testing by approximately 5%. Tests at the higher MedDRA hierarchical levels (ie, HLT and HLGT) should be performed only if such hypotheses contributed substantial extra information, for example, at least 50% statistical information.

The next two chapters presented the results of the meta-analyses of all AEs in the double blind trials with the primary population being statin therapy vs placebo, and the secondary analysis population being more intensive vs less intensive statin therapy in which any apparent effects were examined to be dose-related.

A time-to-event analysis using the log-rank statistical method was used to calculate the observed minus expected (o–e) statistic and its variance (v) for the first occurrence of each AE among participants randomised into each trial. Then, in a meta-analysis of such trials, the fixed effect Peto one-step method was used to calculate the RR and 95% confidence interval. The nominal p-value was calculated from the comparison of the test statistic (of the meta-analyses) against a chi-square with 1 degree of freedom. The Mehrotra & Adewale MHT method controlling for the FDR at 5% (ie, 2p=0.05) level of significance was used to test the multiple null hypotheses (ie, no effect of either statin or of more intensive statin therapy) to identify the true effects of statin therapy.

The meta-analyses of the 19 statin vs placebo trials involved 123,940 participants followed up for an average median of 4.3 years, with these participants having a mean age of 63 years with 28% being women and 48% having a history of vascular disease. The blinded (to

treatment allocation) analyses highlighted the importance of using statistical techniques which account for the problem of MHT, (since no correction would lead to about 180 FPs out of a total of 4,094 AEs examined), with the Mehrotra & Adewale MHT method reducing the number of FP to zero indicating that any effects identified by such method are likely to be real. Then, the initial set of unblinded analyses (ie, reporting true effects of statin therapy), which included all non-lipid AEs, showed that statins significantly reduce the risk of various types of cardiovascular events (eg, RR 0.70 for myocardial infarction and 0.76 for ischaemic stroke) and death by (ie, RR 0.91), with such treatment effects being similar to those reported in previous CTT papers (18, 19, 22, 24, 25). Allocation to statin therapy was found to reduce the rates of two types of coronary procedures, arteriogram (ie, RR 0.82) and angiogram (ie, RR 0.67). On the contrary, statin therapy was shown to increase the risk of some adverse events, including musculoskeletal and connective tissue disorders (ie, increased risk of 4%), hepatobiliary investigations such as liver function analyses (ie, increased risk of 32%) and the very rare outcome of parosmia (ie, relative risk of about 9, but reported by less than 2 in 10,000 patients treated for 5 years). The two further unblinded analyses performed after removing already known effects of statin therapy and very rare outcomes to avoid irrelevant overlap in tests resulted in the identification of two new benefits of statin therapy (ie, reductions in peripheral embolism and thrombosis and carotid endarterectomy) but no additional harmful effects.

The meta-analyses of the 4 more vs less statin trials involved 30,724 participants followed up for an average median of approximately 5 years, with mean age of 62 years, and with 19% being women and 15% having a history of diabetes. The blinded analyses of all AEs resulted in a large number of FPs (ie, 115 FPs out of a total of 2,169 AEs) when not correcting for the problem of MHT, with this number being zero when applying the Mehrotra & Adewale method at the 5% FDR level. The first set of the unblinded analyses on all non-lipid AEs showed that more intensive (compared to less intensive) statin therapy further reduced the risk of cardiovascular outcomes including the known effects of coronary arterial therapeutic

procedures, ischaemic coronary artery disorders and myocardial infarction. In addition, more intensive statin therapy further reduced the risk (compared to less intensive statin therapy) of coronary artery disorders (ie, angina pectoris and unstable angina) by approximately 8%, general system disorders NEC (ie, oedema, general signs and symptoms, pain and discomfort) by 13%, and any surgical and medical procedures by 7%. Further analyses which removed already known effects of statin therapy or very rare outcomes plus any irrelevant overlap in tests showed that more intensive statin therapy reduced the risk of the rare AE of renal therapeutic procedures by about half, and the very low occurence of a blood urea increase by more than 50%. Compared to less intensive statin regimens, more intensive regimens were shown to cause very few additional adverse effects, but did further increase the risk of hepatobiliary investigations by approximately 95%, the very rare adverse event of myopathy by about 4 times, and blood creatine phosphokinase by approximately 50% (with those last two outcomes being clinically correlated).

7.3. Strengths

7.3.1. Large-scale, double-blinded randomised trials

Previous evidence estimating the effects of statin therapy on particular types of AEs has mainly been reported from non-randomised observational studies which may be subject to moderate biases and hence can produce misleading results. A key strength of this study is that it uses data reported in large-scale, double-blind randomised trials. The method of randomisation is the best approach to minimize any systematic errors (ie, biases) and ensure that the observed differences in rates among health outcomes can generally be attributed causally to differences in study treatment, while the use of double-blinded trials (ie, masking assignment for both patient and doctor of study treatment with a matching dummy treatment) minimizes the risk of reporting bias, also known as the nocebo effect. In addition, the use

large-scale randomised trials has the advantage of allowing for the study of a large number of events which helps to avoid moderate random errors.

7.3.2. IPD meta-analyses of all AEs reported in large-scale trials

Another strength of this study is the use of IPD meta-analyses of all AEs reported across multiple large-scale trials of statin therapy. The use of a meta-analytic methodological approach which combines data from multiple trials (preserving the randomisation principle) allows for reliable assessment of the effect of moderate treatment effects on an outcome or the examination of very rare outcomes. Also, the availability of IPD (as opposed to tabular data) allowed for harmonization of the heterogeneous IPD received from the individual trialists into a single common-format dataset based on SDTM standards, which enabled exploration of whether the treatment effects differed over time or across various types of patient. In addition, the standardization of AEs into a single medical dictionary based on MedDRA (ie, a well-established medical dictionary widely used for regulatory communication and evaluation of AE data) is another key strength of this study which could only be achieved by using patient-level data. Furthermore, the collection of all the adverse events received from the multiple trials allowed us to examine the effects of statins on *all* types of adverse events reported in trials of statin therapy.

7.3.3. Validation of methodologies applied

Another asset of this thesis is the fact that various validation approaches were used to confirm that the new statistical and data handling methodologies applied were appropriate and performing reliably.

In particular, a systematic search of the literature was conducted to identify all the MHT methods suitable for use in this project. Then, simulation analyses were implemented to monitor the performance of those MHT methods suitable for use with the CTT AE data (and various scenarios for the treatment effects) in order to identify the best method to use in the actual meta-analysis. In addition, the simulation analyses were used to estimate the statistical

power of the various types of AEs and develop an appropriate strategy for the selection of hypotheses across the MedDRA tree included for analysis.

Although the data handling methodologies (ie, harmonization of IPD into a single dataset with a common format and standardization of AEs into a single medical dictionary) were based on global standards (ie, CDISC SDTM and MedDRA), this project had the advantage of being able to validate the results generated from these new data by comparing them against already known effects of statin therapy from previous CTT papers. The first set of unblinded analyses included all non-lipid AEs, with results showing that statins reduce the risk of various types of cardiovascular disease (ie, myocardial infarction, angina pectoris and ischaemic stroke) and death, which are known statin benefits from previous CTT papers (18, 19). On the contrary, statins were shown to increase the risk of musculoskeletal and connective tissue by 4%, while high intensity statins caused an excess in the risk of myopathy, with such treatment effects being similar to those recently reported in the CTT analyses of the effects of statin therapy on all muscle symptoms (ie, 3% increase in risk of any muscle pain or weakness when taking statin therapy ; 5 times more likely to experience myopathy with high intensity statin therapy) (122). The above findings confirm that the data handling methods are reliable and can be used safely to detect any new real effects of statin therapy.

