MAJOR BLOOD LIPIDS AND RISK OF ISCHAEMIC STROKE AND INTRACEREBRAL HAEMORRHAGE IN CHINESE ADULTS



Luanluan Sun

MBBS, MPH

Exeter College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2017

Abstract

Major blood lipids, and risk of ischaemic stroke and intracerebral haemorrhage in Chinese adults

Luanluan Sun
Exeter College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2017

Background Stroke is the second most common cause of premature death worldwide, and the leading cause of death and disability in China. Stroke is a clinical syndrome that includes three distinct pathological types (i.e., ischaemic stroke [IS], intracerebral haemorrhage [ICH] and subarachnoid haemorrhage [SAH]) that have a different underlying pathophysiology. Previous observational studies and randomised trials have reported conflicting results on the associations of major blood lipids, including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides with stroke pathological types. Moreover, the causal relevance of many of these associations is still not fully understood.

Methods The China Kadoorie Biobank (CKB) prospective study collected questionnaire, physical measurement data, and blood samples on 512 891 adults from 10 diverse regions in China in 2004-2008. By 1st January 2016, over 32 000 incident IS, and 8000 ICH cases were recorded. A nested case-control study of stroke included 16 541 adults (5475 IS, 4776 ICH, and 6290 controls), with measurements of baseline plasma concentrations of LDL-C, HDL-C, and triglycerides, and genotyping of known lipid-related genetic variants. Repeat measurements of blood lipids were obtained in random samples of participants at resurveys (2008 and 2013-2014) to correct for regression dilution bias. The associations of major blood lipids with stroke pathological types were examined after adjustment for confounders and correction for regression dilution bias. The causal relevance of these associations was assessed by Mendelian randomisation (MR) using genetic risk scores (GRSs) involving lipid-related genetic variants identified in Western populations. Multivariable logistic and Cox regressions were used to yield odds ratios (ORs) and hazard ratios (HRs), respectively, relating major blood lipids or genetic variants with risks of stroke pathological types.

Results Overall, the incidence rates of first IS and ICH were 761 and 187 cases per 100 000 person-years, respectively, with ~20% of the incident strokes attributable to ICH. The mean (SD) plasma concentrations of LDL-C, HDL-C, and triglycerides were 2.3 (0.8) mmol/L, 1.3 (0.3) mmol/L, and 1.6 (1.1) mmol/L, respectively. The regression dilution ratios at the mid-point of follow-up (around 3.7 years after baseline) were 0.68 for LDL-C, 0.72 for HDL-C, and 0.62 for triglycerides, respectively. Throughout the range of LDL-C examined (1.7-3.2 mmol/L), there was a positive log-linear association of usual LDL-C with risk of IS (HR 1.17; 95% CI 1.10-1.25 per 1 mmol/L higher LDL-C), and an inverse association with ICH (0.86; 0.80-0.92). Likewise, after adjusting for non-lipid vascular risk factors, usual plasma triglycerides were positively associated with risk of IS (1.02; 1.00-1.04 per 30% higher triglycerides), albeit weaker than for LDL-C, and were inversely associated with risk of ICH (0.94; 0.92-0.96), independent of LDL-C. HDL-C was inversely associated with risk of IS (0.93; 0.89-0.97 per 0.3 mmol/L higher HDL-C), but not with ICH. The proportional relevance of LDL-C, triglycerides, and HDL-C with IS differed significantly from those for ICH ($p_{heterogeneity}$: <0.00001, <0.00001, 0.0006, respectively). In the MR analyses, an externally weighted GRS of 8 known LDL-C related genetic variants was associated with plasma LDL-C ($p=1.15x10^{-272}$), but not associated with other lipid fractions or non-lipid major vascular risk factors. Each 1 mmol/L higher genetically-determined LDL-C was associated with an OR of 1.28 (1.05-1.56) for IS, similar to the risk estimates in both observational studies and randomised trials. There was no significant association of genetically-determined LDL-C with ICH (OR 1.04; 0.87-1.25), but the estimates did not differ significantly from those in the observational studies ($p_{heterogeneity}=0.06$), and randomised trials ($p_{heterogeneity}=0.8$). Neither genetically-determined HDL-C nor triglycerides were significantly associated with either IS or ICH.

Conclusion The associations of major blood lipids with stroke differed qualitatively between pathological types. Lower LDL-C is causally associated with lower risk of IS, and probably, higher risk of ICH. Better genetic instruments are needed to assess the causal relevance, if any, of the observed associations of HDL-C, and of triglycerides for stroke pathological types presented in this thesis.

Acknowledgements

Firstly, I would like to express my gratitude to participants in the China Kadoorie Biobank.

I would like to thank Prof. Zhengming Chen, Prof. Robert Clarke, and Dr. Derrick Bennett for supervising the thesis, and for their patient guidance, encouragement, and advice.

I would also like to thank Prof. Richard Peto for his time, encouragement, and lucid explanations of statistics, epidemiology, and astronomy, along the course; his teaching on reporting of the evidence, and discussion of the findings. I would like to thank Prof. Sasha Shepperd for her time, care, and her organised professionalism that helped me during my DPhil.

I would like to thank Prof. Colin Baigent, and colleagues in the Cholesterol Treatment Trialists' Collaboration for allowing me to conduct analyses on the most recent data, and for their academic advice on several aspects of the study that has been invaluable to my research.

I would like to acknowledge the wider China Kadoorie Biobank team, in Oxford, and China; particularly, Dr. Fiona Bragg for her support during my DPhil course, and Dr. Robin Walters, Dr. Iona Millwood, Dr. Ling Yang, and Dr. Huaidong Du. I have also been supported by the CKB Data, and IT groups.

I would like to acknowledge the Clarendon Scholarship, and the Nuffield Department of Population Health for providing financial support for my DPhil studies. I would like to acknowledge the peer support from the scholars' community, particularly, the helpful support from Gurdeep Mannu and Deon Simpson.

Finally, I would like to thank my family; my mum, for her support and friendship, 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. I performed all aspects of the literature review and planned and designed the analyses. I conducted all statistical analyses (some analyses used statistical programmes developed by colleagues within the Clinical Trial Service Unit and Epidemiological Studies Unit, as reported at the relevant sections in the thesis), and produced all the tables and figures presented. I interpreted the findings and wrote-up all aspects of the study.

Contents

Abstract I
Acknowledgements III
Statement of contribution to work IV
List of Tables IX
List of Figures XII
List of Abbreviations and Acronyms XVII
Chapter 1 Introduction 1
1.1 Stroke 3
1.1.1 Definition 3
1.1.2 Stroke pathological types and subtypes 4
1.1.3 Diagnosis 5
1.1.4 Pathophysiology 6
1.2 Burden of stroke 6
1.2.1 Stroke burden worldwide 6
1.2.2 Stroke burden in China 10
1.2.3 Risk factors for stroke 11
1.3 Lipids metabolism and homeostasis 12
1.3.1 Lipoprotein metabolism 12
1.3.2 Lipids and atherosclerosis 13
1.3.3 Determinants of blood lipids 15
1.4 Major lipids and risk of stroke pathological types 16
1.4.1 Major lipids and risk of ischaemic stroke 19
1.4.2 Major lipids and risk of haemorrhagic stroke 20
1.5 Overview of the thesis 21
1.5.1 Chapter structure 21
1.5.2 Summary 23
Chapter 2 Literature review 27
2.1 Objectives 27
2.2 Methodology 27
2.3 Prospective epidemiological evidence 28
2.3.1 Ischaemic stroke 28
2.3.2 Haemorrhagic stroke 39

2.4 Mendelian randomisation evidence 42
2.5 Randomised controlled trial evidence 46
2.5.1 Low-density lipoprotein cholesterol 46
2.5.2 High-density lipoprotein cholesterol 48
2.5.3 Triglycerides 50
2.6 Summary of findings and main research questions 51
2.7 Aims and objectives of thesis 51
Chapter 3 Overall study design and methods 62
3.1 Background 62
3.2 Overall study design and study population 64
3.2.1 Overall study design 64
3.2.2 Overall study population 65
3.3 Data collection and procedures 67
3.3.1 Baseline survey (2004-2008) 67
3.3.2 Resurveys 69
3.3.3 Long-term follow up for mortality and morbidity 69
3.3.4 Nested case-control study of stroke 71
3.4 Baseline characteristics of the study population 74
3.4.1 The study population 74
3.4.2 The resurvey participants 76
3.5 Summary 77
3.6 Tables 78
Chapter 4 Major lipids and risk of ischaemic stroke 83
4.1 Introduction 83
4.2 Study design and methods 84
4.2.1 The nested case-control study 84
4.2.2 Statistical analyses 85
4.3 Results 98
4.3.1 Incidence rates of ischaemic stroke in CKB participants 98
4.3.2 Baseline characteristics of the nested case-control study participants 99
4.3.3 Cross-sectional correlations of major lipids with other vascular factors 101
4.3.4 Regression dilution ratios of lipid-related biomarkers 107
4.3.5 Prospective associations of major lipids with ischaemic stroke 108

4.4 Discussion 115

4.4.1 Incidence rates of ischaemic stroke in CKB participants 115
4.4.2 Trends of changes in major lipids 119
4.4.3 Prospective associations of major lipids with ischaemic stroke 120
4.4.4 Strengths and limitations 128

4.5 Summary 130
4.6 Tables 132
4.7 Figures 152

Chapter 5 Major lipids and risk of intracerebral haemorrhage 175

5.1 Introduction 175
5.2 Study design and methods 176

5.2.1 The nested case-control study 176
5.2.2 Statistical analyses 177

5.3 Results 180

5.3.1 Incidence rates of intracerebral haemorrhage in CKB participants 180
5.3.2 Baseline characteristics of participants in the nested case-control study 182
5.3.3 Prospective associations of major lipids with intracerebral haemorrhage 184

5.4 Discussion 188

5.4.1 Incidence rates of intracerebral haemorrhage in CKB participants 188
5.4.2 Prospective associations of major lipids and intracerebral haemorrhage 192
5.4.3 Strengths and limitations 198

5.5 Summary 200
5.6 Tables 201
5.7 Figures 216

Chapter 6 Major lipids and stroke subtypes: Mendelian randomisation studies 237

6.1 Introduction 237
6.2 Study design and study methods 239

6.2.1 Study population and study design 239
6.2.2 Instrumental variables 239
6.2.3 Lipid measurements 243
6.2.4 Stroke ascertainment 244
6.2.6 Statistical analysis 244

6.3 Results 244

6.3.1 Baseline characteristics of participants in the nested case-control study 246
6.3.2 Causal estimates of major lipids with risk of strokes 247

6.4 Discussion 252

6.4.1 Low-density lipoprotein cholesterol 252
6.4.2 High-density lipoprotein cholesterol 256
6.4.3 Triglycerides 260
6.4.4 Strengths and limitations of the genetic studies 262

6.5 Summary 264
6.6 Tables 265
6.7 Figures 273

Chapter 7 Discussion 281

7.1 Summary of main findings 281

7.1.1 Incidence rates of stroke in Chinese adults 281
7.1.2 Prospective associations of major lipids with stroke risks 282
7.1.3 Mendelian randomisation studies of major lipids with stroke risks 284

7.2 Comparison with previous studies 286

7.2.1 Epidemiology of stroke 286
7.2.2 Global trends in total cholesterol 290
7.2.3 Low-density lipoprotein cholesterol 291
7.2.4 High-density lipoprotein cholesterol 300
7.2.5 Triglycerides 304

7.3 Strengths and limitations of the study 307
7.4 Public health and clinical implications 310

7.4.1 Stroke and low-density lipoprotein cholesterol 310
7.4.2 Stroke and high-density lipoprotein cholesterol 311
7.4.3 Stroke and triglycerides 311
7.4.4 Stroke pathological types and major blood lipids 312
7.4.5 Stroke prevention in China and worldwide 313

7.5 Future research 315
7.6 Conclusions 317
7.7 Tables 319
7.8 Figures 322

References 333

List of Tables

Table 1.1	Stroke deaths (age >30 years) in China and all the other countries	10
Table 1.2	A glossary of the genetic terms used in the thesis	24
Table 1.3	A list of terms in the thesis relating to Mendelian randomisation study	26
Table 2.1	Literature review search terms	53
Table 2.2	Characteristics of the meta-analyses examining the associations of major blood lipids with stroke, by pathological type	56
Table 2.3	Characteristics of the prospective studies examining the associations of major blood lipids with ischaemic stroke	57
Table 2.4	Characteristics of the prospective studies examining the associations of major blood lipids with haemorrhagic stroke	58
Table 2.5	Key findings of the meta-analyses examining the associations of major blood lipids with stroke, by pathological type	59
Table 2.6	Key findings of the prospective studies examining the associations of major blood lipids with ischaemic stroke	60
Table 2.7	Key findings of the prospective studies examining the associations of major blood lipids with haemorrhagic stroke	61
Table 3.1	Biochemical methods used for lipid-related biomarkers	78
Table 3.2	Selected demographic and lifestyle characteristics of the CKB study participants at baseline survey (2004-2008), by sex	79
Table 3.3	Selected anthropometric measurements of the CKB study participants at baseline survey (2004-2008), by sex	81
Table 3.4	Comparisons of the selected baseline characteristics of the resurvey participants with those of the overall cohort population	82
Table 4.1	Comparison of the disease status censored in 2014 and in 2016	132
Table 4.2	Exclusion criteria of lipid measurements in the present analyses	133
Table 4.3	Number and crude incidence rate of ischaemic stroke, by sex and area (among 489 762 participants)	134
Table 4.4	Selected baseline characteristics of the nested case-control study participants	135
Table 4.5	Selected baseline physical measurements of the nested case-control study participants	136
Table 4.6	Distribution of lipid-related biomarkers of the baseline samples among the nested case-control study participants	138
Table 4.7	Cross-sectional correlations of major blood lipids with blood pressure, adiposity, and kidney function, among the nested case-control study participants (n=16 541)	139

Table 4.8.1	Mean (SD) baseline values of the lipid biomarkers among the nested case-control study participants (n=11 765)	140
Table 4.8.2	Selected baseline characteristics of the nested case-control study participants by levels of total cholesterol (n=11 765)	141
Table 4.8.3	Selected baseline characteristics of the nested case-control study participants by levels of LDL cholesterol (n=11 765)	142
Table 4.8.4	Selected baseline characteristics of the nested case-control study participants by levels of HDL cholesterol (n=11 765)	143
Table 4.8.5	Selected baseline characteristics of the nested case-control study participants by levels of triglycerides (n=11 765)	144
Table 4.8.6	Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein A1 (n=11 765)	145
Table 4.8.7	Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein B100 (n=11 765)	146
Table 4.8.8	Selected baseline characteristics of the nested case-control study participants by levels of non-HDL cholesterol	147
Table 4.8.9	Selected baseline characteristics of the nested case-control study participants by levels of remnant cholesterol (n=11 765)	148
Table 4.9	Non-parametric and parametric estimates of regression dilution for lipoprotein-related biomarkers in the China Kadoorie Biobank study	149
Table 4.10	Comparison of hazard ratios (HR) of major lipid-related biomarkers with ischaemic stroke, before and after adjustment for body mass index (BMI)	150
Table 4.11	Discordance analyses of LDL cholesterol and Apolipoprotein B100 in Women's Health Study (WHS) and China Kadoorie Biobank study (CKB)	151
Table 5.1	Number and crude incidence rate of intracerebral haemorrhage, by sex and area (among 489 762 participants)	201
Table 5.2	Selected baseline characteristics of the nested case-control study participants	202
Table 5.3	Selected baseline physical measurements of the nested case-control study participants	203
Table 5.4	Distribution of lipoprotein-related biomarkers of baseline samples among the nested case-control study participants	204
Table 5.5.1	Mean values of the lipoprotein-related biomarkers among the nested case-control study participants (n=11 066)	205
Table 5.5.2	Selected baseline characteristics of the nested case-control study participants by levels of total cholesterol (n=11 066)	206
Table 5.5.3	Selected baseline characteristics of the nested case-control study participants by levels of LDL cholesterol (n=11 066)	207

Table 5.5.4	Selected baseline characteristics of the nested case-control study participants by levels of HDL cholesterol (n=11 066)	208
Table 5.5.5	Selected baseline characteristics of the nested case-control study participants by levels of triglycerides (n=11 066)	209
Table 5.5.6	Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein A1 (n=11 066)	210
Table 5.5.7	Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein B100 (n=11 066)	211
Table 5.5.8	Selected baseline characteristics of the nested case-control study participants by levels of non-HDL cholesterol (n=11 066)	212
Table 5.5.9	Selected baseline characteristics of the nested case-control study participants by levels of remnant cholesterol (n=11 066)	213
Table 5.6	Comparison of hazards ratios of major lipoprotein-related biomarkers with intracerebral haemorrhage, with and without adjustment for body mass index (BMI)	214
Table 6.1	Basic information on loci primarily associated with LDL cholesterol selected in the present China Kadoorie Biobank study	265
Table 6.2	Linkage disequilibrium statistics of the genetic markers for LDL cholesterol, within the same chromosome	266
Table 6.3	Linkage disequilibrium matrix (R^2) of the *CETP* genetic variants	267
Table 6.4	Basic information on loci primarily associated with triglycerides selected in the present China Kadoorie Biobank study	268
Table 6.5	Selected baseline characteristics of the study population (n=20 220)	269
Table 6.6	Associations of genetic variants for LDL cholesterol with ischaemic stroke (IS) and intracerebral haemorrhage (ICH) in the CKB study	270
Table 6.7	Associations of common *CETP* haplotypes with major blood lipids, blood pressure, and adiposity measurements	271
Table 6.8	Associations of common *CETP* haplotypes with HDL cholesterol, ischaemic stroke (IS), and intracerebral haemorrhage (ICH)	272
Table 7.1	Comparison of associations of major blood lipids with ischaemic stroke and intracerebral haemorrhage	319
Table 7.2	Age-specific rates of ischaemic stroke and haemorrhagic stroke in the China Kadoorie Biobank study, by presence or absence of cardiovascular disease at baseline	320
Table 7.3	Age-specific rates of ischaemic stroke and haemorrhagic stroke in the China Kadoorie Biobank study, by baseline hypertensive status	321

List of Figures

Figure 1.1	Comparison of randomised controlled trials and Mendelian randomisation studies	18
Figure 2.1	A flowchart of literature search and the identification of studies relating to the research question	54
Figure 2.2	Venn Diagram for three main individual participant data based meta-analyses included in the thesis	55
Figure 3.1	Locations of the CKB study sites and number of individuals recruited at the baseline survey (2004-2008)	66
Figure 4.1	Mean values of major lipoprotein-related biomarker by age at risk	152
Figure 4.2	Mean values of major lipoprotein-related biomarker by age at risk and sex	153
Figure 4.3	Serial shrinkages of ranges for major lipids and apolipoproteins among individuals with paired measurements	154
Figure 4.4	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol	155
Figure 4.5	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol, progressively adjusted for potential confounding factors	156
Figure 4.6	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol in population subgroups of baseline vascular risk factors (I)	157
Figure 4.7	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol in population subgroups of baseline vascular risk factors (II)	158
Figure 4.8	Adjusted hazards ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol, by region	159
Figure 4.9	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein B100 (Apo B100)	160
Figure 4.10	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein B100 (Apo B100), progressively adjusted for potential confounding factors	161
Figure 4.11	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol	162
Figure 4.12	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol, progressively adjusted for potential confounding factors	163
Figure 4.13	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol in population subgroups of baseline vascular risk factors (I)	164

Figure 4.14	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol in population subgroups of baseline vascular risk factors (II)	165
Figure 4.15	Adjusted hazards ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol, by study region	166
Figure 4.16	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein A1 (Apo A1)	167
Figure 4.17	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein A1 (Apo A1), progressively adjusted for potential confounding factors	168
Figure 4.18	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides	169
Figure 4.19	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides, progressively adjusted for potential confounding factors	170
Figure 4.20	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides in population subgroups of baseline vascular risk factors (I)	171
Figure 4.21	Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides in population subgroups of baseline vascular risk factors (II)	172
Figure 4.22	Adjusted hazards ratios (HRs) for risk of ischaemic stroke by usual triglycerides, by study region	173
Figure 4.23	Comparison of associations of baseline lipid-related biomarkers with risk of ischaemic stroke, with adjustment for baseline or usual systolic blood pressure (SBP)	174
Figure 5.1	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol	216
Figure 5.2	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol, progressively adjusted for potential confounding factors	217
Figure 5.3	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol in population subgroups of baseline vascular risk factors (I)	218
Figure 5.4	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol in population subgroups of baseline vascular risk factors (II)	219
Figure 5.5	Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol, by region	220
Figure 5.6	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein B100 (Apo B100)	221

Figure 5.7	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein B100 (Apo B100), progressively adjusted for potential confounding factors	222
Figure 5.8	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol	223
Figure 5.9	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol, progressively adjusted for potential confounding factors	224
Figure 5.10	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol in population subgroups of baseline vascular risk factors (I)	225
Figure 5.11	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol in population subgroups of baseline vascular risk factors (II)	226
Figure 5.12	Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol, by region	227
Figure 5.13	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein A1 (Apo A1)	228
Figure 5.14	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein A1 (Apo A1), progressively adjusted for potential confounding factors	229
Figure 5.15	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides	230
Figure 5.16	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides, progressively adjusted for potential confounding factors	231
Figure 5.17	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides in population subgroups of baseline vascular risk factors (I)	232
Figure 5.18	Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides in population subgroups of baseline vascular risk factors (II)	233
Figure 5.19	Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides, by region	234
Figure 5.20	Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by (A) total cholesterol, (B) non-HDL cholesterol, and (C) remnant cholesterol	235
Figure 5.21	Comparison of associations of baseline lipid-related biomarkers with risk of intracerebral haemorrhage, with adjustment for baseline or usual systolic blood pressure (SBP)	236
Figure 6.1	Diagrammatic representation of the *CETP* region, showing the relative position of informative variants, based on Chinese population (n=208) in 1000 Genome Project (Phase 3)	273

Figure 6.2	Graphical display of the beta estimates in ratio method	274
Figure 6.3	Comparison of effect sizes of LDL cholesterol related genetic variants on LDL cholesterol concentrations estimated in GLGC and CKB studies	275
Figure 6.4	Associations of the genetic risk scores (GRSs) for LDL cholesterol with major blood lipids, systolic blood pressure (SBP), and body mass index (BMI)	276
Figure 6.5	Comparison of associations of LDL cholesterol with (A) Ischaemic stroke, and (B) Intracerebral haemorrhage, in observational study, and genetic study	277
Figure 6.6	Comparison of associations of LDL cholesterol with Ischaemic stroke, by levels of baseline systolic blood pressure, in observational study, and genetic study	278
Figure 6.7	Comparison of associations of LDL cholesterol with intracerebral haemorrhage, by levels of baseline systolic blood pressure, in observational study, and genetic study	279
Figure 6.8	Comparison of associations of LDL cholesterol with (A) lacunar stroke, and (B) non-lacunar stroke, in observational study, and genetic study	280
Figure 6.9	Comparison of effect sizes of genetic variants for triglycerides on plasma concentrations of triglycerides, and HDL cholesterol, estimated in GLGC and the CKB studies	281
Figure 6.10	Associations of the genetic risk scores (GRSs) for triglycerides with major blood lipid fractions, systolic blood pressure (SBP), and body mass index (BMI)	282
Figure 7.1	Incidence rates of ischaemic stroke and intracerebral haemorrhage in the China Kadoorie Biobank participants (n=489 762)	322
Figure 7.2	Observational associations of usual LDL cholesterol with risk of (A) ischaemic stroke and (B) intracerebral haemorrhage	323
Figure 7.3	Observational associations of usual LDL cholesterol with risk of (A) ischaemic stroke (B) intracerebral haemorrhage, in population subgroups of other risk factors	324
Figure 7.4	Observational associations of usual triglycerides with risk of (A) ischaemic stroke and (B) intracerebral haemorrhage	325
Figure 7.5	Observational associations of usual HDL-C with risk of (A) ischaemic stroke, and (B) intracerebral haemorrhage	326
Figure 7.6	Mortality rates at ages 35-70, by stroke pathological type, between 1990-2015, for China and the UK	327
Figure 7.7	Incidence rates at ages 35-70, by stroke pathological type, between 1990-2015, for China and the UK	328

Figure 7.8	Effects of lowering LDL cholesterol with statin therapy on risk of stroke pathological types in the Cholesterol Treatment Trialists' Collaboration, by years of follow-up	329
Figure 7.9	Effects of lowering LDL cholesterol on ischaemic stroke and haemorrhagic stroke from meta-analysis of the CTT, SPARCL, and FOURIER trials	330
Figure 7.10	Comparison of the associations of LDL cholesterol with (A) ischaemic stroke and (B) intracerebral haemorrhage, from randomised trials, observational studies, and genetic studies	331
Figure 7.11	Predicted absolute effects on ischaemic stroke (avoidance) and haemorrhagic stroke (excess) by lowering LDL cholesterol with statin therapy (per 10 000 patients treated for 5 years) by different characteristics in the Chinese population	332

List of Abbreviations and Acronyms

AMORIS	Apolipoprotein-related MOrtality RISk study
APCSC	Asia Pacific Cohort Studies Collaboration
ARIC	The Atherosclerosis Risk in Communities Study
BMI	Body mass index
CCHS	Copenhagen City Heart Study
CCS	Causative Classification System
CDC	Centres for Disease Control and Prevention
CETP	Cholesteryl ester transfer protein
CGPS	Copenhagen General Population Study
CHD	Coronary heart disease
CHS	Cardiovascular Health Study
CI	Confidence interval
CKB	China Kadoorie Biobank
CT	Computed tomography
CTSU	Clinical Trial Service Unit and Epidemiological Studies Unit
CTT	Cholesterol Treatment Trialists' collaboration
CVD	Cardiovascular disease
DALY	Disability-adjusted life-years
ERFC	Emerging Risk Factors Collaboration
FM	Family hypercholesterolemia
GGT	Gamma-glutamyl transferase
GLGC	Global Lipids Genetic Consortium
GRS	Genetic risk score
GWAS	Genome-wide association studies
HDL	High-density lipoprotein
HDL-C	High-density lipoprotein cholesterol

HIC	High-income countries
HMG-CoA	3-hydroxy-3-methyl-glutaryl-coenzyme A reductase
HR	Hazards ratio
HS	Haemorrhagic stroke
HWE	Hardy-Weinberg equilibrium
ICD	International Classification of Diseases
ICH	Intracerebral haemorrhage
IDL	Intermediate-density lipoprotein
IHD	Ischaemic heart disease
IPD	Individual participant data
IS	Ischaemic stroke
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDL-C	Low-density lipoprotein cholesterol
Log_e	The natural logarithm
LMICs	Low-income and middle-income countries
LPL	Lipoprotein lipase
MAF	Minor allele frequency
MET	Metabolic equivalent of task
MI	Myocardial infarction
MONICA	Multinational MONItoring of trends and determinants in CArdiovascular disease project
MR	Mendelian randomisation
MRFIT	Multiple Risk Factor Interventional Trial
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
OCSP	Oxfordshire Community Stroke Project
OR	Odds ratio

PSC	Prospective Studies Collaboration
RBG	Random blood glucose
RCT	Randomised controlled trials
RDR	Regression dilution ratios
RR	Relative risk
SAH	Subarachnoid haemorrhage
SBP	Systolic blood pressure
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
TC	Total cholesterol
TIA	Transient ischaemic attack
TOAST	Trial of Org 10172 in Acute Stroke Treatment
TS	Total stroke
VLDL	Very low-density lipoprotein
WHO	World Health Organisation

Chapter 1 Introduction

Stroke is a clinical syndrome, including ischaemic stroke (IS), intracerebral haemorrhage (ICH), and subarachnoid haemorrhage (SAH), and both IS and ICH are further divided into subtypes with different underlying pathophysiology. Stroke is the second leading cause of death and the third leading cause of adult disability, worldwide. In 2015, stroke accounted for 6.3 million deaths, and 118 million disability-adjusted life-years (DALYs).[1-4] Over the last two decades, despite stable incidence rates and decreased mortality rates, the absolute number of stroke cases has increased substantially, and the estimated number of stroke deaths is set to increase to over 8.6 million by 2030.[3, 5, 6] The increasing global impact of stroke in the decades ahead is projected to be greater in developing counties, including China, which accounts for one-fifth of the world's population, and where stroke is already the leading cause of death and adult disability.[7] Furthermore, compared with white populations of European ancestry, studies of Chinese population have reported higher incidence rates of total stroke, higher proportions of ICH,[8-10] and probably, a different distribution of IS subtypes.[11]

Stroke is largely preventable, and approximately 80-90% of the risk of stroke has been attributed to 10 modifiable risk factors.[12] However, the relevance of major blood lipids, including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and triglycerides, for stroke pathological types remains uncertain. Previous observational epidemiological studies have reported positive associations of LDL-C with IS (albeit substantially weaker than those for ischaemic heart disease [IHD]), and inverse association with haemorrhagic stroke (HS).[13, 14] Randomised controlled trials (RCTs) of lowering LDL-C with statin

treatment have demonstrated similar proportional reductions in risks of IS and IHD by lowering LDL-C using statin treatment, with also, probably, a slight increase in risk of HS.[15-18] The substantially weaker associations of LDL-C with IS than IHD in observational studies differ from the comparable results of lowering LDL-C for IS and IHD in large-scale reliable RCTs. The precise reasons for the discrepancy between the observational evidence and RCTs are not fully understood, and concerns have been raised about the inverse associations of LDL-C with risk of HS, especially among Asian populations where the underlying HS rates are high. There are even greater uncertainties about the associations of HDL-C and of triglycerides with stroke pathological types. Furthermore, most of the previous studies addressing these questions have been conducted in Western populations, and substantial uncertainties remain about the relevance of major blood lipids for stroke risks in other populations. To address these questions, we conducted a nested case-control study of stroke pathological types in the China Kadoorie Biobank prospective study. Therefore, the aims of this thesis were: (**i**) to investigate the shape and strength of associations of major blood lipids (i.e., LDL-C, HDL-C, and triglycerides) with risk of two main stroke pathological types (i.e., IS, and ICH), before and after adjusting for various vascular risk factors, and correcting for regression dilution bias; and (**ii**) to examine the causal relevance of major blood lipids for stroke risks using a Mendelian randomisation (MR) approach.

This chapter provides a definition of stroke and the main stroke pathological types, and provides an overview of stroke epidemiology, together with an introduction to major blood lipids and lipoproteins, and their relevance for stroke pathological types. A detailed literature search of previous studies of the associations of major

blood lipids with stroke pathological types is provided in **Chapter 2**. An overview of the thesis is also included at the end of this chapter.

1.1 Stroke

Stroke is caused by an acute focal injury to the central nervous system due to a vascular cause, and includes IS and HS (ICH/SAH).[19] Despite the stable incidence rates and decreased mortality rates worldwide, the last two decades have witnessed a substantial increase in the absolute number of incident stroke cases, deaths attributable to stroke, prevalent stroke survivors, and DALYs lost due to stroke, with the majority of the stroke burden concentrated in developing countries.[1-3]

1.1.1 Definition

Clinically, stroke is defined as a focal neurological deficit caused by a local disturbance of the cerebral circulation, due either to an obstruction of cerebral blood flow or a rupture of a blood vessel supplying the brain. The World Health Organisation (WHO) definition of stroke is 'a clinical syndrome consisting of rapidly developing clinical signs of focal (at times global) disturbance of cerebral function, lasting more than 24 hours (unless interrupted by surgery or death) with no apparent cause other than that of vascular origin'.[19-21] The WHO definition of stroke, which focuses on the duration of symptoms and signs, and does not require corroboration by computed tomography (CT) or magnetic resonance imaging (MRI), has been widely used in epidemiological studies and clinical practice worldwide.[6, 22] The 24 hour threshold was defined arbitrarily to increase the diagnostic accuracy of stroke and distinguish stroke from transient ischaemic attack (TIA), as conditions of primary vascular origin that are likely to result in

permanent neurological injury.[23] Clinical diagnosis with assistance from imaging findings, using either CT scan or MRI, or autopsy evidence, have now been commonly applied to distinguish IS from HS, and to exclude other conditions (such as hypoglycaemia, epilepsy, multiple sclerosis, and intracranial tumours) that mimic stroke.[24, 25]

1.1.2 Stroke pathological types and subtypes

Stroke is a clinical syndrome comprising a heterogeneous group of diseases. Broadly, stroke can be divided into IS, resulting from occlusion or stenosis of the arteries supplying the brain, or HS, characterised by rupture of the arteries supplying the brain. The two main stroke pathological types can be further divided, based either on the underlying pathology of causative lesions, or sites within the brain. Several classification systems for IS have been proposed, including the Trial of Org 10172 in Acute Stroke Treatment (TOAST: large-artery atherosclerosis, cardio-embolism, small-vessel occlusion, stroke of other determined aetiology, and stroke of undetermined aetiology),[26] the Oxfordshire Community Stroke Project (OCSP) classification scheme (lacunar infarct, total anterior circulation infarct, partial anterior circulation infarct, and posterior circulation infarct),[27] the ASCOD phenotyping system (A: atherosclerosis; S: small-vessel disease; C: cardiac pathology; O: other causes),[28] and the Causative Classification System (CCS).[29] Of these, the TOAST aetiological and the OCSP anatomical classification systems have been most widely used in both epidemiological studies and clinical practice. Likewise, HS can be further subdivided, according to its anatomical site or presumed aetiology, into ICH, characterised by arterial bleeding directly into the

intra-parenchymal region of the brain, or SAH, in which bleeding occurs into the subarachnoid space surrounding the brain.

1.1.3 Diagnosis

Reliable diagnosis of stroke pathological types is important because some treatments for IS may not be recommended for HS.[30, 31] The diagnosis of stroke pathological types requires clinical criteria in conjunction with appropriate autopsy or imaging (CT/MRI) findings. Typical clinical symptoms of stroke include sudden paralysis or weakness, numbness or tingling, visual loss, diplopia, and non-orthostatic vertigo. The diagnosis of stroke is confirmed by CT/MRI for detection of lesion(s) from ischaemia or haemorrhage, and to refute conditions that mimic stroke, whilst SAH is typically diagnosed by clinical findings that are confirmed by lumbar puncture or cerebrospinal fluid examination for presence of blood in the cerebrospinal fluid.[32, 33] CT scanning is feasible and readily used in all patients, and non-contrast head CT has near-perfect sensitivity to detect fresh ICH, but is relatively poor for detection of IS if the ischaemia is recent-onset, small, or is located in the posterior fossa.[34] Diffusion weighted MRI detects acute ischaemia in approximately 90% of cases with IS, and about 30% of those with TIA.[28, 35] In addition, haemorrhagic transformation of cerebral ischaemia may be difficult to detect through CT/MRI alone, where information on the medical history is necessary to interpret the results of imaging. Plasma biomarkers have not been widely used for the diagnosis of cerebral infarction,[36] although troponin and creatine kinase concentrations are widely used to diagnose and quantify the severity of myocardial infarction (MI). Accurate diagnosis of stroke is essential for appropriate clinical management, and also for accurate ascertainment of the

distribution and determinants of stroke pathological types in epidemiological studies.

1.1.4 Pathophysiology

Overall, 75-80% of all strokes are due to IS.[37] Typically, IS results from thrombosis, due to atheroma of the major arteries in brain or neck, or from embolism, where a thrombus detaches from a lesion in another part of vascular system and results in occlusion elsewhere in the blood supplying the brain. The pathophysiology of IS (or some IS subtypes) is similar to that of IHD, with both sharing a similar atherosclerotic process as one of the main underlying pathological mechanisms.