7.3.4. Rapid assessment for hypotheses related to statin therapy

The updated CTT database created for the new program of work (35) contains IPD from 28 trials (of which 23 are double-blind trials) requesting all individual-level datasets reporting information on AEs, reasons for stopping study treatment, co-medications, laboratory or physical measurements, lifestyle characteristics and demographics. Such a rich database can be used in the future to explore a range of additional research questions related to statin therapy based on previous or upcoming studies including (but not limited to) estimating the effect of statins (ie, including subdivisions by dose) on specific outcomes or subpopulations of patients. In addition, the analyses produced for this thesis allows for a rapid assessment of

the overall treatment effects for particular types of AEs (reported across the whole MedDRA tree) which may be further examined using the CTT AE database in the future.

7.4. Limitations

This study has some limitations. Firstly, the analyses performed included AEs reported only through the 'standard' AE datasets (including AEs derived from reasons for stopping study treatment). Although these datasets might include all events related to the vast majority of AE endpoints (ie, >99%), there are particular types of AEs which may be informed further by outcomes reported in other datasets, such as the co-medication and laboratory datasets. A specific example is the outcome of incidence of diabetes which was not identified as FDR significant in both types of trial meta-analysis (ie, statin vs placebo and more intensive vs less intensive statin therapy), even when performing a sensitivity analyses removing participants with diabetes at baseline (ie, the RR for the PT of 'Diabetes mellitus' in trials of statin vs placebo was 1.10 [95% CI 1.02 - 1.18, p=0.016] and in trials of more vs less statin therapy 1.11 [95% CI 1.01– 1.23, p= 0.030]) (unpublished data). This is because most of the events for incident diabetes were not reported as 'standard' AEs (eg, only 2,797 out of 6,760 events being reported in the AE datasets of trials of statin vs placebo), but were instead identified using laboratory measurements (ie, abnormal values for glucose and HbA1C) and diabetes-related co-medications (ie, metformin and insulin). In addition, the excess in risk when taking statin therapy was higher in diabetes determined biochemically or from co-medication (eg, 20% and 14% respectively in trials of statin vs placebo) than when determined by the actual AE of diabetes (ie, only 10% increase in risk with statin therapy) (unpublished data).

As shown above (chapters 5 and 6), the analyses of AEs were performed based on the hierarchical structure of the MedDRA dictionary with analyses starting from the most granular PT terms and moving up to more generic HLT and HLGT terms and finally the very broad SOC terms. Although this approach allowed for examination of a very large number of correlated

hypotheses based on a well-established tool for classifying AE endpoints, it has a few limitations. The MedDRA tree sometimes reports hypotheses which are closely related with each other into two different SOCs. For example, the PT 'Myocardial infarction' reported under the SOC 'Cardiac disorders' might be related to the PT 'Infarction' reported under the SOC 'Vascular disorders'. In addition, the approach of using Standardised MedDRA Queries (SMQs: a tool developed by experts in various disease areas and those working in pharmacovigilance and clinical development to group sets of related MedDRA terms together) (123) was not considered for these analyses. However, the CTT database contains a very rich AE dataset which allowed for more detailed analyses than the SMQs might for the effects of statins on various types of AEs. Also, further testing of the approximately 100 different SMQs would endanger breaking the correlation structure incorporated within the MedDRA dictionary.

The analyses produced for this thesis did not explore the heterogeneity of the effects of statin on *all* AEs across different subpopulations (ie, different age groups, or by gender and medical history) or over time. However, performing such detailed analyses would increase the number of hypotheses to be tested substantially (ie, tens of thousands of extra hypotheses would be required) endangering their ability to detect any true effects of statin therapy across the whole population examined.

Finally, on-treatment analyses for particular rare types of AEs (identified as FDR significant) were planned but not feasible because data were not consistently available for discontinuation of treatment.

7.5. Future work

The new hypotheses (ie, new effects of statin therapy) generated from this thesis can become the basis for future exploratory work. Specifically, more detailed analyses may be needed to estimate the treatment effects of standard and high intensity statins on a set clinically important endpoints related to hepatobiliary investigations by experts in the field, including the impact (if

any) from the extra data reported in laboratory or other datasets. Also, similar analyses may be needed for various other AEs detected as FDR significant, such as the apparently beneficial effects of statin on surgical or medical procedures or more intensive statin therapy on general system disorders. In addition, further analyses are needed to estimate adverse effects of statin therapy on hepatobiliary disease across various subgroups of patients or to assess if such effects differ over time. Moreover, more detailed analyses may be required to explore the effect of statin therapy on clinical endpoints closely related to parosmia, as well as, if such adverse effect appear only to specific types of patient. Also, as shown in section 7.3.4, the current analyses can be used to assess fairly rapidly treatment effects for various AEs reported in the CTT AE database which may be of interest in future, for example, testing hypotheses related to effects of statin therapy (or more intensive statin therapy) on particular endpoints generated from future studies.

7.6. Novelty

7.6.1. Ability to address public health concerns

Despite the fact that statins reduce the risk of major vascular events by about one-quarter for each mmol/L reduction in LDL cholesterol during each year (after the first) that it continues to be taken, hence preventing life-threatening cardiovascular morbidity and mortality (2), some people stop or never start taking them due to concerns regarding their adverse effects. This is a significant public health problem associated with an increased risk of cardiovascular events. However, these concerns have typically been based on non-randomised observational evidence which may be subject to bias. To the best of my knowledge, this is the first IPD meta-analyses of AEs conducted in large-scale randomised trials of a therapy to allow for examination of the causal effects of such a therapy on all AEs (in this case statin therapy). This study combined the AE data from multiple trials to maximize the likelihood of detecting

any real unknown effects (beneficial or harmful) of statin therapy which were not able to be detected in a single trial.

Results confirmed the already known cardiovascular benefits of statins and also showed that statins reduced the risk of undergoing a coronary arteriogram or angiogram, as well as peripheral embolism and thrombosis, and carotid endarterectomy. On the contrary, statins increased the risk of musculoskeletal and connective tissue disorders (with detailed analyses presented in the recently published CTT muscle outcomes paper), hepatobiliary investigations including liver function analyses (including the AEs of hepatic enzymes, alanine and aspartate aminotransferase increased); and the very rare occurrence of parosmia. The meta-analyses of more intensive vs less intensive statin therapy trials, apart from confirming the known benefits on major vascular events, detected new benefits of more intensive statin therapy including reducing the risk of coronary artery disorders, general system disorders, any surgical or medical procedures and renal therapeutic procedures. However, more intensive statin regimens increased the risk hepatobiliary investigations, myopathy) and blood creatine phosphokinase.

The above findings highlight the important benefits of statin therapy in preventing cardiovascular disease and also identify some new important beneficial effects including (but not limited to) the reduction of the number of some medical or surgical procedures and the outcomes of general system and coronary artery disorders, at the cost of a minimal increase in risk of a few AEs including hepatobiliary investigations, musculoskeletal and connective tissue disorders and parosmia. Consequently, the nature and very low number of adverse effects caused by statin therapy are greatly outweighed by the numerous important, often life-saving, beneficial effects of such treatment. These findings suggest that there is a need to review the information in the drug label for statins and to further inform doctors and patients regarding the safety of this drug class, and the importance of starting or continuing to take them if indicated so as to improve the health of people at risk of cardiovascular disease.

7.6.2. Methodological novelties

To the best of my knowledge, this is the first time that bespoke systems to convert large amount of patient-level data from numerous RCTs into a single analysable dataset based on global standards (ie, CDISC SDTM and MedDRA) have been created for IPD meta-analyses (107). Also, this study initiated some innovative simulation studies which accounted for the complex dependency structures (ie, the correlation between hypotheses and the meta-analytical effect). Those simulations were conducted based on the *exact* structure of the CTT AE data (ie, using a large subset of the data) and avoiding having to make any assumptions (ie, about how the data might look). This allowed for evaluation of the performance of the various MHT approaches and estimation of the statistical power of the various types of AEs examined based on different scenarios for the treatment effects. In addition, this was the first study to examine the effects of a treatment on such a large safety dataset involving thousands of hypotheses related to AEs. Furthermore, some creative statistical programming had to be developed to extract and organize the summary estimates (for all AEs) so as to present the results in a concise and easily understood way.

7.7. Conclusion

This DPhil investigated the effects of statin therapy on all AEs using IPD from large-scale, randomised, double-blind trials, taking into consideration the substantial challenges related to multiple testing. The findings highlighted the importance of using appropriate statistical methods to control for MHT. The results confirmed the already known benefits of statin therapy while identifying some new potential benefits and harms. However, the nature and frequency of the newly detected harms are greatly outweighed by the important, often life-saving, benefits of statin therapy.

References

1. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. Lancet (London, England). 2001;357(9253):373-80.
2. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. The Lancet. 2016;388(10059):2532-61.
3. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol. 2020;76(25):2982-3021.
4. Mensah GA, Roth GA, Fuster V. The Global Burden of Cardiovascular Diseases and Risk Factors. Journal of the American College of Cardiology. 2019;74(20):2529-32.
5. Prospective Studies Collaboration. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. Lancet (London, England). 2007;370(9602):1829-39.
6. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. The Lancet. 2010;375(9733):2215-22.
7. Petersen LK, Christensen K, Kragstrup J. Lipid-lowering treatment to the end? A review of observational studies and RCTs on cholesterol and mortality in 80+-year olds. Age and Ageing. 2010;39(6):674-80.
8. Ravnskov U, Diamond DM, Hama R, Hamazaki T, Hammarskjöld B, Hynes N, et al. Lack of an association or an inverse association between low-density-lipoprotein cholesterol and mortality in the elderly: a systematic review. BMJ Open. 2016;6(6):e010401.
9. Law MR, Thompson SG, Wald NJ. Assessing possible hazards of reducing serum cholesterol. BMJ. 1994;308(6925):373-9.
10. Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? BMJ. 1994;308(6925):367-72.
11. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. British Medical Journal. 1990;301(6747):309-14.
12. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
13. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of Long-Term Exposure to Lower Low-Density Lipoprotein Cholesterol Beginning Early in Life on the Risk of Coronary Heart Disease: A Mendelian Randomization Analysis. Journal of the American College of Cardiology. 2012;60(25):2631-9.
14. Postmus I, Deelen J, Sedaghat S, Trompet S, de Craen AJ, Heijmans BT, et al. LDL cholesterol still a problem in old age? A Mendelian randomization study. International Journal of Epidemiology. 2015;44(2):604-12.
15. Ference BA, Bhatt DL, Catapano AL, Packard CJ, Graham I, Kaptoge S, et al. Association of Genetic Variants Related to Combined Exposure to Lower Low-Density Lipoproteins and Lower Systolic Blood Pressure With Lifetime Risk of Cardiovascular Disease. JAMA. 2019;322(14):1381-91.
16. Tobert JA. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. Nature Reviews Drug Discovery. 2003;2(7):517-26.
17. Schachter M. Chemical, pharmacokinetic and pharmacodynamic properties of statins: an update. Fundamental & clinical pharmacology. 2005;19(1):117-25.
18. Cholesterol Treatment Trialists Collaboration. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials. The Lancet. 2010;376(9753):1670-81.

19. Cholesterol Treatment Trialists Collaboration. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. The Lancet. 2005;366(9493):1267-78.
20. Cholesterol Treatment Trialists Collaboration (https://www.cttcollaboration.org, last visited on 31 Oct 2019).
21. Cholesterol Treatment Trialists Collaboration. Efficacy of cholesterol-lowering therapy in 18 686 people with diabetes in 14 randomised trials of statins: a meta-analysis. The Lancet. 2008;371(9607):117-25.
22. Cholesterol Treatment Trialists Collaboration. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174,000 participants in 27 randomised trials. The Lancet. 2015; 10.1016/S0140-6736(14)61368-4.
23. Cholesterol Treatment Trialists C. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. The Lancet. 2012;380(9841):581-90.
24. Cholesterol Treatment Trialists Collaboration. Efficacy and safety of statin therapy in older people: a meta-analysis of individual participant data from 28 randomised controlled trials. The Lancet. 2019;393(10170):407-15.
25. Cholesterol Treatment Trialists Collaboration. Lack of Effect of Lowering LDL Cholesterol on Cancer: Meta-Analysis of Individual Data from 175,000 People in 27 Randomised Trials of Statin Therapy. PLOS ONE. 2012;7(1):e29849.
26. Bruckert E, Hayem G, Dejager S, Yau C, Bégaud B. Mild to Moderate Muscular Symptoms with High-Dosage Statin Therapy in Hyperlipidemic Patients —The PRIMO Study. Cardiovascular Drugs and Therapy. 2005;19(6):403-14.
27. Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database. BMJ. 2010;340:c2197.
28. Kang J-H, Kao L-T, Lin H-C, Tsai M-C, Chung S-D. Statin use increases the risk of depressive disorder in stroke patients: A population-based study. Journal of the Neurological Sciences. 2015;348(1):89-93.
29. Evans MA, Golomb BA. Statin-Associated Adverse Cognitive Effects: Survey Results from 171 Patients. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 2009;29(7):800-11.
30. Dormuth CR, Hemmelgarn BR, Paterson JM, James MT, Teare GF, Raymond CB, et al. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases. BMJ: British Medical Journal. 2013;346(7900):14-.
31. Nielsen SF, Nordestgaard BG, Bojesen SE. Statin Use and Reduced Cancer-Related Mortality. New England Journal of Medicine. 2012;367(19):1792-802.
32. van de Garde EMW, Hak E, Souverein PC, Hoes AW, van den Bosch JMM, Leufkens HGM. Statin treatment and reduced risk of pneumonia in patients with diabetes. Thorax. 2006;61(11):957-61.
33. Chan KA, Andrade SE, Boles M, Buist DSM, Chase GA, Donahue JG, et al. Inhibitors of hydroxymethylglutaryl-coenzyme A reductase and risk of fracture among older women. The Lancet. 2000;355(9222):2185-8.
34. Lee Y-C, Lin C-H, Wu R-M, Lin M-S, Lin J-W, Chang C-H, et al. Discontinuation of statin therapy associates with Parkinson disease. A population-based study. 2013;81(5):410-6.
35. Cholesterol Treatment Trialists C. Protocol for analyses of adverse event data from randomised controlled trials of statin therapy. Am Heart J. 2016;176:63-9.
36. Fisher RA. Statistical Methods for Research Workers. Journal of the Royal Statistical Society. 1929;92(1):101-3.
37. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. BMJ. 1995;311(7003):485.

38. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
39. Aubert J, Bar-Hen A, Daudin JJ, Robin S. Determination of the differentially expressed genes in microarray experiments using local FDR. BMC Bioinformatics. 2004;5 (no pagination)(125).
40. Nguyen DV. On estimating the proportion of true null hypotheses for false discovery rate controlling procedures in exploratory DNA microarray studies. Computational Statistics and Data Analysis. 2004;47(3):611-37.
41. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. Journal of Educational and Behavioral Statistics. 2000;25(1):60-83.
42. Sedgwick P. Multiple hypothesis testing and Bonferroni's correction. BMJ : British Medical Journal. 2014;349.
43. Murphy C, Deplazes E, Cranfield CG, Garcia A. The Role of Structure and Biophysical Properties in the Pleiotropic Effects of Statins. Int J Mol Sci. 2020;21(22):8745.
44. Chou R, Dana T, Blazina I, Daeges M, Bougatsos C, Grusing S, et al. Statin Use for the Prevention of Cardiovascular Disease in Adults: A Systematic Review for the U.S. Preventive Services Task Force: Agency for Healthcare Research and Quality (US), Rockville (MD); 2016 2016.
45. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;139(25):e1082-e143.
46. NHS. NHS Business Authority: NHS Prescription Services; Visited in 22 June 2021; [Available from: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff].
47. Staffa JA, Chang J, Green L. Cerivastatin and Reports of Fatal Rhabdomyolysis. New England Journal of Medicine. 2002;346(7):539-40.
48. Dunnett CW, Gent M. An alternative to the use of two-sided tests in clinical trials. Statistics in Medicine. 1996;15(16):1729-38.
49. Bauer P, Hackl P, Hommel G, Sonnemann E. Multiple testing of pairs of one-sided hypotheses. Metrika. 1986;33(1):121-7.
50. Gu J, Shen S. Oracle and adaptive false discovery rate controlling methods for one-sided testing: theory and application in treatment effect evaluation. The Econometrics Journal. 2017;21(1):11-35.
51. Shaffer JP. Modified sequentially rejective multiple test procedures. Journal of the American Statistical Association. 1986;81(395):826-31.
52. Sarkar SK. Stepup procedures controlling generalized FWER and generalized FDR. Annals of Statistics. 2007;35(6):2405-20.
53. Lehmann EL, Romano JP. Generalizations of the Familywise Error Rate. In: Rojo J, editor. Selected Works of E L Lehmann. Boston, MA: Springer US; 2012. p. 719-35.
54. Sarkar SK, Guo W. On a generalized false discovery rate. Annals of Statistics. 2009;37(3):1545-65.
55. Genovese CR, Roeder K, Wasserman L. False discovery control with p-value weighting. Biometrika. 2006;93(3):509-24.
56. Farcomeni A. Generalized Augmentation to Control the False Discovery Exceedance in Multiple Testing. Scandinavian Journal of Statistics. 2009;36(3):501-17.
57. Genovese CR, Wasserman L. Exceedance control of the false discovery proportion. Journal of the American Statistical Association. 2006;101(476):1408-17.
58. Farcomeni A. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. Statistical Methods in Medical Research. 2007;17(4):347-88.
59. Benjamini Y. Discovering the false discovery rate. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2010;72(4):405-16.