By contrast, HS (ICH and SAH) accounts for 10-20% of all strokes (varying in populations), with slightly higher proportions in Asian than in Western populations.[10] ICH results from rupture of weakened cerebral blood vessels, due chiefly to hypertension or cerebral amyloid angiopathy that commonly correlates with ageing. SAH typically follows rupture of one or more aneurysms (generally of unknown aetiology, but also results from hypertension) or rupture of a vascular malformation.

1.2 Burden of stroke

1.2.1 Stroke burden worldwide

Worldwide, stroke is the second leading cause of death, and the third leading cause of disability.[1, 3] In 2015, there were 16.9 million incident stroke cases, 6.3 million stroke deaths, 118 million DALYs lost due to stroke worldwide.[1, 3] The global burden of stroke has increased substantially over the last two decades, including a 67% increase in new-onset stroke cases, 84% increase in stroke

survivors, 26% increase of stroke-related deaths, and 19% increase in DALYs lost due to stroke,[6] with most of such increases in stroke occurring in low- and middle-income countries (LMICs). Population growth and ageing, and in LMICs, an increase in the prevalence of stroke risk factors following the epidemiological transition has accounted for most of the increase in the burden of stroke.

Between 1990 and 2013, the overall mortality rates for stroke have declined from 141.6 per 100 000 per year to 110.1 per 100 000 per year, after standardisation to the WHO standard population,[38] with a somewhat steeper decline in mortality in developed than in developing countries. By contrast, over the last two decades, the overall age-standardised stroke incidence has remained unchanged (250.6 in 1990 versus 258.0 in 2010).[6] These global patterns and trends in stroke incidence rates have masked a diverging trend between populations, with a significant 12% (6-17%) decrease in high-income countries (HICs), and albeit non-significant, a 12% (-3 to 22%) increase in LMICs.[6, 39] Moreover, in the first decade of the 21st century, the overall stroke incidence rates in LMICs, for the first time, have exceeded those in HICs.[39] Furthermore, stroke affects individuals at a relatively younger age in LMICs than in HICs (a mean difference of 3-5 years), despite in both settings, an increase in the mean age at first-ever stroke over the last two decades.[5, 6, 40]

Both stroke incidence and stroke mortality rates vary substantially between countries, with an approximately ten-fold difference between geographical regions.[6, 41, 42] In 2015, the stroke incidence rates ranged from 38 (per 100 000 person-years) in Canada to 315 in Lithuania, and stroke mortality rates ranged from 29 in Canada to 348 in Afghanistan.[4] Comparisons of the patterns and trends in stroke incidence and mortality between countries, suggests that most of the

differences in stroke mortality are largely accounted for by changes in incidence, although differences in treatment and case severity could also contribute to some of the decline in stroke mortality.[5, 6, 41] Variation in stroke incidence and mortality between geographical regions could also reflect differences in exposure to environmental risk factors, genetic factors, and differences in stroke management practices, or differences in the methodology used for reporting such statistics.[43, 44]

Stroke pathological types

Investigation of the epidemiology of stroke pathological types could help prioritise region-specific stroke prevention strategies and also contribute to knowledge of the health consequences arising from the epidemiological transition in such populations. Consistent with the patterns for total stroke, the burden of both stroke pathological types has increased substantially, particularly in developing countries. While IS is the chief pathological type of stroke in developed countries, HS accounts for a much higher proportion of stroke deaths and greater burden of disability worldwide.[1, 3, 10, 45]

Ischaemic stroke

In 2015, worldwide, there were 11.5 million incident IS events, 3.3 million deaths from IS, and 65.5 million DALYs lost due to IS, of which, 63%, 57%, and 64% respectively, occurred in LMICs. From 1990 to 2010, there was a significant increase in the absolute number of incident IS (37%), death from IS (21%), and DALYs lost due to IS (18%). In 2015, the region with the highest median country-specific mortality was Eastern Europe (96 per 100 000 person-years), followed by the Oceanic islands (63) and Central Europe (62). The lowest mortality rates were

reported for high-income North America (19), Australasia (19), and Andean Latin America (21).[4, 5, 46]

In HICs, between 1990 and 2010, the rates of IS incidence declined by 13%, mortality declined by 37%, and DALYs declined by 34%, respectively. In contrast, in LMICs, there was a modest 6% increase in the age-standardised incidence of IS, including East Asia, despite modest reductions in mortality and DALYs, reflecting perhaps the epidemiological transition in LMICs.[47, 48] However, it has not been feasible to obtain reliable estimates of the temporal trends of stroke pathological types in most LMICs, since brain imaging has not been widely available for stroke diagnosis in these countries.

Haemorrhagic stroke

HS, including ICH and SAH, accounted for approximately 10-25% of all strokes worldwide. In 2015, there were around 3.5 million incident HS cases, over 3.0 million deaths, and 62.8 million DALYs lost due to HS, of which, 80% of the incident cases and 63% of the deaths, occurred in LMIC. Likewise, there has been substantial geographical variation in the incidence of HS between countries. The incidence of HS was highest in Central and East Asia (101-158 per 100 000 person-years), and East and Southern sub-Sahara Africa (73-101), while high-income North America, Western Europe, and Oceania (25-40) had the lowest rates of HS.[4]

Since 1990, in HIC, age-standardised rates of HS incidence, mortality, and DALYs lost have declined, by 8%, 38%, and 39%, respectively. In contrast, in LMICs, there was a significant 22% increase in HS incidence, despite the reductions in mortality and disability related to HS. Partly because of this, but chiefly due to

population growth and ageing, the absolute number of incident HS cases has doubled in LMICs (from 2.0 million to 4.2 million).[5]

1.2.2 Stroke burden in China

Stroke is the leading cause of death in China,[3, 7] and in 2010, stroke accounted for over 1.9 million deaths (approximately one-third of overall stroke deaths worldwide; **Table 1.1**), and 25 million DALYs lost.[7] The epidemiology of stroke in China is characterised by a higher overall stroke incidence, and a higher proportion of ICH, compared to Caucasian populations.[10, 42, 49-51] ICH was reported to account for approximately 30% of all stroke cases in China, compared with around 10-15% in Western populations,[10, 52] and for IS, studies have reported a higher proportion of small vessel disease in Chinese compared to Western populations,[11] although the evidence on this has been inconclusive.[10]

Table 1.1: Stroke deaths (age >30 years) in China and all the other countries *

	Ischaemic stroke	Haemorrhagic stroke
China	1 M	1 M
All other countries	2 M	2 M

M=Million; *Numbers in the table were rounding to the nearest million; Source: Global Burden of Disease Study 2015.[4]

In China, a few well-designed studies have examined the differences in stroke incidence in different parts of China.[42, 49, 51, 53, 54] The most reliable estimates of regional differences in stroke incidence in China are provided by the WHO Multinational MONItoring of trends and determinants in CArdiovascular disease project (MONICA), which was initiated in 1981, to investigate trends in the major determinants of cardiovascular diseases in a selection of populations, including China.[22] As part of the project, the Sino-MONICA project recorded stroke incidence in Beijing between 1985 and 1987, and expanded to 17 regions in 1987 across China.[53] Substantial geographical variation in overall stroke incidence rates

has been recorded, with ten-fold differences between regions in China. The incidence rates of stroke at age 35-64 years ranged from 74 to 751 per 100 000 person-years for men, and 41 to 406 per 100 000 person-years for women. In both sexes, the incidence rates tended to be higher in Northern than in Southern regions.[42, 51, 55, 56] From 1980s to 2000s, in China, there has been an increase in overall stroke incidence, consisting of an increase in IS and a decrease in HS.[57] In 2013, a national survey of around 480 000 Chinese adults reported that the crude incidence rates of stroke were around 345 per 100 000 person-years, with incidence rates of 240 and 82 for IS, and ICH, respectively.[58] Despite the large sample size, the population involved in this survey (2013-2014) was relatively young (mean age: 46.6), and was unable to exclude the recurrent from the first-ever-in-a-lifetime stroke events.

1.2.3 Risk factors for stroke

Given the differences in the underlying pathophysiology of stroke, major risk factor profiles of stroke pathological types also differ. Risk factors for IS are similar to those for IHD, differing somewhat in strength of associations, including advancing age,[59] elevated blood pressure, smoking, diabetes, adiposity,[60-62] lack of physical activity,[63] atrial fibrillation,[64, 65] and carotid artery atherosclerosis.[66] Risk factors for HS include elevated blood pressure,[67] advancing age,[59] smoking,[68] heavy alcohol consumption,[69] and coagulation disturbance.[70] Several of these risk factors for stroke are modifiable or treatable, providing scope for both primary and secondary prevention. The international study of risk factors for stroke – the INTERSTROKE study – suggested that approximately 90% of the stroke burden could be explained by a few modifiable risk factors.[12, 40] In addition, multiple genetic

variants have been identified demonstrating associations with increased stroke risk, but the effect sizes are modest for most of such genetic variants.[71-73]

1.3 Lipids metabolism and homeostasis

1.3.1 Lipoprotein metabolism

Cholesterol serves as a structural component of cell membranes, and as a precursor of the bile salts and steroid hormones. Cholesterol is obtained from the diet or is synthesized by most cells in the body, but chiefly in cells in the liver and intestine. HMG-CoA reductase is the major rate-limiting enzyme involved in cholesterol synthesis. Competitive inhibitors of the HMG-CoA reductase enzyme (e.g., statins) can inhibit cholesterol synthesis, and cause the upregulation of LDL receptors in the liver, and therefore decrease the plasma concentrations of cholesterol.

Triglycerides are the major energy store, and due to their insolubility, are transported with cholesterol in macromolecular complexes called lipoproteins. A typical lipoprotein particle consists of a core of triglycerides, and cholesteryl esters, with a surface monolayer of phospholipids and free cholesterol. Each lipoprotein particle is associated with one or more protein molecules (i.e., apolipoproteins). The large chylomicrons and VLDLs are collectively referred to as triglyceride-rich lipoproteins, while the smaller LDL and HDL particles typically transport most of the cholesterol in the circulation.

Dietary cholesterol and triglycerides are both absorbed and re-esterified in the intestine, and combine with phospholipids and one molecule of apolipoprotein B48 to form chylomicron particles in the circulation. Meanwhile, cholesterol and

triglycerides are also synthesized in the liver, and packaged together with phospholipids, one molecule of apolipoprotein B100, and a small amount of other apolipoproteins as VLDL. Plasma triglycerides are digested by lipoprotein lipase (LPL), an enzyme found attached to capillary endothelial cells. Fatty acids generated during the lipolysis process are oxidized to produce energy, or taken up by adipose tissue and stored as triglycerides. LPL converts chylomicrons to chylomicron remnants, and VLDLs to intermediate-density lipoproteins (IDLs). These products can be taken up by liver through receptors (i.e., LDL-receptor related protein), or are converted to LDL through further digestion of triglycerides in VLDLs. Such LDL particles are rich in cholesterol and cholesterol esters. It is estimated that approximately 60-70% of cholesterol in the bloodstream is transported in LDL particles, delivering cholesterol to peripheral tissues.[74] Most of the remaining measured cholesterol is transported in HDL particles, of which the principal function is to transport the excess cholesterol from peripheral tissues back to the liver (i.e., the reverse cholesterol transport process), and to exchange proteins and lipids with chylomicrons and VLDLs.

1.3.2 Lipids and atherosclerosis

Atherosclerosis is a chronic disease of the arteries that are characterized by initial deposition of fatty streaks in the major arteries. Fatty streaks in arterial walls gradually develop into atheroma and characteristic atheromatous plaques. Acute rupture of these atheromatous plaques causes local thrombosis, leading to partial or total occlusion of the affected artery. The clinical consequences of atheromatous plaques depend on their site and the degree of arterial occlusion.

Atherosclerosis has a long latency, involving several decades, and frequently coexists in several vascular beds.

Low-density lipoprotein cholesterol

LDL initiates atherosclerosis by accumulation in the sub-intimal space of the arterial wall where the particles are modified by oxidation and aggregation. Oxidized LDL is taken up by macrophages much more rapidly and sufficiently to cause cholesterol accumulation, and is believed to induce inflammation and a cascade of events that culminates in plaque formation. This pathophysiological model is consistent with the epidemiological evidence that higher LDL-C is associated with an increased risk of occlusive vascular diseases in different populations,[13, 75] and provides a biological rationale for strategies to lower LDL-C in RCTs.[76, 77]

High-density lipoprotein cholesterol

In contrast, HDL is the smallest and most protein-rich lipoprotein class, and is involved in transport of excess cholesterol from peripheral tissues and returning it to the liver.[78] HDL particles have several other functions with the potential to protect against atherosclerosis, including inhibition of vascular oxidation, inflammation, and promotion of endothelial repair, etc.[79] The HDL fraction in human plasma is heterogeneous, consisting of several subpopulations of particles varying in size, density, composition, and probably, function.[80] HDL-C is a measure of the overall cholesterol content in these HDL particles, which are now frequently grouped into two general sub-fractions based on density: the larger, less dense HDL_2 particles, and the smaller and denser HDL_3 particles.[81] The protective effect of HDL system may reflect both the metabolic and inflammatory

processes, in addition to its role in reverse cholesterol transport process. This may partly explain the consistently inverse association of HDL-C, as a biomarker of the HDL system, with CHD in large-scale epidemiological studies throughout different populations.[13, 14] Despite previous studies,[79, 80] the relevance of the different HDL subpopulations for protection against atherosclerosis is not fully understood.

Triglycerides

The overall evidence tends to support the cholesterol content in triglyceride-rich lipoproteins, rather than triglyceride per se, that relates to atherosclerosis.[82, 83] Low HDL-C concentrations may be a marker for some defects in the metabolism of the triglyceride-rich lipoproteins. The triglyceride-rich lipoprotein remnants remain for longer in circulation until they become sufficiently small for receptor-mediated uptake. These triglyceride-rich lipoproteins can also become cholesterol enriched through the action of cholesteryl ester transfer protein (CETP), and the cholesterol-rich remnants themselves may be taken up to initiate the formation of atherosclerotic lesions.[84, 85]

1.3.3 Determinants of blood lipids

The principal environmental determinants of blood lipid concentrations include dietary intake of saturated fat, polyunsaturated fat, and dietary cholesterol,[86-89] albeit plasma lipid concentrations are also affected by weight loss,[90] physical activity,[91] genetic variation,[92, 93] and other factors. However, in observational epidemiological studies, the absence of associations of dietary intake of total fats and major types of fatty acids with major blood lipids may not be unexpected. In an ethnic group where the dietary habits are relatively homogenous, the intra-individual variation in dietary fats is approximately of the same order of magnitude

as the inter-individual variation, and thus, the assignment of individuals within groups defined by differences in diet tends to be less reliable.[94] The most reliable evidence for assessing the quantitative relationships between intake of saturated fat and plasma lipid concentrations has been provided by "metabolic ward" studies, and meta-analyses of such studies, which assess effects on major lipids in participants consuming diets that exchange a particular dietary fat by another or by complex carbohydrates under controlled isocaloric conditions.[87]

In addition, genome-wide association studies (GWAS) have identified 157 loci associated with plasma lipid concentrations. Each has only minor effects, and together they account for only modest differences in plasma lipid concentrations (8-17%).[92] **Table 1.2** defines the genetic terms that will be used in this thesis.[95-97] The strongest associations reported in previous GWAS studies have almost exclusively been reported with common variants (minor allele frequency [MAF] >5%) in populations of European ancestry, and much less is known about their effects in other populations. Although the underlying genetic architecture of blood lipids is, for a large part, similar across ethnic groups, population-specific lipid loci have also been identified.[98] In contrast, previously published variants known to cause Mendelian dyslipidaemia (e.g., familial hypercholesterolemia) and, more generally, variants with known functional impact on lipids frequently have a low MAF (<5%), and consequently, their effects on population mean plasma lipid concentrations may be trivial.

1.4 Major lipids and risk of stroke pathological types

Despite the increasing burden of stroke worldwide, the relationship of major blood lipids with stroke pathological types remains unclear. Previous studies including

observational epidemiological studies, genetic epidemiological studies, and RCTs have each provided complementary evidence for the associations of different lipid fractions with risk of either stroke pathological types. However, substantial uncertainty persists about the public health implications and impact on clinical practice of treatment of individual lipid fraction. Thus, further research is needed to investigate the associations of major blood lipids with stroke pathological types.

Conventional observational epidemiological studies have an important role in generating hypotheses, and in providing insights into the risks of disease in particular circumstances that can help to generalise from clinical trials to clinical practice of the effects of particular treatments.[99] But, due to their inherent potential for moderate or large biases, confounding, reverse causality, and measurement error, observational studies cannot draw conclusions on the causal nature of the associations of commonly moderate differences in risks of major vascular disease outcomes. Use of a genetic approach in observational epidemiology – Mendelian randomisation (MR) – affords an alternative approach that overcomes many of the limitations of conventional observational studies.