60. Chi Z. False discovery rate control with multivariate p-values. Electronic Journal of Statistics. 2008;2:368-411.
61. Follmann D. Multivariate tests for multiple endpoints in clinical trials. Statistics in Medicine. 1995;14(11):1163-75.
62. Westfall PH, Young SS, Wright SP. On Adjusting P-Values for Multiplicity. Biometrics. 1993;49(3):941-5.
63. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics. 2001;29(4):1165-88.
64. Sarkar SK. Some results on false discovery rate in stepwise multiple testing procedures. Annals of Statistics. 2002;30(1):239-57.
65. Dunnett CW. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. Journal of the American Statistical Association. 1955;50(272):1096-121.
66. Dmitrienko A, D'Agostino Sr R. Traditional multiplicity adjustment methods in clinical trials. Statistics in Medicine. 2013;32(29):5172-218.
67. Maurer W, Bretz F. Multiple Testing in Group Sequential Trials Using Graphical Approaches. Statistics in Biopharmaceutical Research. 2013;5(4):311-20.
68. Mathias D, Michael DP. Multiple Testing and Error Control in Gaussian Graphical Model Selection. Statistical Science. 2007;22(3):430-49.
69. Liu J, Peissig P, Zhang C, Burnside E, McCarty C, Page D. Graphical-model Based Multiple Testing under Dependence, with Applications to Genome-wide Association Studies. Uncertain Artif Intell. 2012;2012:511-22.
70. Millen BA, Dmitrienko A. Chain Procedures: A Class of Flexible Closed Testing Procedures With Clinical Trial Applications. Statistics in Biopharmaceutical Research. 2011;3(1):14-30.
71. Wiens BL, Dmitrienko A. The Fallback Procedure for Evaluating a Single Family of Hypotheses. Journal of Biopharmaceutical Statistics. 2005;15(6):929-42.
72. Dmitrienko A, Wiens B, Westfall P. Fallback Tests in Dose-Response Clinical Trials. Journal of Biopharmaceutical Statistics. 2006;16(5):745-55.
73. Dmitrienko A, Wiens BL, Tamhane AC, Wang X. Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. Statistics in Medicine. 2007;26(12):2465-78.
74. Dmitrienko A, Tamhane AC. Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. Statistics in Medicine. 2011;30(13):1473-88.
75. Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. Journal of Statistical Planning and Inference. 2001;99(1):25-40.
76. Hommel G, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. Statistics in Medicine. 2007;26(22):4063-73.
77. Hommel G. Adaptive Modifications of Hypotheses After an Interim Analysis. Biometrical Journal. 2001;43(5):581-9.
78. Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes Analysis of a Microarray Experiment. Journal of the American Statistical Association. 2001;96(456):1151-60.
79. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. Annals of Statistics. 2003;31(6):2013-35.
80. Ausín MC, Gómez-Villegas MA, González-Pérez B, Rodríguez-Bernal MT, Salazar I, Sanz L. Bayesian analysis of multiple hypothesis testing with applications to microarray experiments. Communications in Statistics - Theory and Methods. 2011;40(13):2276-91.
81. Bertolino F, Piccinato L, Racugno W. Multiple Bayes factors for testing hypotheses. Journal of the American Statistical Association. 1995;90(429):213-9.
82. Van Der Laan MJ, Birkner MD, Hubbard AE. Empirical Bayes and resampling based multiple testing procedure controlling tail probability of the proportion of false positives. Statistical Applications in Genetics and Molecular Biology. 2005;4(1):i-30.
83. Whittemore AS. A Bayesian false discovery rate for multiple testing. Journal of Applied Statistics. 2007;34(1):1-9.

84. Berry SM, Berry DA. Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model. Biometrics. 2004;60(2):418-26.
85. Berry DA, Hochberg Y. Bayesian perspectives on multiple comparisons. Journal of Statistical Planning and Inference. 1999;82(1):215-27.
86. Amy Xia H, Ma H, Carlin BP. Bayesian Hierarchical Modeling for Detecting Safety Signals in Clinical Trials. Journal of Biopharmaceutical Statistics. 2011;21(5):1006-29.
87. Ahmed I, Thiessard F, Miremont-Salamé G, Bégaud B, Tubert-Bitter P. Pharmacovigilance data mining with methods based on false discovery rates: A comparative simulation study. Clinical Pharmacology and Therapeutics. 2010;88(4):492-8.
88. Bland JM, Altman DG. Multiple significance tests: The Bonferroni method. BMJ. 1995;310(6973):170.
89. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979;6(2):65-70.
90. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. American Journal of Public Health. 1996;86(5):726-8.
91. Shaffer JP. Multiple hypothesis testing. Annual Review of Psychology1995. p. 561-84.
92. Brown BW, Russell K. Methods of correcting for multiple testing: operating characteristics. Statistics in Medicine.16(22):2511-28.
93. Hochberg Y. A sharper bonferroni procedure for multiple tests of significance. Biometrika. 1988;75(4):800-2.
94. Simes RJ. An improved bonferroni procedure for multiple tests of significance. Biometrika. 1986;73(3):751-4.
95. Levin B. On the Holm, Simes, and Hochberg multiple test procedures. American journal of public health. 1996;86(5):628-9.
96. Hochberg Y, Rom D. Extensions of multiple testing procedures based on Simes' test. Journal of Statistical Planning and Inference. 1995;48(2):141-52.
97. Rom DM. An improved Hochberg procedure for multiple tests of significance. British Journal of Mathematical & Statistical Psychology.66(1):189-96.
98. Holland B. On the application of three modified Bonferroni procedures to pairwise multiple comparisons in balanced repeated measures designs. Computational Statistics Quarterly. 1991.
99. Rom DM. An improved Hochberg procedure for multiple tests of significance. British Journal of Mathematical and Statistical Psychology. 2013;66(1):189-96.
100. Hommel G. A stagewise rejective multiple test procedure based on a modified bonferroni test. Biometrika. 1988;75(2):383-6.
101. Blakesley RE, Mazumdar S, Dew MA, Houck PR, Tang G, Reynolds CF, 3rd, et al. Comparisons of methods for multiple hypothesis testing in neuropsychological research. Neuropsychology. 2009;23(2):255-64.
102. Dunnett CW, Tamhane AC. A step-up multiple test procedure. Journal of the American Statistical Association. 1992;87(417):162-70.
103. Ludbrook J. Multiple comparison procedures updated. Clinical and Experimental Pharmacology and Physiology. 1998;25(12):1032-7.
104. Benjamini Y, Liu W. A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. Journal of Statistical Planning and Inference. 1999;82(1-2):163-70.
105. Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. Statistical Methods in Medical Research. 2004;13(3):227-38.
106. Mehrotra DV, Adewale AJ. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. Statistics in Medicine. 2012;31(18):1918-30.

107. Cholesterol Treatment Trialists Collaboration. Harmonization of large-scale, heterogeneous individual participant adverse event data from randomized trials of statin therapy. Clinical Trials. 2022.
108. Murdoch DJ, Tsai Y-L, Adcock J. P-Values are Random Variables. The American Statistician. 2008;62(3):242-5.
109. Medical Dictionary for Regulatory Activities (https://www.meddra.org/ , last visited on 14 Sep 2018).
110. Tobert JA, Newman CB. The nocebo effect in the context of statin intolerance. Journal of Clinical Lipidology. 2016;10(4):739-47.
111. Gupta A, Thompson D, Whitehouse A, Collier T, Dahlof B, Poulter N, et al. Adverse events associated with unblinded, but not with blinded, statin therapy in the Anglo-Scandinavian Cardiac Outcomes Trial—Lipid-Lowering Arm (ASCOT-LLA): a randomised double-blind placebo-controlled trial and its non-randomised non-blind extension phase. The Lancet. 2017;389(10088):2473-81.
112. Early Breast Cancer Trialists' C, Group. Treatment of early breast cancer: worldwide evidence 1985–1990. Oxford: Oxford University Press. 1990.
113. Bland JM, Altman DG. The logrank test. Bmj. 2004;328(7447):1073.
114. Peto R, Pike M, Armitage P, Breslow N, Cox D, Howard SV, et al. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. I. Introduction and design. British journal of cancer. 1976;34(6):585-612.
115. Peto R, Pike M, Armitage P, Breslow NE, Cox D, Howard S, et al. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. II. analysis and examples. British journal of cancer. 1977;35(1):1-39.
116. Sackrowitz H, Samuel-Cahn E. P Values as Random Variables—Expected P Values. The American Statistician. 1999;53(4):326-31.
117. Hoeffding W. On the Distribution of the Number of Successes in Independent Trials. Ann Math Statist. 1956;27(3):713-21.
118. Millstein J, Volfson D. Computationally efficient permutation-based confidence interval estimation for tail-area FDR. Front Genet. 2013;4(SEP).
119. Blocker C, Conway J, Demortier L, Heinrich J, Junk T, Lyons L, et al. Simple facts about p-values. CDF8023. 2006.
120. Nedelman J, Wallenius T. Bernoulli Trials, Poisson Trials, Surprising Variances, and Jensen's Inequality. The American Statistician. 1986;40(4):286-9.
121. Habibzadeh F, Habibzadeh P. Shuffle randomization. Croatian Medical Journal. 2015;56(4):383.
122. Cholesterol Treatment Trialists Collaboration. Effect of statin therapy on muscle symptoms: an individual participant data meta-analysis of large-scale, randomised, double-blind trials. The Lancet. 2022.
123. Standardised MedDRA Queries: https://www.meddra.org/standardised-meddra-queries. (accessed on 25 September 2022).

Appendix

Appendix Table 1: CDISC SDTM (V3.2) domains and variables created for the completion of the CTT AE project including appropriate modifications

Domain model	Variable Name	Variable Label	Type (length)	Description
Adverse Event [AE]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	AETERM*	Reported Term for the Adverse Event	Char ($200)	Verbatim name of the event.
	AEDECOD*	Dictionary-Derived Term	Char ($200)	Dictionary-derived text description of AETERM or AEMODIFY. Equivalent to the Preferred Term (PT in MedDRA). The sponsor is expected to provide the dictionary name and version used to map the terms utilizing the define.xml external codelist attributes.
	AESEV	Severity/Intensity	Char ($20)	The severity or intensity of the event. Examples: MILD, MODERATE, SEVERE.
	AESER	Serious Event	Char ($1)	Is this a serious event?
	AEREL	Causality	Char ($200)	Records the investigator's opinion as to the causality of the event to the treatment. ICH E2A and E2B examples include NOT RELATED, UNLIKELY RELATED, POSSIBLY RELATED, RELATED. Controlled Terminology may be defined in the future. Check with regulatory authority for population of this variable.
	AESCONG†	Congenital Anomaly or Birth Defect	Char ($1)	Was the serious event associated with congenital anomaly or birth defect?
	AESDISAB†	Persist or Signif Disability/Incapacity	Char ($1)	Did the serious event result in persistent or significant disability/incapacity?
	AESDTH†	Results in Death	Char ($1)	Did the serious event result in death?
	AESHOSP†	Requires or Prolongs Hospitalization	Char ($1)	Did the serious event require or prolong hospitalization?
	AESLIFE†	Is Life Threatening	Char ($1)	Was the serious event life threatening?
	AESMIE†	Other Medically Important Serious Event	Char ($1)	Do additional categories for seriousness apply?
	AESTDY	Study Day of Start of Selected Adverse Event	Num	Study Day of Start of Selected Adverse Event, measured as integer days.
	SUPP_CTT_AE_TERM	Adverse event term selected for CTT analyses	Char ($200)	Supplementary variable: Adverse event term selected for CTT analyses, either AETERM or AEDECOD.
	SUPP_CTT_AE_STOP	Type of treatment stop‡	Char ($1)	Supplementary variable: Type of treatment stop P = Permanent statin stop T = Temporary statin stop S = Statin stop, unknown if permanent or temporary R = Statin dose reduced N = Did not lead to a statin stop

Clinical Event [CE]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	CETERM*	Reported Term for the Clinical Event	Char ($200)	Term for the medical condition or event. Most likely pre-printed on CRF.
	CEDECOD*	Dictionary-Derived Term	Char ($200)	Controlled terminology for the name of the clinical event. The sponsor is expected to provide the dictionary name and version used to map the terms utilizing the define.xml external codelist attributes
	CESEV	Severity/Intensity	Char ($50)	The severity or intensity of the event. Examples: MILD, MODERATE, SEVERE
	CEDY	Study Day of Event Collection	Num	Study day of clinical event collection, measured as integer days.
	SUPP_CTT_CE_TERM	Clinical event term selected for CTT analyses	Char ($200)	Supplementary variable: Clinical event term selected for CTT analyses, either CETERM or CEDECOD.
Death Details [DD]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	DDORRES	Result or Finding as Collected	Char ($200)	Result of the test defined in DDTEST, as originally received or collected.
	DDDY	Study Day of Collection	Num	Study day of the collection, in integer days.
Health encounters [HO]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	HOTERM*	Reported Term for the Healthcare Encounter	Char ($200)	Verbatim or preprinted CRF term for the healthcare encounter.
	HODECOD*	Dictionary-Derived Term for the Healthcare Encounter	Char ($200)	Dictionary or sponsor-defined derived text description of HOTERM or the modified topic variable (HOMODIFY).
	HOSTDY	Study Day of Start of Healthcare Encounter	Num	Study Day of Start of Selected Healthcare Encounter, measured as integer days.
	SUPP_CTT_HO_TERM	Hospitalisation event term selected for CTT analyses	Char ($200)	Supplementary variable: Hospitalisation event term selected for CTT analyses, either HOTERM or HODECOD.
Procedures [PR]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	PRTRT*	Reported Name of Procedure	Char ($200)	Name of procedure performed, either pre-printed or collected on a CRF.
	PRDECOD*	Standardized Procedure Name	Char ($200)	Standardized or dictionary-derived name of PRTRT. The sponsor is expected to provide the dictionary name and version used to map the terms in the define.xml external codelist attributes. If an intervention term does not have a decode value in the dictionary then PRDECOD will be null.