MR is a term for studies that use genetic variants in epidemiological studies to make causal inferences about modifiable factors for disease and health-related outcomes (**Table 1.3** defines the terms that will be used in this thesis relating to MR studies).[100, 101] Such studies exploit what is known as Mendel's second law that the inheritance of one trait is independent of (i.e., randomised with respect to) the inheritance of other traits. The independent distribution of genetic variants means that a study relating the genetic variants to the health outcome will not suffer from confounding that frequently distorts the interpretation of findings from

observational epidemiology.[102] Furthermore, an individual's genotype is determined at conception and therefore associations between the genotype and the disease outcome are unaffected by reverse causality. Additionally, genetic variants that are related to a modifiable exposure (e.g., LDL-C) will generally be related to it throughout life and therefore their use in causal inference can also avoid attenuation by random measurement errors (regression dilution bias).[100, 103]

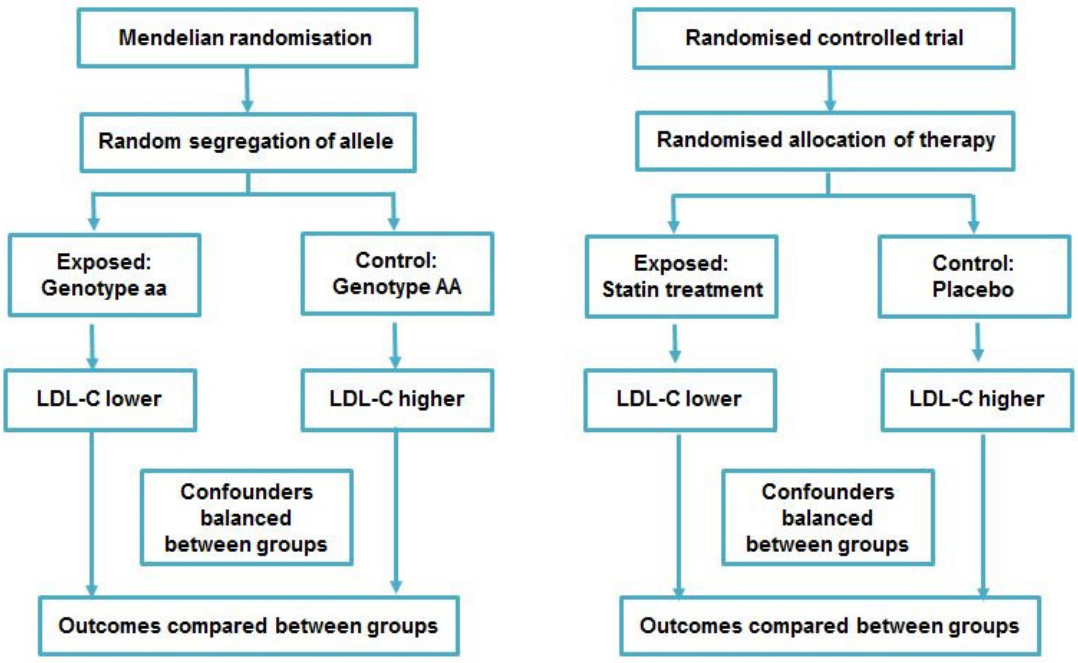


Figure 1.1: Comparison of randomised controlled trials and Mendelian randomisation studies

"aa" and "AA" refer to alleles of *LDLR* gene associated with low and high LDL-C, respectively.[80] For simplicity, only homozygous individuals are shown.

MR studies have been likened to a "natural" RCT,[104] as shown in **Figure 1.1**, in which a difference in disease risk between these subgroups would therefore indicate a causal relationship between the modifiable exposure and disease outcome. The availability of previously identified genetic variants encoding lipids have provided an opportunity to use genetic variants associated with lipid levels,[100, 105] to mimic RCTs using observational data to confirm or refute the

causal relevance of such associations of major blood lipids with stroke pathological types.

However, as with observational studies, the precise interpretation of MR studies relies on the fulfilment of certain assumptions, of which the most important is the selection of valid genetic variants based on biological pathways, and the avoidance of "weak instrument bias" due to the modest effect contributed by each individual or combined genetic variant(s). MR studies are most commonly carried out within general populations, which potentially make them susceptible to distorting factors such as population stratification (**Table 1.3**). Moreover, the fact that the randomisation in MR occurs at conception also means that MR studies may be biased by developmental compensation (**Table 1.3**). Instead, the most reliable evidence for assessing moderate effects on major disease outcomes can only come from sufficiently large-scale RCTs, which requires strict control of bias and random errors, in both the design and analysis of such trials.[18, 106]

1.4.1 Major lipids and risk of ischaemic stroke

Previous epidemiological studies conducted in Western populations have reported that the relevance of LDL-C (or non-HDL-C) for IS was substantially weaker than that for IHD.[13, 14] In contrast, RCTs have demonstrated that lowering LDL-C reduces the proportional risk of IS to a similar extent to that for IHD.[15] Such discrepancies between the results of observational studies and the RCTs have not been fully understood. Furthermore, it is unclear whether the relevance of LDL-C for IS continues down to lower plasma concentrations than those commonly observed in Western populations. Hence, further studies are needed to investigate

the associations of LDL-C with IS in different populations, and possibly, IS subtypes, using both conventional observational and MR study approaches.

In contrast, previous studies assessing the relevance of HDL-C, and of triglycerides for IS have been predominantly conducted in Western populations,[14, 107] and there is a lack of evidence about their relevance in other populations. There is also uncertainty about the relevance of treatment to lower triglycerides concentrations for prevention of cardiovascular disease (CVD) in general.[78, 83] Therefore, understanding the associations of HDL-C and of triglycerides with risk of IS, could help to improve risk prediction, and potentially afford additional therapeutic targets for the prevention of stroke.

1.4.2 Major lipids and risk of haemorrhagic stroke

There is even greater uncertainty about the relevance of major blood lipids to risk of HS. Previous observational epidemiological studies have reported inverse associations of elevated TC with risk of HS, particularly in individuals with high blood pressure.[13, 108] Several meta-analyses involving individual participant data from RCTs have suggested higher risks of HS associated with lowering LDL-C by statin treatment.[15, 18] However, the absolute excess in HS may be outweighed by the beneficial effect on IS in Western populations,[18] but there is a lack of evidence about the balance of benefits and hazards in Asian populations where the underlying rates of HS are higher than in Western populations, or patients with prior cerebrovascular events, or hypertension.[10] Furthermore, available published evidence on the relevance of HDL-C and triglycerides to HS is extremely limited. Hence, appropriately designed studies involving large numbers of well-phenotyped HS cases are needed to address these questions.

1.5 Overview of the thesis

The last two decades has seen an increase in stroke burden worldwide. Despite the fact that the majority of the stroke burden is largely avoidable by improvement in modifiable risk factors, the relevance of major blood lipids to risk of stroke needs more thorough investigation. Although a few large-scale well-designed epidemiological studies have investigated the shape and strength of associations of major blood lipids with stroke pathological types, these have been limited by a relatively small number of confirmed stroke cases, lack of reliable diagnosis of stroke pathological types, or indirect measurement of lipid fractions. Hence, further large-scale well-designed prospective studies with accurate lipid measurements and detailed ascertainment of stroke pathological types are required.

1.5.1 Chapter structure

Chapter 2 presents a systematic literature review of the previously published studies of the relevance of major blood lipids to stroke risks in different types of studies, including observational, MR studies, and RCTs. The aim of the literature review was to identify the evidence gap in previous studies, and to inform research strategy for the present DPhil work. A summary of the aims and objectives of the DPhil thesis is included at the end of the chapter.

Chapter 3 outlines the design and characteristics of the CKB study that are included in this DPhil project. This chapter outlines the design, data collection procedures, and the baseline characteristics of the CKB study population, and resurvey populations, relevant to the subsequent chapters where various exclusions are made.

Chapter 4 examines the shape, strength and independence of the associations of major blood lipids with risk of IS. This chapter describes and estimates the regression dilution ratios of major lipid fractions assayed using the baseline survey, first and second resurveys in the Chinese population. This chapter also presents the cross-sectional correlations of lifestyle factors with lipid-related biomarkers, as well as the prospective associations of major blood lipids with IS, after adjustment for other vascular risk factors, and correction for regression dilution bias.

Chapter 5 assesses the shape, strength and independence of the associations of major blood lipids with ICH, and the relevance of other vascular risk factors, particularly blood pressure, for these associations. Equivalent analyses to those presented in Chapter 4 are conducted in this chapter to allow comparisons of associations of major blood lipids with different stroke pathological types.

Chapter 6 displays the results of MR studies of genetic variants determining plasma concentrations of major lipid fractions with stroke pathological types. This chapter presents the selection of genetic variants and the construction of genetic risk scores (i.e., LDL-C, HDL-C, and triglycerides) as instrumental variables relating major blood lipids to risk of stroke pathological types.

Chapter 7 includes a discussion of the findings of the overall thesis, including a summary of the main findings reported in the thesis, comparisons with previous studies on this topic, and critical commentary on the strengths and limitations of the analyse included in the thesis. This chapter also provides conclusions, and suggests possible directions for future research.

1.5.2 Summary

Despite previous studies, the relevance of major blood lipids for stroke pathological types remains uncertain, and even greater uncertainties exist in Chinese populations that have higher rates of HS, compared to Western populations. This thesis provides important new evidence on the relevance of major blood lipids for risk of two main stroke pathological types. The results should provide evidence for the public health interventions, and guide clinical practice in Chinese populations, but also, will provide insights into the aetiology of stroke and clinical practice in Western populations.

In the following chapter, a detailed literature review of the previous published evidence on these research questions is presented, with the aim of identifying gaps in the evidence and informing the research strategies in the present thesis.

Table 1.2: A glossary of the genetic terms used in the thesis

- Alleles are the variant forms of a single-nucleotide polymorphism (SNP), a specific polymorphic site or a whole gene detectable at a locus. Allele frequency is the relative frequency of an allele at a particular locus in a population, expressed as a fraction or percentage. Minor allele frequency (MAF) refers to the frequency at which the second most common allele occurs in a given population.
- Candidate gene study is conducted with focusing on associations between genetic variations within pre-specified genes of interest and phenotypes or disease states. Candidate genes are most often selected for study based on a priori knowledge of the gene's biological functional impact on the phenotype or disease in question.
- Chromosome carries a collection of genes located on a long string of DNA. A non-homologous chromosome carries a unique collection of genes on a long string of DNA that is different from the gene collection of another non-homologue. Normal non-homologous chromosomes are not attached to each other during meiosis, and move independently of one another, each carrying its own gene collection. Two homologous chromosomes carry the same collection of genes, but each gene can be represented by a different allele on the two homologues (a heterozygous individual). A gamete will receive one of those homologues, but not both. Humans have 22 pairs of autosomal homologous chromosomes and 1 pair of sex chromosomes.
- DNA – deoxyribonucleic acid is a molecule that contains the genetic instructions used in the development and functioning of all living organisms. It contains the instructions needed to construct other components of cells, including proteins and ribonucleic acid (RNA) molecules. DNA has four nucleotide bases A, T, G and C. The two strands of DNA in the double-helix structure are complementary (sense and anti-sense strands) such that A binds with T and G binds with C.
- Gene comprises a DNA sequence, including introns, exons, and regulatory regions, related to transcription of a given RNA.
- Genome-wide association study (GWAS) is a study of common genetic variation across the entire human genome designed to identify genetic associations with observable phenotypes or disease of interest.
- Genotype of an individual refers to the two alleles inherited at a specific locus - if the alleles are the same; the genotype is homozygous, if different, heterozygous.
- Genotype call rate is the proportion of genotypes per marker with non-missing data.
- Haplotype describes the particular combination of alleles from linked loci found on a single chromosome.
- Hardy-Weinberg equilibrium (HWE) is defined by genotype frequencies consistent with the two alleles being independently sampled from a population of alleles.
- Locus is the position in a DNA sequence and can be a SNP, a large region of DNA sequence, or a whole gene.

Table 1.2: A glossary of the genetic terms used in the thesis

- Phenotype refers to measurable characteristics of an individual, other than genotypes. Phenotypes may be binary disease states – the presence or absence of stroke, for example, or quantitative characteristics, such as LDL cholesterol concentrations.
- Polymorphism is the existence of two or more variants (i.e., SNPs, specific polymorphic sites or whole genes) at a locus. Polymorphism is usually restricted to moderately common genetic variants, at least two alleles with frequencies of greater than 1 per cent in the population.
- Recombination is any process that generates new gene or chromosomal combinations not found previously in that cell or its progenitors. During meiosis, recombination is the process that generates haploid cells that have non-parental combinations of genes.
- Single-nucleotide polymorphism(s) (SNPs) are genetic variations in which one base in the DNA is altered (e.g. a T instead of an A).
- Wild type refers to the normal allele, not associated with an increased risk of the phenotype.

Table 1.3: A list of terms in the thesis relating to Mendelian randomisation study

- Association analysis in genetic epidemiology seeks to identify disease or health-related outcome association with measured genetic exposures – the alleles of genotyped genes.
- Canalisation (also known as developmental compensation) is the process by which potentially disruptive influences on normal development from genetic (and environmental) variations are damped or buffered by complementary developmental processes.
- Linkage disequilibrium (LD) is the correlation between allelic states at different loci within the population, usually occurs because the polymorphisms are close together on the chromosome. The term LD describes a state that represents a departure from the hypothetical situation in which all loci exhibit complete independence (linkage equilibrium).
- Pleiotropy is the potential for polymorphisms to have effects on more than one specific phenotype.
- Population stratification is the confounding of the genotype-outcome associations by ethnicity or other population subgroups, which can mimic the signal of association and therefore lead to more false positive results or to missed real effects.
- An instrumental variable (IV) is a variable that is analogous to the treatment allocation process in a randomised controlled trial and is not directly related to outcome, except through the direct effect of treatment, nor to outcome through either measured or unmeasured confounders. In the context of Mendelian randomisation studies, an IV refers to the genetic variants associated with the modifiable factors for a disease or health-related outcome of interest.
- Weak instrument bias occurs when the chance difference in confounders explains more of the variations in exposures between subgroups than the IVs. If the effect of IVs is near zero, then the estimate of the 'causal association' in MR studies approaches the association between exposures and outcome caused by changes in confounders, i.e., the observational confounded association.

Chapter 2 Literature review

The aims of this chapter are to review systematically the previously published studies on the associations of major blood lipids, apolipoproteins with stroke risks, to identify major gaps in the evidence, to inform new research on this topic, and to outline the specific aims and objectives of the thesis. This chapter includes a description of the process of identifying the relevant literature, a critical appraisal and a summary of the published evidence from prospective observational studies, MR studies, and RCT evidence on the relevance of major blood lipids for risk of IS and ICH. The main findings from the previously published studies are summarised in the tables included in this chapter.

2.1 Objectives

The objectives of the literature review were to identify, appraise and summarise the previously published studies on the relevance of major blood lipids (i.e., LDL-C, HDL-C, and triglycerides), apolipoproteins (i.e., apolipoprotein A1 [Apo A1] and apolipoprotein B100 [Apo B100]), for risk of stroke pathological types. The literature review was undertaken also to identify gaps in the current evidence and to inform further research questions to be investigated.

2.2 Methodology

The literature search was undertaken using the PubMed database. Details of the search terms, and screening procedures are provided in **Table 2.1** and **Figure 2.1**. Individual studies that have been included in identified meta-analyses were not reported separately. The search for RCT evidence was undertaken separately, as most of the relevant evidence, including published or unpublished trials, comes

from several reports of comprehensive meta-analyses of all large-scale trials (e.g., Cholesterol Treatment Trialists' Collaboration [CTT]). While the published evidence on HS includes both ICH and SAH, the literature search was not restricted to ICH, as no studies were identified on the associations of major blood lipids with ICH alone, based on the criteria defined at the beginning of the literature search (**Figure 2.1**). The appraisal of published evidence was based on the findings of the large-scale studies included in **Tables 2.2-2.7**. The MR studies, including four published full reports and one solely reported in abstract form, are discussed separately, and are not listed in the tables provided.

2.3 Prospective epidemiological evidence

2.3.1 Ischaemic stroke

Reliable assay methods for direct measurement of LDL-C have only become available in the 1990's. Hence, before this period, total cholesterol (TC) was used as the main indicator of the cholesterol content in atherogenic lipoproteins (e.g., LDL) in previous studies examining associations of major blood lipids with risk of stroke pathological types. Approximately 60-70% of the TC is transported via LDL particles (or Apo B-containing particles, more accurately), whereas the remainder is transported in HDL particles. Studies of the associations of TC or of non-HDL-C (i.e., TC minus HDL-C) with IS are described together with those for LDL-C. Due to "surrogate dilution bias" in observational studies, using TC as an indicator for LDL-C (not LDL-C targeted interventional studies, where the variation of TC is chiefly dependent on the differences in LDL-C) to relate to atherosclerotic disease may have underestimated the relevance of LDL-C for atherosclerotic disease. In

addition, the atherogenic effect of LDL-C may, somehow, be diluted by the probable cardio-protective effect of HDL particles.

Overall, 9 individual participant data (IPD) meta-analyses, and 13 individual prospective studies, were identified that fulfilled the pre-defined criteria on the relevance of major blood lipids or apolipoproteins with IS, in different populations (**Table 2.2**, **2.3**, **2.5**, and **2.6**). Evidence on the prospective associations of major lipids and stroke types were chiefly from three large IPD meta-analysis studies, i.e., the Prospective Study Collaboration (PSC) study,[13] the Emerging Risk Factor Collaboration (ERFC) Study[14] and the Asia Pacific Cohort Studies Collaboration (APCSC) study.[109] **Figure 2.2** shows the overlapping of the studies among the three main IPD meta-analyses.

The totality of evidence indicated a continuous (log-linear) positive association of non-HDL-C (or LDL-C, and TC), and of triglycerides with IS, and an inverse association of HDL-C with IS. Substantial discrepancies existed in the strength of the associations reported in individual studies, perhaps reflecting inadequate statistical power to characterise associations across the full range of lipid values.

2.3.1.1 Low-density lipoprotein cholesterol

Three IPD meta-analyses of prospective studies, of plasma cholesterol concentrations and stroke have been conducted in different populations. The Prospective Studies Collaboration (PSC) included 61 prospective studies recruited chiefly from Western populations (with approximately 1.8% were enrolled from Chinese populations), involving approximately 900 000 individuals with no prior history of vascular disease and with measurements of TC at baseline. Over 12 000 stroke deaths were recorded during a mean follow-up period of 13 years.