	PRSTDY	Study Day of Start of Procedure	Num	Study day of start of procedure, measured as integer days.
	SUPP_CTT_PR_TERM	Procedure event term selected for CTT analyses	Char ($200)	Supplementary variable: Procedure event term selected for CTT analyses, either PRTRT or PRDECOD.
Concomitant/Prior Medications [CM]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	CMTRT*	Reported Name of Drug, Med, or Therapy	Char ($200)	Verbatim medication name that is either pre-printed or collected on a CRF.
	CMDECOD*	Standardized Medication Name	Char ($200)	Standardized or dictionary-derived text description of CMTRT or CMMODIFY. Equivalent to the generic medication name in WHO Drug. The sponsor is expected to provide the dictionary name and version used to map the terms utilizing the define.xml external codelist attributes. If an intervention term does not have a decode value in the dictionary then CMDECOD will be left blank.
	CMENDY	Study Day of End of Medication	Num	Study day of end of medication relative to the sponsor-defined RFSTDTC.
	CMSTDY	Study Day of Start of Selected Co-medication term	Num	Study Day of Start of Medication, measured as integer days.
	SUPP_CTT_CM_TERM	Co-medication term selected for CTT analyses	Char ($200)	Supplementary variable: Co-medication term selected for CTT analyses either CMTRT or CMDECOD.
	SUPP_CMBLFL	Baseline Flag	Char ($1)	Supplementary variable: Indicator used to identify a baseline value. The value should be “Y” or null.
Demographics [DM]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be a compound identifier formed by concatenating STUDYID-SITEID-SUBJID.
	AGE	Age	Num	Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC, but BRTHDTC may not be available in all cases (due to subject privacy concerns).
	AGEU	Age Units	Char ($20)	Units associated with AGE.
	SEX	Sex	Char ($1)	Sex of the subject.
	RACE*	Race	Char ($20)	Race of the subject. Sponsors should refer to “Collection of Race and Ethnicity Data in Clinical Trials” (FDA, September 2005) for guidance regarding the collection of race (http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm) See Assumption below regarding RACE.
	ETHNIC*	Ethnicity	Char ($20)	The ethnicity of the subject. Sponsors should refer to “Collection of Race and Ethnicity Data in Clinical Trials” (FDA, September 2005) for guidance regarding the collection of ethnicity (http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm).
	ACTARMCD	Actual Arm Code	Char ($200)	Code of actual Arm. When an Arm is not planned (not in Trial Arms), ACTARMCD will be UNPLAN. Randomised subjects who were not treated will be given a value of NOTTRT. Values should be “SCRNFAIL” for screen failures and “NOTASSGN” for subjects not assigned to treatment. Restricted to values in Trial Arms in all other cases. ACTARMCD is limited to 20 characters and does not have special character restrictions. The maximum length of ACTARMCD is longer than for other short variables to accommodate the kind of values that are likely to be needed for crossover trials.
	ACTARM	Description of Actual Arm	Char ($200)	Description of actual Arm. When an Arm is not planned (not in Trial Arms), ACTARM will be “Unplanned Treatment”. Randomised subjects who were not treated will be given a value of “Not Treated”. Values should be “Screen Failure” for screen failures and “Not Assigned” for subjects not assigned to treatment. Restricted to values in Trial Arms in all other cases.

	SUPP_CTT_END_DAY	Reference End Day for the subject after randomisation	Num	Supplementary variable: Reference End Day for the subject after randomisation, measured as integer days. Defined as date of death or last date patient known to be alive.
	SUPP_AGEGRP	Age in 5 years categories	Char ($20)	Supplementary variable: Age in 5 years categories. Required only if variable AGE was missing.
Exposure [EX]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	EXADJ	Reason for Dose Adjustment	Char ($200)	Describes reason or explanation of why a dose is adjusted.
	EXENDY	Study Day of End of Treatment	Num	Study day of end of treatment.
	SUPP_CTT_EX_STOP	Type of treatment stop‡	Char ($1)	Supplementary variable: Type of treatment stop P = Permanent statin stop T = Temporary statin stop S = Statin stop, unknown if permanent or temporary R = Statin dose reduced N = Did not lead to a statin stop
Laboratory Test Results [LB]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	LBTESTCD	Lab Test or Examination Short Name	Char ($200)	Short name of the measurement, test, or examination described in LBTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in LBTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g.”1TEST”). LBTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: ALT, LDH.
	LBTEST	Lab Test or Examination Name	Char ($200)	Verbatim name of the test or examination used to obtain the measurement or finding. Note any test normally performed by a clinical laboratory is considered a lab test. The value in LBTEST cannot be longer than 40 characters. Examples: Alanine Aminotransferase, Lactate Dehydrogenase.
	LBORRES	Result or Finding in Original Units	Char ($20)	Result of the measurement or finding as originally received or collected.
	LBORRESU	Original Units	Char ($20)	Original units in which the data were collected. The unit for LBORRES. Example: g/L.
	LBORNRHI	Reference Range Upper Limit in Orig Unit	Char ($20)	Upper end of reference range for continuous measurements in original units. Should be populated only for continuous results.
	LBSTRESN	Numeric Result/Finding in Standard Units	Num	Used for continuous or numeric results or findings in standard format; copied in numeric format from LBSTRESC. LBSTRESN should store all numeric test results or findings.
	LBSTRESU	Standard Units	Char ($20)	Standardized unit used for LBSTRESC or LBSTRESN.
	LBSTNRHI	Reference Range Upper Limit-Std Units	Num	Upper end of reference range for continuous measurements in standardized units. Should be populated only for continuous results.
	LBNRIND	Reference Range Indicator	Char ($20)	1. Indicates where the value falls with respect to reference range defined by LBORNRLO and LBORNRHI, LBSTNRLO and LBSTNRHI, or by LBSTNRC. Examples: NORMAL, ABNORMAL, HIGH, LOW. 2. Sponsors should specify in the study metadata (Comments column in the define.xml) whether LBNRIND refers to the original or standard reference ranges and results. 3.Should not be used to indicate clinical significance.
	LBBLFL	Baseline Flag	Char ($1)	Indicator used to identify a baseline value. The value should be “Y” or null.

	LBFAST	Fasting Status	Char ($1)	Indicator used to identify fasting status such as Y, N, U, or null if not relevant. For CTT purposes, it is only required for glucose tests.
	LBDRVFL	Derived Flag	Char ($1)	Used to indicate a derived record. The value should be Y or null. Records that represent the average of other records, or do not come from the CRF, or are not as originally received or collected are examples of records that might be derived for the submission datasets. If LBDRVFL=Y, then LBORRES may be null, with LBSTRESC, and (if numeric) LBSTRESN having the derived value.
	LBDY	Study Day of Specimen Collection	Num	Study day of specimen collection, measured as integer days.
	LBTPT	Planned Time Point Name	Char ($200)	Planned Time Point Name Required only for Oral Glucose Tolerance Tests (OGTTs) measurements
Medical History [MH]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	MHTERM*	Reported Term for the Medical History	Char ($200)	Verbatim or preprinted CRF term for the medical condition or event.
	MHDECOD*	Dictionary-Derived Term	Char ($200)	Dictionary-derived text description of MHTERM or MHMODIFY. Equivalent to the Preferred Term (PT in MedDRA). The sponsor is expected to provide the dictionary name and version used to map the terms utilizing the define.xml external codelist attributes.
	MHCAT	Category for Medical History	Char ($200)	Used to define a category of related records. Examples: CARDIAC or GENERAL
	MHDY	Study Day of History Collection	Num	Study day of medical history collection, measured as integer days.
	SUPP_CTT_MH_TERM	Medical history term selected for CTT analyses	Char ($200)	Supplementary variable: Medical history term selected for analyses
Substance Use [SU]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	SUTRT	Reported Name of Substance	Char ($200)	Substance name. Examples: Cigarettes, Coffee.
	SUDOSTXT	Substance Use Consumption Text	Char ($200)	Substance use consumption amounts or a range of consumption information collected in text form. Not populated if SUDOSE is populated.
	SUSTDY	Study Day of Start of Substance Use	Num	Study day of start of substance use relative to randomisation date, measured as integer days.
	SUPP_SUBLFL	Baseline flag	Char ($1)	Supplementary variable: Baseline flag
Subject Visit [SV]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	VISITNUM	Visit Number	Num	1. Clinical encounter number. (Decimal numbering may be useful for inserting unplanned visits.) 2. Numeric version of VISIT, used for sorting.
	VISIT	Visit Name	Char ($50)	1. Protocol-defined description of clinical encounter. 2. May be used in addition to VISITNUM and/or VISITDY as a text description of the clinical encounter.