Data on stroke pathological types were available for about half of all the stroke deaths, among which, less than half were attributable to IS (n=2134). Despite the variation in assay methods for TC in each of the individual studies, the mean TC concentrations tended to be lower in Asian (4.6 mmol/L) than in Western populations (5.5 mmol/L in North America and 6.0 mmol/L in Europe). Overall, in the PSC study, after adjustment for age, sex, and study, compared with the bottom group (mean of usual TC: 4.08 mmol/L), the higher risks of IS were restricted to individuals with usual TC of 6 mmol/L or greater (HR 1.14; 95% CI: 1.05-1.23), below which the shape of associations appeared to be flat. After stratification by age-at-risk, there was a weak positive association of usual TC with mortality from IS among individuals in early and late middle age (40-59 years and, perhaps, 60-69 years), but there were no such associations with IS at older ages (>70 years). Furthermore, as noted by the authors, the weak associations observed in middle-aged adults could be wholly or largely explained by the positive correlation of TC with systolic blood pressure (SBP). After stratification by baseline SBP, usual TC levels tended to be positively, albeit weakly, associated with IS only among individuals with baseline SBP levels below average (<145 mmHg), and there was no evidence of any association with stroke among those with higher levels of SBP.[13]

Parallel analyses of the Multiple Risk Factor Interventional Trial (MRFIT) data were included in the PSC report that permitted direct comparisons with all other studies included in the PSC. The MRFIT study included 347 681 individuals with no prior history of vascular disease, and contributed an additional 480 IS deaths.[108] Although slightly more extreme effects were observed in the MRFIT study, the results of combined meta-analysis of both the MRFIT and the PSC

studies were similar to those observed in PSC alone.[13] In PSC, as most of the included studies were established before 1990s, the misclassification of stroke pathological types may have underestimated the relevance of usual TC concentrations with IS.

The Emerging Risk Factor Collaboration study (ERFC) included approximately 180 000 individuals with no prior history of vascular disease from 32 prospective studies recruited from populations in Europe or North America, and did not include any Chinese populations.[14] In the ERFC study, non-HDL-C was used as the main indicator of the cholesterol content in atherogenic lipoprotein particles. Over a median follow-up of 6.3 years, 2534 incident IS events were ascertained. After adjustment for established vascular risk factors (i.e., age, sex, SBP, smoking status, diabetes status, and body mass index [BMI]), each 1 mmol/L higher non-HDL-C concentration was associated with a 12% (95% CI: 7-19%) higher risk of IS. Further adjustment for other lipid fractions (e.g., HDL-C and triglycerides) attenuated the strength of these associations marginally (HR 1.11; 95% CI: 1.04-1.18).

While the PSC and the ERFC IPD meta-analyses were primarily confined to studies conducted in Western populations, the Asia Pacific Cohort Studies Collaboration (APCSC) included around 600 000 individuals from 44 studies, mainly in Asia (83%) and Australasia (17%).[110] The APCSC was initiated to investigate the associations of blood pressure with cardiovascular diseases, and measurements of baseline blood lipids were only available in a subset of the included studies. Measurements of TC were available in 25 studies involving 352 033 individuals.[109] The mean (SD) concentrations of TC were slightly higher in

Australasia than in Asia (5.6 [SD 1.13] versus 4.9 [0.98] mmol/L). Over a mean follow-up of 5.7 years, 403 IS deaths, and 749 fatal or non-fatal IS events were recorded. There was a positive association of TC with risk of IS (fatal and non-fatal combined), with an adjusted HR of HR 1.25 (95% CI: 1.13-1.40 per 1 mmol/L higher of usual total cholesterol). The higher risks of IS appeared to be absent at usual levels of TC below about 5.5 mmol/L, on the logarithmic scale (i.e., log risk).

Three additional large-scale individual studies were not included in the previous meta-analyses. In the Apolipoprotein MOrtality RISK study (AMORIS) of around 148 600 men and women with no prior history of stroke at baseline, 4558 incident IS cases were ascertained over a median follow-up of 11.8 years, by linkage with both hospital discharge and death registers. TC and LDL-C (or non-HDL-C) were each positively associated with IS risk. The associations of LDL-C with IS appeared slightly stronger for non-fatal (n=4136; HR 1.12; 95% CI: 1.08-1.16 per 1 SD higher) than for fatal (n=422; HR 1.04; 95% CI: 0.91-1.18) stroke events, and for individuals aged 65 years or below (n=2645; HR 1.11; 95% CI: 1.07-1.15) than above (n=1825; HR 1.03; 95% CI: 0.98-1.08), but were otherwise similar in men and women. However, in the AMORIS study, measurements of HDL-C, SBP, BMI, and smoking were not recorded at baseline. Hence, adjustment for blood pressure or diabetic status was only available in individuals with relevant hospitalizations, which could have substantially underestimated the true prevalence of hypertension and diabetes in this study. As such, the associations of LDL-C with IS could be subject to residual confounding from these established vascular risk factors.

In the Korean National Health System prospective cohort study (KNHS) of approximately 787 442 public servants with no prior history of MI or stroke at baseline, 6328 incident IS cases were ascertained through routinely collected disease events over a 10-year period between 1990 and 2001.[111] IS was strongly and positively associated with plasma TC concentrations down to 3.3 mmol/L or even lower, with a slightly larger effect size than in other studies (HR 1.20; 95%CI: 1.16-1.24 per 1 mmol/L higher TC), probably reflecting the relatively younger and healthier nature of this employee-based study population. Likewise, in the Kailuan study of approximately 96 000 Chinese individuals with no prior history of cardiovascular disease or stroke at baseline, 1156 IS events were recorded during a 4-year follow-up period.[112] Each 1 mmol/L higher non-HDL-C was associated with a 20% (95% CI: 10-33%) higher risk of IS, after adjustment for the effects of various vascular risk factors (i.e., age, sex, BMI, hypertension, diabetes, smoking status, alcohol consumption, and physical activity, and stratified by hospital site), and other lipids including triglycerides and HDL-C. The relative short duration of follow-up compared with other studies may have contributed to the slightly higher effect sizes observed in the Kailuan study compared with those in previously studies.

Additional studies, involving 200-600 incident cases of stroke per study, reported conflicting results for the associations of LDL-C (non-HDL-C, or TC) with IS (**Table 2.5** and **Table 2.6**).[113-115] In some of such studies with relatively modest number of IS cases, the higher risks of IS associated with higher LDL-C were only present in certain subgroups such as men, or younger age, with no evidence of an overall log-linear shape of associations of LDL-C concentrations with risk of IS across the LDL-C range studied.

Limitations of the evidence and implications for future study

Direct measurements of LDL-C were unavailable for most of the previous studies, and the use of non-HDL-C as a surrogate indicator for LDL-C was confined only to a few studies. The confirmation of stroke pathological types also varied, and was probably less reliable in the early studies prior to the introduction of widespread use of brain imaging techniques (e.g., CT/MRI) for stroke diagnosis. The lack of correction for regression dilution bias in some studies may also have underestimated the associations of prolonged exposure to LDL-C differences with risk of IS. More importantly, this review highlights the need for the large-scale prospective epidemiological studies with a sufficiently large number of definite IS cases.

Uncertainties persist about whether the positive associations of non-HDL-C with IS continue down below the ranges reported in previous studies. Nevertheless, prospective epidemiological studies were predominantly conducted in Western populations, where statin therapy has been widely used in recent decades. In contrast, studies conducted in Asian populations with relatively lower TC and LDL-C concentrations and where statin therapy has not been readily accessible, can improve our understanding of the differences in the risk of stroke associated with prolonged differences in LDL-C concentrations. Such studies can also provide evidence on whether lowering LDL-C down to the range below those commonly seen in Western populations (but could be achieved through interventions) will further reduce the risk of IS.

2.3.1.2 High-density lipoprotein cholesterol

The evidence on the associations of HDL-C with IS comes chiefly from the ERFC and the AMORIS studies. The ERFC study,[14] as previously described, reported that HDL-C was inversely associated with risk of IS (HR 0.89; 95% CI: 0.82-0.97 per 1 SD [0.39 mmol/L] higher HDL-C), after adjustment for non-lipid vascular factors. This inverse relationship appeared flat at the upper end of usual HDL-C concentrations. After adjustment for other lipid fractions (i.e., non-HDL-C and triglycerides), the risks were attenuated (HR 0.93; 95% CI: 0.84-1.02), but were statistically compatible as before. The absence of significant associations could reflect the relatively limited number of IS cases in the ERFC study. The AMORIS study of 4470 incident IS cases also demonstrated an overall inverse association of HDL-C with IS (HR 0.88; 95% CI: 0.85-0.90 per 1 SD [0.40 mmol/L] higher [calculated using published group-specific estimates]).[116] The major limitations of AMORIS study have been previously described (i.e., lack of collection of data on some major vascular risk factors), and as such, the associations of HDL-C with IS could also be susceptible to residual confounding.

The APCSC study included 271 IS cases in 25 studies,[117] but did not find any association of HDL-C with IS. The Framingham study reported an inverse relationship of HDL-C with risk of IS (n=301, HR 0.77; 95% CI: 0.68-0.87 per 1 SD higher HDL-C [Log_e]), but the log-transformation of HDL-C values may suggest the presence of outliers or individuals with extreme values which have complicated the interpretation of the study.[118] One recent study involving combination of the Copenhagen City Heart Study (CCHS) and Copenhagen General Population Study (CGPS) reported a “U-shaped” association of HDL-C with all-cause and

cause-specific mortality (including mortality from IS [n=2399]) for both men and women, suggesting slightly higher risk of IS associated with HDL-C above about 1.9 mmol/L in men, and 2.4 mmol/L in women.[119] Interpretation of such findings requires caution, as the limited number of IS cases was associated with wide confidence intervals in certain subgroups. Further, the lack of information on the population characteristics throughout the range of HDL-C studied may have resulted in potential residual confounding, or reverse causality. Comparisons of such paradoxical results with studies conducted in other populations are not currently available, as the number of individuals with HDL-C above 2 mmol/L in general populations are too few to be informative in many of the previous studies.

Furthermore, in some studies, the absence of any overall associations of HDL-C with IS makes it difficult to interpret isolated associations in certain subgroups other than being due to the play of chance.[120] For example, although a somewhat higher risk of IS was observed in certain circumstances, such as women rather than men,[121] these subgroup-specific findings should be interpreted with caution, in particular, in the absence of any overall effect in such studies.

Limitations of the evidence and implications for future study

Previous studies reported inverse associations of HDL-C with IS, although there was also some attenuation towards the null observed approaching the top of the range of HDL-C concentrations, probably reflecting the limited number of cases with HDL-C measurements around this range. However, the inverse relationship of HDL-C with IS, and its independence from LDL-C, has not been previously reported in other populations, particularly, in populations with relatively low LDL-C.

2.3.1.3 Triglycerides

Compared with studies on associations of cholesterol with IS, the available evidence relating plasma triglyceride concentrations to IS has also been conflicting, prompting further uncertainties about the independent relevance of triglycerides for stroke risks. The ERFC study, which provided more detailed adjustment for confounders and correction for regression dilution bias than previous studies, reported continuous positive associations of plasma triglyceride concentrations with IS (HR 1.26; 95% CI: 1.18-1.34 per SD [0.68] higher triglycerides [Log_e]), albeit the strength of the association was weaker than for IHD (HR 1.51; 95% CI: 1.44-1.59), adjusted for age, and sex. However, the relevance of triglycerides for IS was attenuated substantially (~1/3) after adjustment for SBP (HR 1.17; 95% CI: 1.11-1.24), and further attenuated towards the null after additional adjustment for other vascular risk factors, including smoking, diabetes status, BMI (HR 1.12; 95% CI: 1.05-1.19), and non-HDL-C, and HDL-C (HR 1.02; 95% CI: 0.94-1.11).[83]

The combined analyses of the Copenhagen City Heart Study (CCHS) and the Copenhagen General Population Study (CGPS) had similar statistical power as the ERFC study, and reported similar strength of associations of triglycerides with risk of IS, but up to much higher non-fasting triglyceride concentrations than those observed in ERFC (6.6 versus 2.7 mmol/L).[83] However, adjustments were restricted to age, sex, and study, and no results were reported after additional adjustment for established vascular risk factors. Specifically, in the CCHS study of 13 956 individuals with 1529 IS cases,[107] adjustment for SBP altered the associations of triglycerides with IS marginally, and further adjustment for diabetes, BMI, and HDL-C attenuated the associations, but the associations remained

statistically significant. No adjustment for non-HDL-C was made in the CCHS study. The multivariable adjusted HRs associated with each 1 mmol/L higher plasma triglyceride concentrations were slightly stronger in women (HR 1.28; 95% CI: 1.15-1.43) than in men (HR 1.12; 95% CI: 1.04-1.20). The AMORIS study also described a continuous positive association of plasma triglyceride concentrations with IS (HR 1.17; 95% CI: 1.15-1.21 per 1 SD higher triglycerides [Log_e]), with slightly stronger effects in women than in men.[116] As previously described, in the AMORIS study, some residual confounding may persist, due to the imprecise measurements of SBP, smoking, and diabetes.

The APCSC study reported a weak positive log-linear relationship of triglycerides with IS (HR 1.35; 95% CI: 1.00-1.83 per 1 SD higher triglycerides [Log_e]), after adjustment for age at risk, SBP, smoking status, and the ratio of TC to HDL-C levels.[122]

Limitations of the evidence and implications for future studies

The evidence on associations of plasma triglyceride concentrations with risk of IS has been less consistent than those observed for other cholesterol fractions. It is unclear whether triglycerides play an independent role in IS, and more generally, in atherosclerosis. However, the main issue with previous studies is the inadequate statistical power to confirm or refute any such associations, and when available, to what extent this could be relevant to other vascular risk factors, including SBP or other cholesterol fractions.

Triglycerides per se are less likely to be a causal risk factor for IS, and more generally, for atherosclerosis,[83, 123] but the triglyceride-mediated pathways may potentially be relevant in atherogenesis.[124] However, if the relevance of

triglycerides for risk of IS is mediated through the cholesterol content in the triglyceride-rich lipoprotein particles, or through HDL particles, additional adjustment for cholesterol fractions may represent an over-adjustment when assessing the IS risk associated with differences in plasma triglyceride concentrations. Reliable assessment of the independent relevance of triglycerides for risk of stroke may be difficult to determine in the observational studies alone.

2.3.2 Haemorrhagic stroke

Compared with IS, the available evidence on the associations of cholesterol with HS is much more limited, due to the paucity of HS cases in previous studies conducted in Western populations. Overall, 7 IPD meta-analyses from 4 studies, 1 published data based meta-analysis, and 8 individual prospective report from 6 studies, were identified that examined the associations of major blood lipids or apolipoproteins with HS (**Tables 2.2**, **2.4**, **2.5**, and **2.7**). These studies were primarily conducted in Asian populations. The overall observational evidence appeared to suggest an inverse association of plasma cholesterol concentrations (e.g., TC, or LDL-C) with risk of HS, particularly with high levels of SBP. Greater uncertainties remain about the associations of HDL-C and of triglycerides with risk of HS.

2.3.2.1 Low-density lipoprotein cholesterol

The large KNHS study included 3947 HS cases, involving information on diagnoses from routinely collected data.[111] After adjustment for age and sex only, measurements of serum TC were not associated with HS risk. However, after additional adjustment for other vascular risk factors (i.e., BMI, height, serum glucose, hypertension, alcohol consumption, smoking, physical activity, income,

and area of residence), each 1 mmol/L higher TC was associated with an HR of 0.91 (95% CI: 0.87-0.95) for risk of HS. The higher risks associated with lower TC were only observed among individuals with hypertension (HR 0.89; 95% CI: 0.84-0.94), although it was statistically compatible with those in non-hypertensive individuals (no group-specific results were reported). In a subset of the study population, further stratification by gamma-glutamyl transferase (GGT) levels, a biomarker for alcohol consumption, suggested the inverse relationship of TC with risk of HS was only apparent among individuals with high levels of alcohol consumption, irrespective of hypertension status. However, the confidence intervals were wide, and no comparisons were made about the characteristics of participants with or without GGT measurements in this study.

In the PSC study involving around 2700 HS deaths, there was a weak inverse association of usual TC with HS risk, which was more extreme in individuals with high blood pressure.[13] Estimates obtained by combining the MRFIT and the PSC studies were compatible with those of PSC alone.[108] The APCSC study including 614 HS deaths reported that each 1 mmol/L higher usual TC across the range studied (4.5-5.5 mmol/L) was associated with 20% (95% CI: 8-30%) lower risk of death from HS.[109] Similar associations were observed over a longer follow-up period (HR 0.73; 95% CI: 0.64-0.84 per 1 mmol/L higher TC), in an updated report conducted in men only.[125]

Other studies, with comparatively fewer numbers of HS cases, also suggested inverse associations of TC with risk of HS, although were non-significant,[126, 127] or confined to the elderly.[128] The significant higher risks of HS associated with lower cholesterol concentrations observed in studies with relatively small number of

definite HS cases requires confirmation in well-powered prospective studies. In contrast, neither of the ERFC, nor the AMORIS studies, with, respectively, 513 and 876 HS events, showed clear evidence of association of non-HDL-C or LDL-C with risk of HS.

2.3.2.2 High-density lipoprotein cholesterol and triglycerides

Amongst the large prospective epidemiological studies identified, there was no consensus from the available evidence about the relevance of either HDL-C or triglycerides for risk of HS (see **Tables 2.2-2.7**), perhaps due to the insufficient statistical power in such studies.

Limitations of the evidence and implications for future study

There is limited data available about the shape of associations of LDL-C with risk of HS, and whether the proportional excess risk with lower LDL-C is relevant only in individuals with high blood pressure. Future studies involving a large number of definite HS cases are needed to assess the shape and strength of such associations. Furthermore, reliable evidence from large-scale prospective studies is also required to characterise the associations of HDL-C and of triglycerides with risk of HS.

Observational studies can provide evidence about the associations of risk factors with disease outcomes, but they do not necessarily confirm or refute the causal nature of any such associations. MR studies and RCT evidence are therefore needed to offer further insights into the causality of such observations. If lowering LDL-C would perhaps increase the risk of HS, it is less clear to what extent the probable excess HS risk may be outweighed by the benefits of reductions in IS

risk, especially in populations where the underlying HS rates are comparatively high, such as in Asian populations, with a high prevalence of hypertension, and high rates of stroke.

2.4 Mendelian randomisation evidence

Observational epidemiological studies, as previously described, are constrained by inherent limitations in their design, including confounding, measurement error, reverse causality and other biases. It is not possible to draw conclusions about the causal nature of the associations of lipid fractions with stroke pathological types, with observational evidence alone. One approach which can overcome many these limitations is using MR, which has been widely applied in studies of CHD associated with modifiable risk factors. MR studies have confirmed the causal role for LDL-C,[101, 129] and demonstrated the atherogenic causal relevance of triglyceride-mediated pathways,[124] and appeared to have refuted the causal role of HDL-C for CHD, albeit uncertainties remain regarding the relevance of HDL function.

Compared to the amount of evidence concerning risk of CHD, the available data on stroke risks is much more limited. Overall, four fully-published reports, and one meta-analysis study published only in abstract form were identified. In the METASTROKE Consortium, involving almost 12 000 IS cases and 62 000 controls, 12 genetic variants, which had specific effects on LDL-C but not on other lipoproteins, were included in the genetic risk score (GRS) for LDL-C. The GRS, corresponding to 1 mmol/l higher LDL-C was associated with a non-significant 8% higher IS risk (OR 1.08; 95% CI: 0.93-1.25), with no significant heterogeneity in the effects of the GRS by IS subtypes (i.e., large artery atherosclerotic, cardio-

embolic, small vessel, $p_{heterogeneity}=0.88$).[130] In contrast, the same LDL-C GRS was associated with a two-fold higher IHD risk (OR 2.14; 95% CI: 1.67-2.73), estimated in the HPS and PROCARDIS studies (8419 cases and 9257 controls).[130] In an updated METASTROKE report, the relevance of *PCSK9* genetic variants, which are relatively rare in Western population but have a large effect on LDL-C, was assessed with risk of IS (10 307 IS cases and 19 326 controls), and IHD (60 801 CHD cases and 123 504 controls). The loss-of-function *PCSK9* genetic variant (R46L: rs11591147) was associated with 0.5 mmol/L lower LDL-C concentrations, and 23% lower CHD risk (OR 0.77; 95% CI: 0.69-0.87), but there was no significant association of the genetic variant with IS risk (OR 1.04; 95% CI: 0.84-1.28). A recent meta-analysis of nine studies of blacks (n=17 459) and whites (n=31 306) found *PCSK9* loss-of-function variants were associated with lower LDL-C concentrations in both blacks (0.91 [0.19] mmol/L) and whites (0.34 [0.13] mmol/L), and were associated with lower risk of CHD (In blacks: n=851 cases; OR: 0.51; 95% CI: 0.28-0.92; In whites: n=2662 cases; OR=0.82; 95% CI: 0.63-1.06), but were not significantly associated with stroke (In blacks: n=523 cases; OR: 0.84; 95% CI: 0.48-1.47; In whites: n=1660 cases; OR=1.06; 95% CI:0.80-1.41).[131] In this study, the effects of *PCSK9* loss-of-function variants on LDL-C were stronger in blacks than in whites, due probably to the different sets of *PCSK9* variants used between blacks (Y142X: rs67608943 and C679X: rs28362286) and whites (R46L: rs11591147). It is less informative to compare directly the effects of *PCSK9* variants on stroke in this study with others (plus between blacks and whites), because of the lack of information on stroke pathological types, in addition to the relatively limited stroke cases.

The Framingham study examined the impact of HDL-C on IS using a GRS involving 47 previously identified genetic variants of HDL-C (describing approximately 6% of the variations in HDL-C), and found the GRS was not associated with IS, due probably to the relatively limited number of IS cases.

Another MR analysis from CCHS study examined association of genetic variation in the *CETP* gene (*CETP* -629 C>A [rs1800775] and *Taq*1B G>A [rs708272]) that resulted in lower CETP activity with risk of IS (n=1064) in European populations.[132] In the latter study, compared to those without any of such *CETP* variants, individuals with four HDL-C raising alleles had a lower risk of IS (OR 0.71; 95% CI: 0.58-0.88). However, the *CETP* variants included increase HDL-C, but also decrease both LDL-C and triglycerides in European populations,[92] comparable to the effects of pharmacological CETP inhibitors (e.g., anacetrapib and evacetrapib).[133] As such, it is difficult to distinguish whether the observed cardio-protective effects caused by the *CETP* variants in the CCHS study is due to the changes in HDL-C or LDL-C. Recent MR evidence (although on IHD) has suggested the potential benefits of CETP inhibition be largely proportionate to the reductions of plasma concentrations of LDL-C,[134] or Apo B.[135, 136] Based on the previous statin trials,[15] and other MR studies,[101, 134] the proportional risk reduction of IS estimated in the CCHS study was, however, much greater than the life-long lowering of LDL-C alone caused by the CETP deficiency. In addition, the function of CETP per se is complex, as it may prevent atherosclerosis by promoting clearance of cholesteryl ester from peripheral tissues, and may also cause atherosclerosis by redistributing cholesteryl esters from HDL particles to triglyceride-rich lipoproteins, thereby changing the density and diameter of such particles as well, with LDL particles becoming smaller and denser.[74]

Limitations of the evidence and implications for future study

Previous MR studies have not provided reliable evidence on major blood lipids and stroke risks. The METASTROKE study appeared to suggest that any relevance of LDL-C for risk of IS was weaker than that with IHD, in contrast to the comparable beneficial effect on risk reduction of IS and IHD associated with lowering LDL-C with statin treatment.[15, 130] The METASTROKE study of IS in Western populations has included the largest number of IS cases. Nevertheless, heterogeneity between the individual studies may have probably limited the detection of any significant associations of LDL-C with risk of IS in the METASTROKE study. Future well-powered MR studies with confirmed diagnoses of IS from a single population are needed to assess the relevance of major blood lipids for IS. Furthermore, there have been no MR studies that investigated the causal relevance of LDL-C with HS risk, and no previous MR studies about HDL-C and triglycerides with stroke of any type.

Despite the potential of MR studies to be a useful tool to investigate causal relationships, the assumptions inherent in MR studies have important limitations, including pleiotropic effects of genetic variants, linkage disequilibrium (LD) between genetic variants, and the underlying population stratification (**Table 1.3**). Moreover, differences in the proportional effects of lipids explained by individual genetic variants identified in previous GWAS studies have been only modest,[92, 93] which results in weak instrument bias. Combining multiple genetic variants into a GRS can offset some of the effects of the weak instrument bias.[105] Collectively, the interpretation of MR causal investigation should take account of biological mechanisms in addition to statistical and methodological considerations.

2.5 Randomised controlled trial evidence

Epidemiological studies using conventional observational approaches or MR studies using genetic data can help to generate hypotheses about the likely effects on risk of disease outcomes associated with prolonged differences in such risk factors, and investigate causality. However, to shift overall lipid concentrations to a healthier range at population levels through lifestyle interventions (e.g., dietary factors), or among people with high risk of vascular disease through drugs (e.g., statins), informed decisions require evidence from reliable large-scale RCTs, or meta-analyses of such trials to permit sufficient statistical power.

In this thesis, an appraisal of the RCT evidence of the effects of treatment to lower LDL-C concentrations has been chiefly obtained from the CTT IPD meta-analysis of all of the large cholesterol lowering trials involving at least 1000 participants who were treated for at least two years.[137] The criteria used in CTT study were applied to help the search of relevant RCT evidence on HDL-C and triglycerides, which have not been included in two recent systematic reviews.[78, 83]

2.5.1 Low-density lipoprotein cholesterol

LDL-C has been widely acknowledged as a causal factor for atherosclerosis in animals,[138] monogenic and polygenic associations in humans,[101, 139] and in randomised trials of LDL-C lowering therapy.[18] Statins are a class of drugs that reduce LDL-C concentrations through inhibition of the activity of HMG CoA reductase. Different statins have different potencies, with the newer agents (e.g., atorvastatin and rosuvastatin) able to produce larger reductions in LDL-C than the older agents (e.g., simvastatin and pravastatin).[140] Large-scale evidence from the CTT indicates each 1 mmol/L reduction in LDL-C with statin therapy was

associated with a 25% reduction in the risk of major vascular events (coronary deaths, myocardial infarction [MI], stroke, and coronary revascularisation) during each year (after the first) of allocated statin treatment.[15, 18]

In the CTT meta-analyses, which included around 500 HS events, there was a non-significant excess risk of HS associated with lowering LDL-C with statin treatment (RR 1.15; 95% CI; 0.96-1.36 per 1 mmol/L lower LDL-C). In contrast, in the Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial of 4731 individuals with prior cerebrovascular disease (which was currently not included in the CTT IPD meta-analyses), random allocation to atorvastatin treatment versus placebo was associated with a 44% (95% CI: 6-95%) higher risk of HS.[141] When the SPARCL results were combined with those from the other trials included in the CTT meta-analysis, the rate ratio of HS was 1.21 (95% CI: 1.05-1.41) per 1 mmol/L reduction in LDL-C.[15]

In European and North American populations (rates of HS as estimated in the CTT study of 0.1% per annum), the proportional risk of HS associated with lowering LDL-C by statin therapy would translate into an absolute 5-10 extra HS per 10 000 patients treated in whom LDL-C was reduced by 1-2 mmol/L for every 5 years with statin therapy (calculated by multiplications of the incidence rates of HS [0.1%], the proportional excess risk [10-20%], the treatment duration [5 years], and the number of patients treated [10 000 patients]).[15, 18] However, the absolute excess in HS may be higher in populations (e.g., Asia), where the underlying rates of HS have been reported higher, particularly among those with prior cerebrovascular events, or hypertensive patients.

2.5.2 High-density lipoprotein cholesterol

Although prospective observational epidemiological studies have shown that there is an inverse association of HDL-C with IHD and IS, reports from several RCTs were unable to demonstrate any reduction in risk of composite cardiovascular disease, including stroke of any type, alone, or in the composite outcomes associated with raising HDL-C concentrations used in such trials. Although each of these trials has their own limitations, the consensus from all such trials suggests that raising HDL-C (using the drugs tested so far) alone does not appear associated with a lower risk of stroke or CHD. The Coronary Drug Project (CDP) was conducted in the pre-statin era, and showed the benefits of niacin treatment versus placebo in reducing the risk of CVDs.[142] The significant reductions in LDL-C caused by the administration of niacin, perhaps, contribute to, or even account for the benefits (using the proportional risk reduction by lowering an equivalent amount of LDL-C estimated in statin trials[15]). Two additional trials investigated the effects of the addition of niacin to statins. However, neither the AIM-High trial, nor the large HPS2-THRIVE trial, were able to demonstrate any benefits of niacin therapy in addition to statin treatment,[143, 144] and also reported excess risks of serious adverse effects associated with niacin treatment.

Another approach to raising the HDL-C concentrations is to inhibit CETP that involved in the reverse cholesterol transport process. The HPS3/TIMI55-REVEAL trial of 30 449 men and women who had prior history of atherosclerotic vascular events, and on intensive statin treatment, demonstrated that the addition of the CETP inhibitor (anacetrapib) to intensive statin therapy for around four years resulted in a significant 9% (95% CI: 3-15%) reduction in risk of major coronary

events (had no significant effects on presumed IS), without any of those previously hypothesized adverse effects during the follow-up period.[145] In the REVEAL study, patients who were randomly allocated to anacetrapib treatment had increased HDL-C concentrations by 100%, and decreased non-HDL-C concentrations by 18%, approximately, compared to the placebo group. Consequently, the observed risk reduction in REVEAL trial with anacetrapib treatment is largely consistent with the estimated risk reduction of similar amount of non-HDL-C lowering with statin treatment. The result from this large-scale RCT study has cast doubt on the relevance of raising HDL-C, or other actions (apart from effects on non-HDL-C) that anacetrapib played for modifying the risk of coronary events.

The significant beneficial effects with treatment of anacetrapib added on intensive statins in the REVEAL trial appeared to differ from those of other trials of CETP inhibitors, although the latter studies also had significant limitations. The Investigation of Lipid Level Management to Understand its Impact in Atherosclerotic Events (ILLUMINATE) trial of torcetrapib was stopped prematurely due to the off-target drug effects (significantly elevated SBP and circulating aldosterone concentrations). Another two trials – the Effects of Dalcetrapib in Patients with a Recent Acute Coronary Syndrome (Dal-OUTCOME) trial and the Assessment of Clinical Effects of Cholesteryl Ester Transfer Protein Inhibition with Evacetrapib in Patients at a High Risk for Vascular Outcomes (ACCELERATE) trial – were stopped early (approximately for two years) for futility.[146, 147] Notably, the dalcetrapib tested in the Dal-OUTCOME trial had relatively weak effects on lipid profiles (no effects on LDL-C), whereas evacetrapib in the ACCELERATE trial increased HDL-C by as much as 130% and reduced LDL-C by nearly 35%. The evacetrapib is of similar potency to anacetrapib, but the ACCELARATE trial

involved only about one-third as many participants as the REVEAL trial, and had a much shorter duration of treatment (2 years). The totality of evidence appears to cast doubt on the causal role of HDL-C as a risk factor for stroke or CHD, which probably acts as a biomarker reflecting, perhaps, a combination of various metabolic and inflammatory processes, in addition to the reverse cholesterol transport system. It also suggests that CETP inhibitors are involved more in blood cholesterol re-distribution, than the overall cholesterol reduction caused by other treatments (e.g., statins).

2.5.3 Triglycerides

No large-scale randomised trials have examined the effect of reducing triglycerides on cardiovascular disease risk in individuals with raised plasma triglyceride concentrations. A meta-regression analysis of the effect of triglyceride-lowering using fibrate regimens reported that a 0.1 mmol/L reduction in plasma triglycerides concentrations resulted in a 5% (95% CI: 1-10%) lower risk of coronary disease, but had no effects on stroke.[148] As noted in this study, fibrates can also decrease LDL-C by around 0.1 mmol/L, which would translate to a 3% (95% CI: 2-4%) lower risk of coronary events.[15] This study had access only to published data, and thus, was unable to investigate the effects of triglycerides had after accounting for effects of other lipids on coronary events, and may also be subject to publication bias. Hence, in the absence of individual participant data, substantial uncertainties persist about the relevance of lowering triglycerides on risk of ischaemic vascular diseases. More generally, the relevance of elevated triglycerides for atherosclerosis is also unclear, as it is uncertain if it is triglyceride per se or triglyceride-mediated pathways that are atherogenic.[83]

2.6 Summary of findings and main research questions

This review of the available literature demonstrates considerable gaps in the evidence about the relevance of major lipids to stroke risks by pathological type. In both observational and genetic studies, the substantially weaker effects of LDL-C on risk of IS compared with IHD in Western populations remain unexplained. Moreover, it is unclear whether the relevance of LDL-C to IS continues down to a lower range of values than those observed in Western populations. There is a lack of evidence about whether there is an inverse association of LDL-C with ICH risk, and if this is causal. It is therefore unclear the extent to which the increased risks of ICH associated with lower LDL-C are offset by the risk reduction in IS, and other ischaemic vascular disease outcomes. Likewise, there is a lack of evidence about the shape and strength of the associations of both HDL-C and triglycerides, with IS and ICH in the Chinese population.

2.7 Aims and objectives of thesis

The aims of this thesis were to examine the distribution of major lipids and their associations with two main stroke pathological types (IS and ICH). In particular, the thesis examined the shape, the strength and the independence of the associations of major blood lipids with stroke pathological types, using both conventional and genetic epidemiological approaches in the China Kadoorie Biobank prospective (CKB) study. The specific objectives of this thesis were to:

1. To examine the distribution of major lipids by age, sex and levels of major risk factors for vascular disease (including SBP and cigarette smoking).

2. To assess the shape, strength, and independence of associations of major blood lipids (i.e., LDL-C, HDL-C, and triglycerides), and apolipoproteins (i.e., Apo A1 and Apo B100) with risk of IS.
3. To characterise the shape, strength, and independence of the associations of major blood lipids (i.e., LDL-C, HDL-C, and triglycerides), and apolipoproteins (i.e., Apo A1 and Apo B100) with risk of ICH.
4. To investigate the causal nature, if any, of the associations between major blood lipids with IS using genetic risk scores in MR studies.
5. To investigate the causal nature, if any, of the associations of major blood lipids with ICH using genetic risk scores in MR studies.

Table 2.1: Literature review search terms

Key words	Search terms	Query terms, 08-Sep-2016
Stroke	stroke intracerebral haemorrhage intracerebral haemorrhage intracranial haemorrhage intracranial haemorrhage cerebral haemorrhage cerebral haemorrhage cerebral infraction cerebrovascular disorders brain ischaemia brain ischemia	((lipids[Title/Abstract] OR lipid[Title/Abstract] OR lipoproteins[Title/Abstract] OR lipoprotein[Title/Abstract] OR apolipoproteins[Title/Abstract]) OR apolipoprotein[Title/Abstract] OR cholesterol[Title/Abstract] OR triglycerides[Title/Abstract] OR triglyceride[Title/Abstract] OR triacylglycerols[Title/Abstract] OR triacylglycerol[Title/Abstract] AND (stroke[Title/Abstract] OR intracerebral haemorrhage [Title/Abstract] OR intracerebral haemorrhage [Title/Abstract] OR intracranial haemorrhage [Title/Abstract] OR intracranial haemorrhage [Title/Abstract] OR cerebral haemorrhage [Title/Abstract] OR cerebral haemorrhage [Title/Abstract] OR cerebral infarction [Title/Abstract] OR cerebrovascular disorders [Title/Abstract] OR brain ischaemia [Title/Abstract] OR brain ischemia [Title/Abstract])) AND ("1970"[CRDAT] : "3000"[CRDAT])
Lipids	lipid lipids cholesterol triglyceride triglycerides triacylglycerol triacylglycerols lipoprotein lipoproteins apolipoprotein apolipoproteins	
Time periods	1970 -	

Searches were undertaken in PubMed (1970 to 08-September-2016) and repeated weekly until 31 August 2017. Only literature published after 1970 was included, when the first case of clinical computed tomography scan on patients took place approximately (1971).