	SVSTDY	Study Day of Start of Visit	Num	Study day of start of visit relative to randomisation date, measured as integer days.
Vital Signs [VS]	STUDYID	Study Identifier	Char ($20)	Unique identifier for a study.
	DOMAIN	Domain Abbreviation	Char ($2)	Two-character abbreviation for the domain.
	USUBJID	Unique Subject Identifier	Char ($50)	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
	VSTESTCD	Vital Signs Test Short Name	Char ($200)	Short name of the measurement, test, or examination described in VSTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in VSTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g."1TEST"). VSTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: SYSBP, DIABP, BMI.
	VSTEST	Vital Signs Test Name	Char ($200)	Verbatim name of the test or examination used to obtain the measurement or finding. The value in VSTEST cannot be longer than 40 characters. Examples: Systolic Blood Pressure, Diastolic Blood Pressure, Body Mass Index.
	VSORRES	Result or Finding in Original Units	Char ($20)	Result of the vital signs measurement as originally received or collected.
	VSORRESU	Original Units	Char ($20)	Original units in which the data were collected. The unit for VSORRES. Examples: IN, LB, BEATS/MIN.
	VSSTRESN	Numeric Result/Finding in Standard Units	Num	Used for continuous or numeric results or findings in standard format; copied in numeric format from VSSTRESC. VSSTRESN should store all numeric test results or findings.
	VSSTRESU	Standard Units	Char ($20)	Standardized unit used for VSSTRESC and VSSTRESN.
	VSBLFL	Baseline Flag	Char ($1)	Indicator used to identify a baseline value. The value should be "Y" or null.
	VSDY	Study Day of Vital Signs	Num	Study day of vital signs measurements, measured as integer days.
	SUPP_VSORRESGRP	Results as originally received in categories	Char ($20)	Supplementary variable: Result of the vital signs measurement as originally received or collected (ie, categorized). Required only if variable VSORRES was missing

* At least one of the two variables within each domain was required for the completion of this project † Only to be created if variable AESER cannot be created.

Appendix Table 2: Adverse event terms reported in MedDRA (V20.0) where a hypothesis of association with statin is known or has been previously examined in CTT

MedDRA hierarchical level	Categorization of adverse event terms			
	Lipid	**Muscle**	**Diabetes**	**Major vascular event**
Preferred Terms	Dyslipidaemia, Hypo hdl cholesterolaemia, Hyper hdl cholesterolaemia, Lipid metabolism disorder, Hypolipidaemia, Hypotriglyceridaemia, Hypocholesterolaemia, Cholesterosis, Hypercholesterolaemia, Remnant hyperlipidaemia, Acquired mixed hyperlipidaemia, Hyperchylomicronaemia, Autoimmune hyperlipidaemia, Hyperlipidaemia, Hypertriglyceridaemia, Free fatty acids increased, Carnitine abnormal, Carnitine decreased, Carnitine increased, Carnitine normal, Lipids, Lipids abnormal, Lipids decreased, Lipids increased, Lipids normal, Apolipoprotein normal, Apolipoprotein abnormal, Apolipoprotein increased, Apolipoprotein decreased, Apolipoprotein, Apolipoprotein b/apolipoprotein a-1 ratio, Apolipoprotein a-ii abnormal, Carnitine, Apolipoprotein b decreased, Apolipoprotein a-ii normal, Apolipoprotein b, Apolipoprotein a-ii decreased, Apolipoprotein b abnormal, Apolipoprotein a-ii, Apolipoprotein b increased, Apolipoprotein b normal, Apolipoprotein a-i, Apolipoprotein a-ii increased, Apolipoprotein a-i abnormal, Apolipoprotein a-i decreased, Apolipoprotein a-i increased, Apolipoprotein a-i normal, Apolipoprotein c abnormal, Phytosterol level increased, Phytosterol level, Apolipoprotein c, Apolipoprotein e abnormal, Free fatty acids, Apolipoprotein e, Free fatty acids abnormal, Serum pristanic acid increased, Phytanic acid increased, Lecithin-cholesterol cyltransferase activity decreased, Apolipoprotein b/apolipoprotein a-1 ratio increased, Lecithin-cholesterol acyltransferase activity increased, Free fatty acids decreased, Lipaemic index score, Blood triglycerides increased, Blood triglycerides normal, Blood triglycerides decreased, Blood triglycerides abnormal, Blood triglycerides, Low density lipoprotein,	Myoglobinuria, Myalgia, Myalgia intercostal, Fibromyalgia, Muscle spasms, Muscle fatigue, Myoglobinaemia, Myopathy, Myopathy toxic, Rhabdomyolysis, Muscle necrosis, Necrotising myositis, Myositis, Muscular weakness, Sarcopenia, Groin pain, Limb discomfort, Musculoskeletal chest pain, Musculoskeletal discomfort, Pain in extremity, Musculoskeletal pain, Neck pain, Flank pain, Back pain, Blood creatine phosphokinase mm increased, Blood creatine phosphokinase abnormal, Blood creatine phosphokinase increased, Muscle enzyme increased, Myoglobin blood present, Myoglobin urine present, Myoglobin blood increased.	Diabetic gastropathy, Diabetic gastroenteropathy, Diabetic enteropathy, Diabetic gastroparesis, Gestational diabetes, Mauriac syndrome, Diabetic complication, Diabetic dyslipidaemia, Diabetic endorgan damage, Diabetic metabolic decompensation, Diabetic ketoacidosis, Hyperglycaemic hyperosmolar nonketotic syndrome, Diabetes mellitus malnutrition-related, Insulin resistant diabetes, Pancreatogenous diabetes, Latent autoimmune diabetes in adults, Insulin-requiring type 2 diabetes mellitus, Diabetes mellitus, Diabetes mellitus inadequate control, Diabetes with hyperosmolarity, Diabetes complicating pregnancy, Acquired lipoatrophic diabetes, Monogenic diabetes, Type 3 diabetes mellitus, Fulminant type 1 diabetes mellitus, Type 2 diabetes mellitus, Type 1 diabetes mellitus, Diabetic gangrene, Diabetic foot infection, Diabetic vascular disorder, Diabetic macroangiopathy, Diabetic microangiopathy, Diabetic arteritis, Diabetic cheiroarthropathy, Diabetic dermopathy, Diabetic bullosis, Necrobiosis lipoidica diabeticorum, Diabetic neuropathic ulcer, Diabetic foot, Diabetic ulcer, Cataract diabetic, Diabetic keratopathy, Diabetic eye disease, Diabetic glaucoma, Diabetic uveitis, Macular oedema, Diabetic retinal oedema, Diabetic ophthalmoplegia, Maculopathy, Retinal neovascularisation, Vitreous haemorrhage, Retinopathy	Sudden cardiac death, Death, Pituitary infarction, Percutaneous coronary intervention, Coronary brachytherapy, Coronary artery bypass, Coronary endarterectomy, Coronary revascularisation, Coronary arterial stent insertion, Coronary angioplasty, Peripheral revascularisation, Vascular operation, Post procedural myocardial infarction, Coronary vascular graft occlusion, Coronary bypass stenosis, Periprocedural myocardial infarction, Coronary vascular graft stenosis, Coronary artery reocclusion, Coronary bypass thrombosis, Coronary artery restenosis, Cerebral artery restenosis, Coronary artery stenosis, Coronary artery thrombosis, Coronary ostial stenosis, Coronary artery disease, Coronary artery occlusion, Arteriosclerosis coronary artery, Coronary artery insufficiency, Angina unstable, Angina pectoris, Acute myocardial infarction, Subendocardial ischaemia, Postinfarction angina, Silent myocardial infarction, Acute coronary syndrome, Microvascular coronary artery disease, Myocardial ischaemia, Myocardial infarction, Papillary muscle infarction, Cardiac failure, Cardiac failure acute, Haemorrhagic cerebral infarction, Lateral medullary syndrome, Cerebral artery occlusion, Cerebral artery thrombosis, Cerebral artery embolism, Cerebellar infarction,