Figure 2.1: A flowchart of literature search and the identification of studies relating to the research question

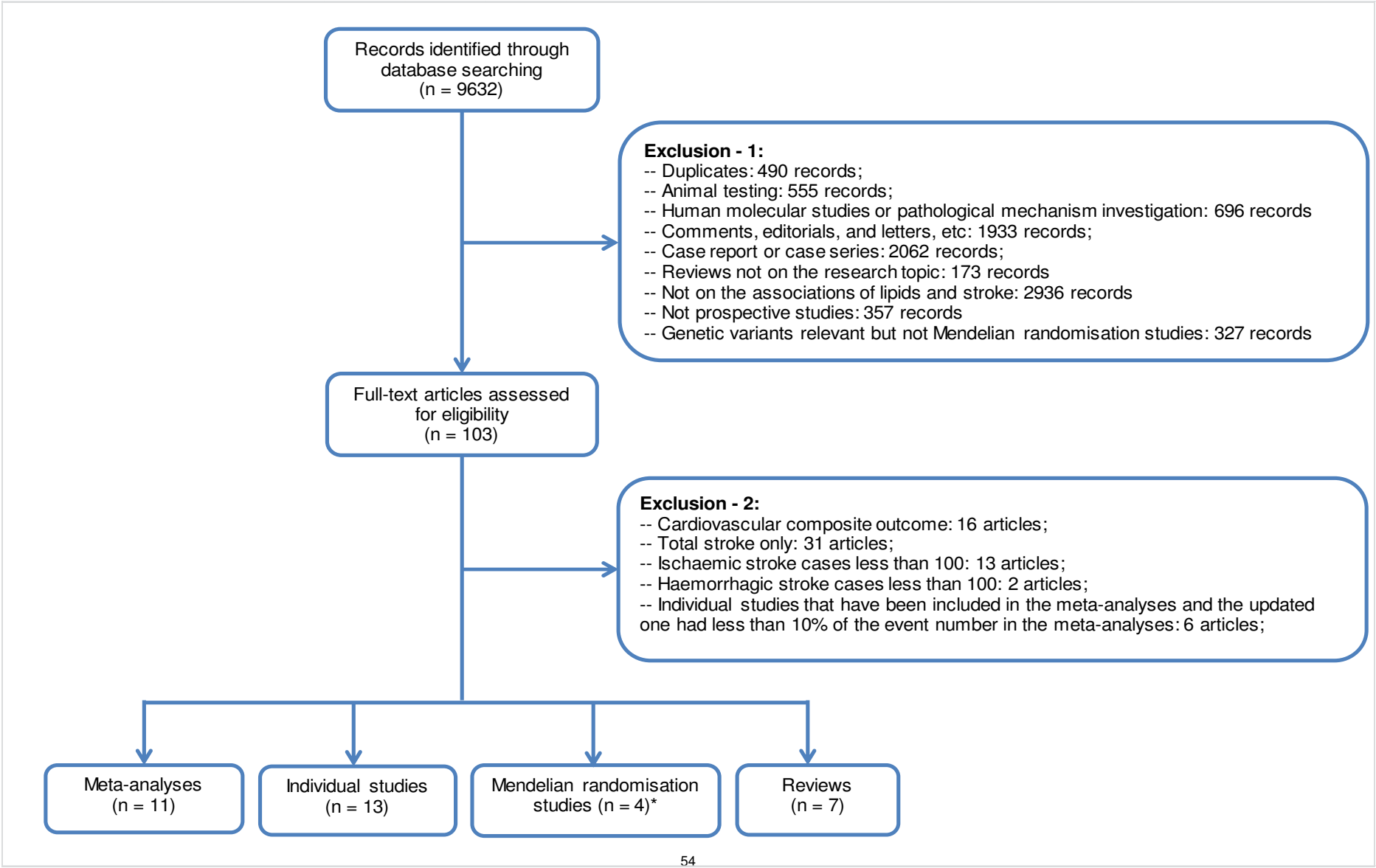
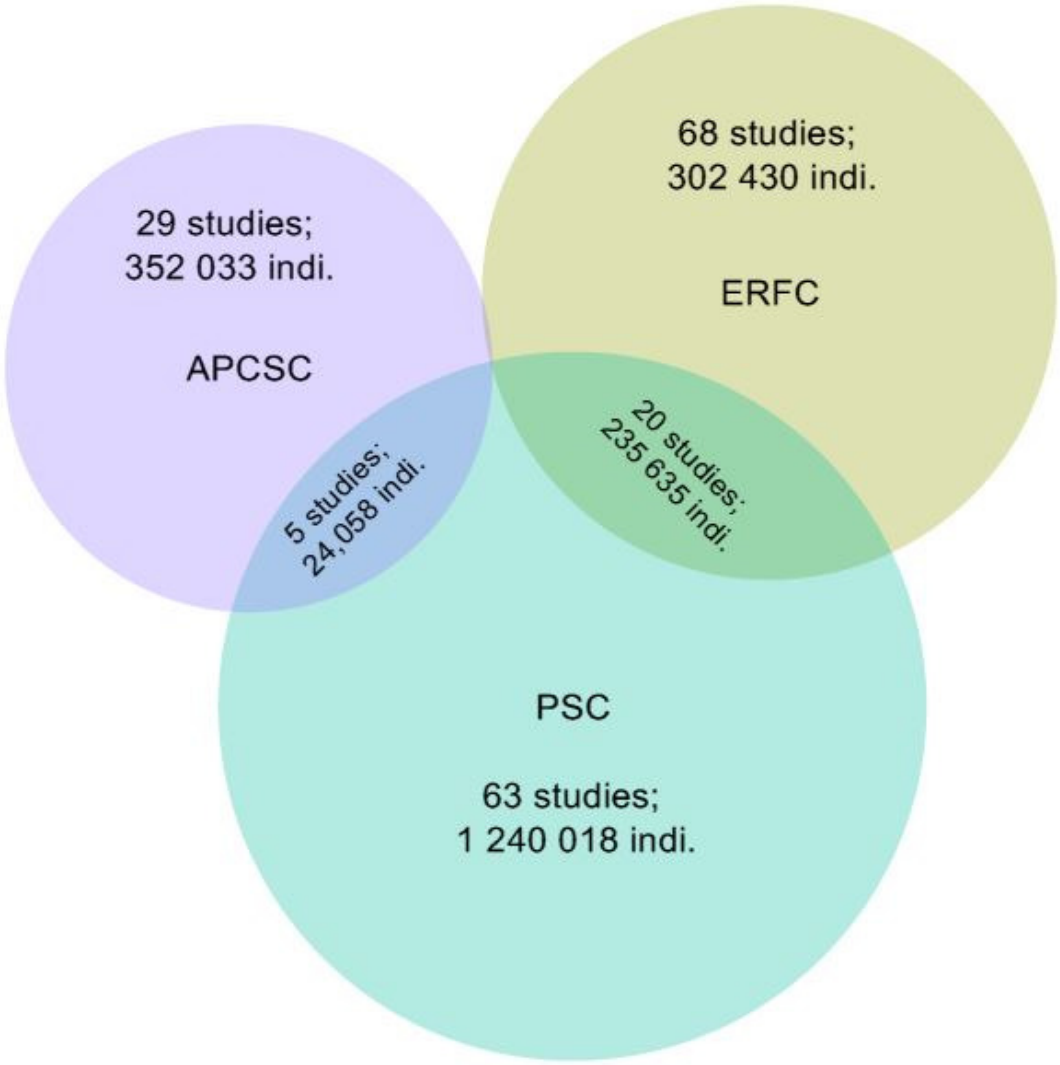


Figure 2.2: Venn Diagram for three main individual participant data based meta-analyses included in the thesis



PSC=Prospective Studies Collaboration; ERFC=The Emerging Risk Factors Collaboration; APCSC=Asia Pacific Cohort Studies Collaboration; Indi.=individuals. Number of individuals overlapped between studies was taken from PSC.[13]

Table 2.2: Characteristics of the meta-analyses examining the associations of major blood lipids with stroke, by pathological type

Study (Year published)	No. of studies	No. of participants (female, %)	Geographical locations	Mean age at baseline, years	Mean follow-up, years	No. of events*	Diagnostic tool, %
IPD meta-analysis							
ERFC (2007)[14]	IS: 32 HS: 19	173 312 (49) 142 916 (51)	Global	NA	NA	2534 513	MR + DC
PSC (2007)[13]	IS: 61 HS: 61	NA	Global	NA	13	2134 F 2700 F	DC
APCSC-1 (2003)[109]	IS: 29 HS: 29	352 033 (43)	Asia (81%), Australasia (19%)	47	8.5	749 614 F	MR + DC DC
APCSC-2 (2004)[122]	IS: 26 HS: 26	96 224 (48)	Asia, Australasia	49	7.9	273 195	MR + DC
APCSC-3 (2007)[117]	IS: 25 HS: 25	79 694 (46)	Asia, Australasia	49	6.8	271 222	MR + DC
APCSC-4 (2013)[125]	IS: 44 HS: 44	286 411 (0)	Asia (86%), Australasia (14%)	52	9.4	1448 1314	MR + DC
WPP (2002)[149]	IS: 8	24 343 (100)	America	52	13.9	461 F	DC
EUROSTROKE (2002)[121]	IS: 4	1251 (34)	Europe	62	NA	346	MR
EPOCH-JAPAN (2012)[127]	IS: 10 HS: 10	65 594 (59)	Japan	58 58	10.1 10.1	457 F 212 F	DC
JALS-ECC (2010)[150]	IS: 10	22 430 (60)	Japan	58	7.6	224	MR + DC
Pub meta-analysis							
Wang X (2013)[151]	HS: 23	1 430 141 (NA)	Varied	Varied	Varied	HS: 7960	Varied

IPD=Individual participant data; Pub=Published literature or aggregate study level meta-analyses; IS=Ischaemic stroke; HS=Haemorrhagic stroke; NA=Not available; MR=Medical records; DC=Death certificates; F=Fatal events; * Fatal events were labelled as "F", where otherwise the events consisted of both fatal and non-fatal events.

Table 2.3: Characteristics of the prospective studies examining the associations of major blood lipids with ischaemic stroke

Study (Year screened)	Bio-sample characteristics				Baseline population characteristics				Follow-up characteristics		
	Lipids	Fasting (F) / non-fasting (NF)	Serum (S) / Plasma (P)	Assay methods	Geographical locations	No. of participants (female, %)	Mean age at baseline, years	Lipid levels, mean (SD), mmol/L	Mean follow-up, years	No. of events	Diagnostic tool, %
CCHS (1976)	TG	NF	P	Enz.	Denmark	19 356 (54)	NA	N/S	31	1529	MR
CCHS (1991-1994) +CGPS (2003)	HDL-C	NF	P	Col.	Denmark	116 508	M: 58 (10) W: 57 (10)	M: 1.4 (0.3) W: 1.7 (0.3)	6	1204	MR
CRCS (1985)	HDL-C	NF	S	Gel.	Japan	310 (49)	65	N/S	11	155	CT/MRI (100%)
AMORIS, (1986)[116]	TC	F (62%)	S	Enz.	Sweden	148 600 (44)	M: 49 (11) W: 53 (13)	M: 6.1 (1.2) W: 6.0 (1.1)	12	4558	MR + DC
	HDL-C			Cal.				M: 1.4 (0.4) W: 1.7 (0.4)	12	4558	
	TG			Enz.				M: 1.5 (0.8) W: 1.2 (0.7)	12	4558	
	LDL-C			Cal.				M: 6.1 (1.2) W: 6.0 (1.1)	12	4558	
7 Chinese cohorts (1987)[113]	TC	F	S	Enz.	China	10 825 (55)	55 (10)	5.1 (1.4)	8.4	250	CT/MRI (63%)
	TG							1.5 (0.8)	8.4	250	CT/MRI (63%)
JPHC (1990)[114]	TC	NF	S	Enz.	Japan	33 469 (65)	54	4.9	12	354	MR + DC
CMPCS (1992)[115, 152, 153]	TC	NA	S	Enz.	China	30 384 (NA)	47	4.7	10	314	MR + DC.
	HDL-C	NA	S	Enz.	China	30 384 (NA)	47	M: 1.4 (0.4) W: 1.5 (0.4)	10	427	
	non-HDL-C	F	S	Cal.	China	29 937 (47)	47	M: 3.3 (1.1) W: 3.4 (1.2)	8	420	
IPHS (1993)[126]	LDL-C	NF	S	Cal	Japan	912 19 (66)	59	3.04	10	539	DC
FPS	TC	NA	P	Enz.	U.S.	6276 (56)	64 (10)	5.7 (1.1)	10	301	MR + DC
	HDL-C					6276 (56)	64 (10)	1.3 (0.4)	10	301	
The Three-City Study (1999)	TG	F	S	Enz.	France	8046 (61)	74	NA	5	130	MR + DC
The Kailuan Study (2006)	non-HDL-C	F	S	Cal.	China	95 916 (20)	51	3.38	4	1156	MR

Enz.= Enzymatic methods; Cal.=Calculated; Col.= Colorimetric; Gel.= Gel-permeation chromatography; M=Men; W=Women; Individual studies that have been included in identified individual-participant-based meta-analyses (Table 2.2) were not reviewed separately. Other annotations are the same as Table 2.2, unless otherwise stated.

Table 2.4: Characteristics of the prospective studies examining the associations of major blood lipids with haemorrhagic stroke

	Bio-sample characteristics				Baseline population characteristics				Follow-up characteristics		
Study (Year screened)	**Lipids**	**Fasting (F) / non-fasting (NF)**	**Serum (S) / Plasma (P)**	**Assay methods**	**Geographical locations**	**No. of participants (female, %)**	**Mean age at baseline, years**	**Lipid levels, mean (SD), mmol/L**	**Mean follow-up, years**	**No. of events**	**Diagnostic tool, %**
KPMCP (1977)	TC	F	S	Enz.	U.S.	61 756 (54)	M: 54 (10) F: 55 (10)	M: 6.0 F: 6.2	11	386	MR + DC
KNHS (1986)	TC	NA	S	Enz.	Korea	787 442 (16)	42	5.2	10	3947	MR + DC
AMORIS (1986)[116]	TC	F (62%)	S	Enz.	Sweden	148 600 (44)	M: 49 (11) F: 53 (13)	M: 6.1 (1.2) F: 6.0 (1.1)	12	865	MR + DC
	HDL-C			Cal.				M: 1.4 (0.4) F: 1.7 (0.4)			
	TG			Enz.				M: 1.5 (0.8) F: 1.2 (0.7)			
	LDL-C			Cal.				M: 6.1 (1.2) F: 6.0 (1.1)			
7 Chinese cohorts (1987)[113]	TC TG	F	S	Enz.	China	10 825 (55)	55 (10)	5.1 (1.4) 1.5 (0.8)	8.4	216	CT/MRI (63%)
JPHC (1990)	TC	NF	S	Enz.	Japan	33 469 (65)	54	4.9	12	207	MR + DC
CMPCS (1992)[115, 152, 153]	TC	NA	S	Enz.	China	30 384 (NA)	47	4.7	10	128	MR + DC
	HDL-C	NA		Enz.		30 384 (NA)	47	M: 1.4 (0.4) F: 1.5 (0.4)	10	159	
	non-HDL-C	F		Cal		29 937 (47)	47	M: 3.3 (1.1) F: 3.4 (1.2)	8	420	
IPHS (1993)[126]	TC HDL-C TG	NF	S	Enz.	Japan	91 219 (66)	59	5.2 1.4 1.4	10	264 F	DC
	LDL-C			Cal.				3.0			
The Kailuan Study (2006)[112]	non-HDL-C	F	S	Enz.	China	95 916 (20)	51	3.4	4	416	MR

Enz.= Enzymatic methods; Cal.=Calculated; Col.= Colorimetric; Gel.= Gel-permeation chromatography; M=Men; W=Women; Individual studies that have been included in identified individual-participant-based meta-analyses (Table 2.2) were not reviewed separately. Other annotations are the same as Table 2.2, unless otherwise stated.