	Low density lipoprotein increased, Low density lipoprotein normal, Low density lipoprotein decreased, Low density lipoprotein abnormal, Lipoprotein (a) abnormal, Lipoprotein (a) normal, Lipoprotein (a) decreased, Lipoprotein (a), Lipoprotein (a) increased, Total cholesterol/hdl ratio abnormal, Total cholesterol/hdl ratio normal, Total cholesterol/hdl ratio, Total cholesterol/hdl ratio decreased, Total cholesterol/hdl ratio increased, Ldl/hdl ratio increased, Very low density lipoprotein normal, Very low density lipoprotein abnormal, Very low density lipoprotein increased, Very low density lipoprotein decreased, Very low density lipoprotein, Chylomicrons, Ldl/hdl ratio decreased, Ldl/hdl ratio, Blood cholesterol decreased, Blood cholesterol abnormal, Blood cholesterol, Blood cholesterol normal, Blood cholesterol increased, Blood cholesterol esterase increased, Non-high-density lipoprotein cholesterol increased, Non-high-density lipoprotein cholesterol, Intermediate density lipoprotein decreased, Intermediate density lipoprotein increased, Non-high-density lipoprotein cholesterol decreased, Chylomicron decreased, Chylomicron increased, Cholesterol absorption efficiency decreased, Remnant-like lipoprotein particles, Remnant-like lipoprotein particles increased.		proliferative, Retinopathy, Diabetic retinopathy, Diabetic blindness, Retinal laser coagulation, Diabetes mellitus management, Diabetic hepatopathy, Diabetic cystopathy, Diabetic nephropathy, Diabetic end stage renal disease, Diabetic foetopathy, Diabetic arthropathy, Diabetic amyotrophy, Diabetic mastopathy, Diabetic cardiomyopathy, Diabetic encephalopathy, Diabetic coma, Diabetic hyperosmolar coma, Diabetic ketoacidotic hyperglycaemic coma, Diabetic hyperglycaemic coma, Diabetic autonomic neuropathy, Diabetic neuropathy, Diabetic mononeuropathy, Acute painful neuropathy of rapid glycaemic control.	Cerebellar artery thrombosis, Cerebrovascular accident, Cerebral ischaemia, Cerebral thrombosis, Brain stem ischaemia, Brain stem infarction, Embolic stroke, Thalamic infarction, Brain stem thrombosis, Basilar artery thrombosis, Thrombotic cerebral infarction, Cerebellar embolism, Embolic cerebral infarction, Ischaemic cerebral infarction, Ischaemic stroke, Cerebral vascular occlusion, Precerebral artery thrombosis, Brain stem embolism, Cerebral microembolism, Basal ganglia infarction, Cerebellar ischaemia, Precerebral artery occlusion, Lacunar infarction, Cerebellar artery occlusion, Thrombotic stroke, Basilar artery occlusion, Basilar artery stenosis, Cerebral arteriosclerosis, Cerebral artery stenosis, Cerebrovascular stenosis, Precerebral arteriosclerosis, Delayed ischaemic neurological deficit, Cerebral small vessel ischaemic disease, Reversible ischaemic neurological deficit, Transient ischaemic attack.
Higher Level Terms	Lipid metabolism and deposit disorders NEC, Elevated cholesterol hyperlipidaemias NEC, Elevated triglycerides, lipoprotein and lipid tests NEC, Triglyceride analyses, Cholesterol analyses	Muscle pains, Myopathies, Muscle weakness conditions.		Death and sudden death, Site specific vascular disorders nec, Vascular therapeutic procedures nec, Cardiac therapeutic procedures nec, Coronary artery disorders nec, Ischaemic coronary artery disorders, Transient cerebrovascular events.
Higher Level Group Terms	Lipid metabolism disorders, Lipid analyses			Vascular disorders nec, Vascular therapeutic procedures, Cardiac therapeutic procedures, Cardiac and vascular investigations (excl enzyme tests), Coronary artery disorders.
System Organ Class Terms		Musculoskeletal and connective tissue disorders.		Vascular disorders.

MedDRA V20.0 is used to code all adverse event terms.

Appendix B

Mathematical calculations: showing the dependency between the minimal 2-sided relative risk which can be detected as significant by a chi-square distribution at α-level and the total number of events.

Annotations:

- E_1 = events in statin arm
- E_2 = events in placebo arm
- N_1 = number of participants in statin arm
- N_2 = number of participants in placebo arm
- RR = Relative Risk

When conducing the simulation analysis, events in statin arm are generated from a binomial distribution:

$$E_1 \sim \text{Binomial}(C, p), \quad \text{where} \quad C = E_1 + E_2 \quad \text{and} \quad p = \frac{RR}{1+RR} \qquad (1)$$

The OE statistic and its' variance V can be calculated from the following formulas:

$$OE = E_1 - \frac{N_1 C}{Z} \quad \& \quad V = \frac{\frac{N_1 C}{Z} * (1 - \frac{N_1}{Z}) * (Z - C)}{Z - 1}, \quad \text{where} \quad C = E_1 + E_2, \; Z = N_1 + N_2$$

Based on (1) and when assuming that $N_1 = N_2 = \frac{Z}{2}$:

$$E_1 = pC = \frac{RR}{1 + RR} C$$

$$\Leftrightarrow$$

$$OE = \frac{RR}{1+RR} C - \frac{C}{2} = \frac{C}{2}\left(\frac{RR-1}{RR+1}\right) \quad \&$$

$$V = \frac{\frac{C}{2} * (1 - \frac{1}{2}) * (Z - C)}{Z - 1} = \frac{C * (Z - 1) * (Z - C)}{2 * Z * (Z - 1)} = \frac{C * (Z - C)}{2 * Z}$$

The chi-square test can then be calculated as below:

$$x_1^2 = \frac{OE^2}{V}$$

$$\Leftrightarrow$$

$$x_1^2 = \frac{\frac{C^2}{4}\left(\frac{RR-1}{RR+1}\right)^2}{\frac{C*(Z-C)}{2*Z}} = \frac{C*Z\left(\frac{RR-1}{RR+1}\right)^2}{2*(Z-C)}$$

$$\Leftrightarrow$$

$$\left(\frac{RR-1}{RR+1}\right)^2 = \frac{C*Z}{2*(Z-C)*x_1^2}$$

$$\Leftrightarrow$$

$$\frac{RR-1}{RR+1} = \sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}$$

$$RR*\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}} - \sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}} = RR+1$$

$$RR*\left(\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}} - 1\right) = 1 + \sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}$$

$$RR = \frac{1+\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}}{\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}-1} = \left(\frac{\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}-1}{\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}+1}\right)^{-1}$$

$$= \left(\frac{2\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}} - \left(\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}+1\right)}{\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}+1}\right)^{-1}$$

$$= \left(\frac{2\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}}{\sqrt{\frac{C*Z}{2*(Z-C)*x_1^2}}+1} - 1\right)^{-1}$$

$$= \left(\frac{1}{\frac{1}{2}+\sqrt{\frac{(Z-C)}{2*C*Z}}\sqrt{x_1^2}} - 1\right)^{-1}$$

$$RR = \left(\frac{1}{\frac{1}{2}+\sqrt{\frac{(Z-C)}{2*C*Z}}\sqrt{x_1^2}} - 1\right)^{-1}$$

$$\Leftrightarrow$$

$$\mathbf{RR} = \left(\frac{\mathbf{1}}{\frac{\mathbf{1}}{\mathbf{2}}+\frac{\sqrt{\mathbf{x_1^2}}}{\mathbf{2C}}\sqrt{\frac{\boldsymbol{C(Z-C)}}{\boldsymbol{Z-1}}}} - \mathbf{1}\right)^{-1}$$