Table 2.5: Key findings of the meta-analyses examining the associations of major blood lipids with stroke, by pathological type

Study (Year published)	No. of studies	Lipids	Lipid levels, mean (SD), mmol/L	HR (95% CI)	Adjustment*	RDR	Variables investigated for effect modification
IPD meta-analysis							
ERFC (2007)	IS: 32	HDL-C Non-HDL-C TG	NA	0.83 (0.64-1.05) per 1 mmol/L higher 1.11 (1.04-1.18) per 1 mmol/L higher 1.02 (0.94-1.11) per SD (68%) higher	++++++	Yes	None
	HS: 19	HDL-C Non-HDL-C TG	NA	0.98 (0.84-1.15) per 1 mmol/L higher 1.08 (0.93-1.26) per 1 mmol/L higher 1.04 (0.82-1.32) per SD (68%) higher			
PSC (2007)	IS: 61	TC	5.8 (1.3)	Q1: 1.00 (0.85-1.18); Q2: 1.04 (0.94-1.15) Q3: 1.02 (0.94-1.10); Q4: 1.14 (1.05-1.23)	+	Yes	Sex, age, SBP, BMI, and smoking
	HS: 61	TC	5.8 (1.3)	Q1: 1.29 (1.13-1.48); Q2: 1.00 (0.92-1.09) Q3: 0.93 (0.87-0.99); Q4: 0.93 (0.87-1.00)			
APCSC-1 (2003)	IS: 29 HS: 29	TC TC	5.0 (1.1)	1.25 (1.13-1.40) per 1 mmol/L higher 0.80 (0.70-0.92) per 1 mmol/L higher	++++	Yes	None
APCSC-2 (2004)	IS: 26	TG	Asia, 1.5 (1.0) Australasia, 1.4 (1.2)	1.35 (1.00-1.83) per 1 SD higher of $\log_e$ TG	++++	Yes	None
	HS: 26	TG	Asia, 1.5 (1.0) Australasia, 1.4 (1.1)	Not significant			
APCSC-3 (2007)	IS: 25 HS: 25	HDL-C	1.4 (0.4) 1.4 (0.4)	0.76 (0.48-1.19) per 1 mmol/L higher 0.74 (0.46-1.19) per 1 mmol/L higher	+	Yes	Region, sex, and age
APCSC-4 (2013)[125]	IS: 44	TC	Asia 4.8 (0.9) Australasia 5.6 (1.1)	1.25 (1.13-1.38) per 1 mmol/L higher;	+++++	Yes	Region
	HS: 44	TC	Asia 4.8 (0.9) Australasia 5.6 (1.1)	0.73 (0.64-0.84) per 1 mmol/L higher			
WPP (2002)[149]	IS: 8	TC	5.8	<55 years: 1.23 (1.02-1.49) per 1 mmol/L higher ≥55 years: 0.95 (0.87-1.04) per 1 mmol/L higher	+++++	No	None
EUROSTROKE (2002)[121]	IS: 4	TC	6.2 (1.3)	1.05 (0.93-1.18) per 1 mmol/L higher	++	No	None
		HDL-C	1.3 (0.4)	M: 0.65 (0.35-1.19) per 1 mmol/L higher W: 2.52 (1.15-5.50) per 1 mmol/L higher	++++		Sex
EPOCH-JAPAN (2012)[127]	IS: 10	TC	M: 5.0 (0.9) W: 5.3 (1.0)	Per 1 mmol/L higher: M: 40-69 years 0.92 (0.74-1.14); 70-89 years 1.04 (0.87-1.25) W: 40-69 years 1.08 (0.83-1.39) ; 70-89 years 0.97 (0.80-1.16)	+++++	No	Sex, and age
	HS: 10	TC	M: 5.0 (0.9) W: 5.3 (1.0)	0.84 (0.72-0.97) per 1 mmol/L higher			None
JALS-ECC (2010)[150]	IS: 10	TC Non-HDL-C	5.2 (0.9) 3.7 (1.0)	1.03 (0.89-1.20) per 1 mmol/L higher 1.07 (0.94-1.23) per 1 mmol/L higher	+++++	No	None
Pub meta-analysis							
Wang X (2013)[151]	HS: 23	TC HDL-C LDL-C	NA	TC 0.85 (0.80-0.91) per 1 mmol/L higher HDL-C 1.11 (0.99-1.25) per 1 mmol/L higher LDL-C 0.90 (0.77-1.05) per 1 mmol/L higher	Varied	Varied	NA

Conventions as in Table 2.3, unless otherwise stated in the above table. * Levels of adjustments for confounders in the above studies were as follows:
- No adjustment; + adjusted for age and sex; ++ plus blood pressure; +++ plus diabetes; ++++ plus other cardiovascular risk factors (e.g., smoking, alcohol consumption, and physical activity); +++++ plus body mass index or other adiposity measurements; ++++++ plus other lipids.

Table 2.6: Key findings of the prospective studies examining the associations of major blood lipids with ischaemic stroke

Study (year screened)	Lipids	Association characteristics			
		HR (95% CI)	Adjustment	RDR	Variable investigated for effect modification
CCHS (1976)[107]	TG	M:1.12 (1.04-1.20) per 1 mmol/L higher W:1.28 (1.15-1.43) per 1 mmol/L higher	++++++	No	Sex
CRCS (1985)[154]	HDL-C	0.77 (0.68-0.87) per SD higher of $\log_e$ HDL-C	+++++	No	None
AMORIS (1986)[116]	TC	Per 1 mmol/L higher: M: Non-fatal events: 1.11 (1.07-1.14) Fatal events: 1.04 (0.92-1.16) W: Non-fatal events: 1.08 (1.04-1.12) Fatal events: 1.03 (0.92-1.16)	+++	No	None
	HDL-C	Per 1 mmol/L higher: M: Non-fatal events: 0.76 (0.68-0.83) Fatal events: 0.95 (0.68-1.31) W: Non-fatal events: 0.66 (0.60-0.74) Fatal events: 0.58 (0.42-0.83)			
	TG	Per 1 mmol/L higher: M: Non-fatal events: 1.18 (1.13-1.23) Fatal events: 0.97 (0.83-1.13) W: Non-fatal events: 0.79 (0.75-0.85) Fatal events: 0.74 (0.61-0.90)			
	LDL-C	Per 1 mmol/L higher: M: Non-fatal events: 1.11 (1.08-1.17) Fatal events: 1.04 (0.91-1.18) W: Non-fatal events: 1.09 (1.05-1.15) Fatal events: 1.06 (0.94-1.21)			
JPHC (1990)	TC	M: 1.14 (1.02-1.28) per 1 mmol/L higher W: 1.05 (0.92-1.21) per 1 mmol/L higher	++++++	No	Sex
FPS (1977-9 & 1986-90)[118]	TC	1.08 (0.96-1.21) per 1 mmol/L higher	+++++	No	None
	HDL-C	0.85 (0.62-1.16) per SD higher of $\log_e$ HDL-C	++++++	No	None
7 Chinese cohorts (1987)	TC	< 4.40 mmol/L reference: 4.40 – 5.40 mmol/L 1.62 (1.15-2.28) ≥ 5.40 mmol/L 1.59 (1.13-2.23)	+++++	No	None
	TG	< 1.10 mmol/L reference: 1.10 – 1.60 mmol/L 1.22 (0.86-1.73) ≥ 1.60mmol/L 1.44 (1.03-2.02)			
CMPCS (1992)[115, 152, 153]	TC	< 3.64 mmol/L as reference: RRs were 1.43 (0.82-2.51), 1.55 (0.91-2.62), 1.59 (0.95-2.67), 2.09 (1.23-3.55), 1.40 (0.76-2.58), 1.93 (0.99-3.75), 2.59 (1.29-5.23), and 1.97 (0.91-4.29) for each 0.5 mmol/L higher	++++++	No	None
	HDL-C	≥ 1.56 mmol/L as reference: < 1.00 mmol/L 1.69 (1.18-2.43) 1.00-1.03 mmol/L 1.74 (1.25-2.41) 1.04-1.55 mmol/L 1.24 (0.97-1.59)			
	non-HDL-C	< 3.37 mmol/L as reference: 3.37-4.13 mmol/L 1.34 (1.07-1.68) 4.14-4.91 mmol/L 1.38 (1.04-1.83) ≥ 4.92 mmol/L 1.38 (0.97-1.94)	+++++	No	None
IPHS (1993)[126]	LDL-C	< 2.06 mmol/L as reference: 2.06-2.57 mmol/L 0.75 (0.55-1.02) 2.58-3.09 mmol/L 0.77 (0.58-1.04) 3.10-3.61 mmol/L 0.75 (0.55-1.02) ≥ 3.62 mmol/L 0.85 (0.62-1.15)	+++++	No	None
CCHS (1991-1994) +CGPS (2003) [119]	HDL-C	"U-shaped" with lowest point around 1.9 mmol/L	++++++	Yes	Sex
The Three-City Study (1999)[155]	TG	1.13 (0.88 - 1.47) per 1 mmol/L higher	+++++	No	None
The Kailuan Study (2006)	non-HDL-C	1.20 (1.10 - 1.33) per 1 mmol/L higher	++++++	No	Sex, age, diabetes, smoking, hypertension, and BMI.

Conventions as in Table 2.5, unless otherwise stated in the above table.

Table 2.7: Key findings of the prospective studies examining the associations of major blood lipids with haemorrhagic stroke

Study (Year screened)	Lipids	Association characteristics			
		HR (95% CI)	Adjustment	RDR	Variable investigated for effect modification
KPMCP (1978)[128]	TC	M: 40-64 years: 0.71 (0.33-1.54) 65-89 years: 2.72 (1.46-5.02) W:40-64 years: 1.04 (0.45-2.40) 65-89 years: 2.23 (0.90-5.54)	+++++	No	Age, sex, and hypertension
KNHS (1986)[111]	TC	0.91 (0.87-0.95) per 1 mmol/L higher	+++++	No	SBP and GGT
AMORIS (1986)[116]	TC	Per 1 mmol/L higher: M: Non-fatal events: 0.94 (0.85-1.03) Fatal events: 0.93 (0.79-1.07) W: Non-fatal events: 0.88 (0.79-1.00) Fatal events: 0.97 (0.83-1.12)	+++	No	Sex
	HDL-C	Per 1 mmol/L higher: M: Non-fatal events: 0.92 (0.72-1.19) Fatal events: 1.08 (0.74-1.60); W: Non-fatal events: 0.86 (0.61-1.18) Fatal events: 0.57 (0.37-0.88)			
	TG	Per 1 mmol/L higher: M: Non-fatal events: 0.96 (0.86-1.08) Fatal events: 1.03 (0.87-1.25) W: Non-fatal events: 0.92 (0.75-1.10) Fatal events: 0.73 (0.56-0.93)			
	LDL-C	Per 1 mmol/L higher: M: Non-fatal events: 0.97 (0.88-1.07) Fatal events: 0.92 (0.80-1.8) W: Non-fatal events: 0.91 (0.80-1.04) Fatal events: 1.04 (0.89-1.23)			
JPHC (1990)[114]	TC	M: 1.05 (0.90-1.23) per 1 mmol/L higher W: 0.93 (0.81-1.07) per 1 mmol/L higher	++++++	No	Sex
IPHS (1993)[126]	TC	< 4.15 mmol/L as reference: 4.15-4.65 mmol/L 0.97 (0.65-1.46) 4.66-5.16 mmol/L 0.63 (0.42-0.96) 5.17-5.67 mmol/L 0.62 (0.40-0.95) 5.68-6.18 mmol/L 0.59 (0.37-0.95) ≥ 6.19 mmol/L 0.55 (0.33-0.91)	+++++	No	None
	HDL-C	< 1,04 mmol/L as reference: 1.04-1.29 mmol/L 0.71 (0.47-1.06) 1.30-1.54 mmol/L 0.68 (0.46-1.03) 1.55-1.79 mmol/L 0.99 (0.65-1.51) ≥ 1.79 mmol/L 0.98 (0.62-1.53)			
	TG	< 1.13 mmol/L as reference: 1.13-1.68 mmol/L 1.06 (0.79-1.42) 1.69-2.25 mmol/L 0.76 (0.51-1.14) 2.26-2.81 mmol/L 0.77 (0.46-1.29) 2.82-3.38 mmol/L 0.81 (0.41-1.64) ≥ 3.39 mmol/L 1.13 (0.63-2.02)			
	LDL-C	< 2.06 as reference: 2.06-2.57 mmol/L 0.65 (0.44-0.96) 2.58-3.09 mmol/L 0.48 (0.32-0.71) 3.10-3.61 mmol/L 0.50 (0.33-0.75) ≥ 3.62 mmol/L 0.45 (0.30-0.69)			
The Kailuan Study (2006)[112]	non-HDL-C	1.06 (0.92-1.20) per 1 mmol/L	++++++	NO	None
CMPCS (1992)[115, 152, 153]	TC	< 3.64 mmol/L as reference: RRs were 0.75 (0.34-1.64), 0.89 (0.44-1.80), 1.10 (0.57-2.15), 1.16 (0.57-2.35), 1.18 (0.54-2.60), 0.71 (0.23-2.22), 1.97 (0.77-5.05), and 0.60 (0.13-2.71) for each 0.5 mmol/L higher	++++++	No	None
	HDL-C	≥ 1.56 mmol/L as reference: < 1.00 mmol/L 1.00 (0.51-1.99) 1.00-1.03 mmol/L 1.38 (0.79-2.42) 1.04-1.55 mmol/L 1.25 (0.85-1.85)			
	non-HDL-C	< 3.37 mmol/L as reference; 3.37-4.13 mmol/L 1.49 (1.03-2.18) 4.14-4.91 mmol/L 1.69 (1.08-2.63) ≥ 4.92 mmol/L 1.28 (0.71-2.31)	+++++	No	None
7 Chinese cohorts (1987)[113]	TC	< 4.40 mmol/L reference: 4.40-5.40 mmol/L 0.88 (0.64-1.22) ≥ 5.40 mmol/L 0.68 (0.48-0.95)	+++++	No	None
	TG	< 1.10 mmol/L reference: 1.10-1.60 mmol/L 0.91 (0.65-1.27) ≥1.60mmol/L 0.98 (0.70-1.38)			

Conventions as in Table 2.5, unless otherwise stated in the above table.

Chapter 3 Overall study design and methods

Data from the CKB prospective study were used to assess the relevance of major blood lipids for stroke pathological types using both (**i**) conventional observational studies of the associations of major blood lipids with risk of stroke pathological types; and (**ii**) MR studies of genetic variants encoding blood lipids with risk of strokes. Details of the design and population characteristics of the CKB study have been previously reported elsewhere.[156, 157] This chapter provides a summary of the key elements of the study design, data collection, and the baseline characteristics of study population relevant to this thesis. Additional details of the methodology for both (**i**) and (**ii**) are reported in subsequent chapters, where relevant.

3.1 Background

Over the last few decades, China has undergone a rapid epidemiological transition in the main disease profiles, with substantial decreases in maternal and child mortality and adult mortality from infectious diseases, and a moderate increase in non-communicable chronic diseases (e.g., stroke, IHD, cancer, or chronic lung disease).[158-160] Compared with Western populations, Chinese populations differ in distributions of certain risk factors, and patterns of diseases, due perhaps to differences in lifestyle or environmental exposures, or genetic background.[3, 7] For example, Chinese populations have relatively lower LDL-C concentrations,[75, 160, 161] and higher rates of stroke, compared to white populations of European ancestry. Even within China, there were large unexplained variation in rates of certain diseases, suggesting the potential of certain avoidable causes.[162] Several previous very large studies have studied the causes of chronic

diseases in China,[75, 158, 160, 163-165] but these studies have been constrained by retrospective study design, inadequate sample sizes, restriction to specific study populations or had relatively limited data available on exposures and outcomes, particularly blood and other bio-samples. In order to address this, the CKB prospective study of over 0.5 million Chinese adults was established in 2004 by the University of Oxford in the United Kingdom, in collaboration with the Chinese Centre for Disease Control and Prevention in China, which combined a large sample size with a wide range of data on exposures and outcomes to improve our understanding, prevention, and treatment of chronic diseases in middle and old age.

The CKB is an open-ended, blood-based, and general population-based prospective study of the environmental and genetic risk factors for chronic diseases in China. With the large sample size, and the wide range of exposures collected, the CKB aims to assess reliably the effects of both established and emerging risk factors for many different diseases. Moreover, with the long-term storage of blood samples from a large number of adult Chinese, the CKB study allows reliable assessment of the relevance of many genetic and other factors for various common chronic diseases. The CKB prospective study initially uses a “nested” case-control approach for blood assay analyses, whereby, when sufficient number of cases accrues during follow-up, their stored blood will be retrieved and compared with those who have not developed relevant diseases. To avoid selecting as controls any individuals who later become cases, in selecting controls, the cases were ordered by the reverse of the dates on which they developed the disease of interest (i.e., starting with the most recent, and working backwards to the earliest cases). Proceeding in this order, for each case the

appropriate number of suitably matched controls can then be selected from the survivors who were still being followed when that case developed (and who had not already been selected as cases or controls).[166] The nested case-control approach also allows many factors, without being necessary to know at the onset but will be of chief interest many years thereafter, to be studied in relation to many different diseases at relatively low cost. This will, however, eventually turn into a blood assay measurement of the whole study population in the future.

Moreover, the CKB study will not only provide reliable evidence for prevention and treatment of diseases among Chinese people, who account for one-fifth of the world's population, but also, provide complementary evidence and scope for both primary and secondary prevention of certain diseases in Western populations, and permit comparisons with other large "biobanks" prospective studies, such as the UK Biobank,[167-169] the Million Women Study,[170, 171] and the Mexico City Prospective Study.[172]

3.2 Overall study design and study population

3.2.1 Overall study design

The CKB is a large, blood-based prospective study of over half a million Chinese adults (age range: 30-79 years), that was established in 2004, as a collaboration between the Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU) in the University of Oxford, and the China National Centre for Disease Control and Prevention (Sino-CDC).[156, 157] The aim of CKB study was to assess the effects of established and novel risk factors collected either by questionnaire, physical measurements or blood samples collected at recruitment for cause-specific morbidity and mortality over the subsequent follow-up period. In addition, every

4-5 years, approximately 5% of randomly selected individuals from the overall study population were invited to resurveys, which included repeat data collection of similar items from the baseline survey to correct for regression dilution bias, and collection of additional information to address particular research questions.

Ethical approval for the China Kadoorie Biobank study was obtained from the Oxford Tropical Research Ethics Committee at the University of Oxford (approval number 025-04, 3rd February 2005) and the Chinese Centre for Disease Control and Prevention (CDC) Ethical Review Committee (approval number 005/2004, 9th July 2004). Approval was also granted by institutional research boards at the CDCs in each of the survey sites.

3.2.2 Overall study population

The baseline survey took place between 2004 and 2008 in 10 regions across China (**Figure 3.1**), with five urban and five rural areas, based on standard governmental definitions. The study regions were carefully selected to provide diversity in disease patterns, exposures, cover individuals of a range of social-economic status, and demographic characteristics, as well as logic considerations, including population stability, quality of death and disease registries, and local capacity (e.g., shipment of blood sample and internet accessibility) and long-term commitment. In each area, all permanent residents aged 35-74 years from 100 to 150 administrative units (i.e., rural villages or urban residential committees) were identified through official residential records, then invited to participate by letter (accompanied by study information leaflets) following extensive publicity campaigns. Individuals who were aged slightly outside the target age range (35-74

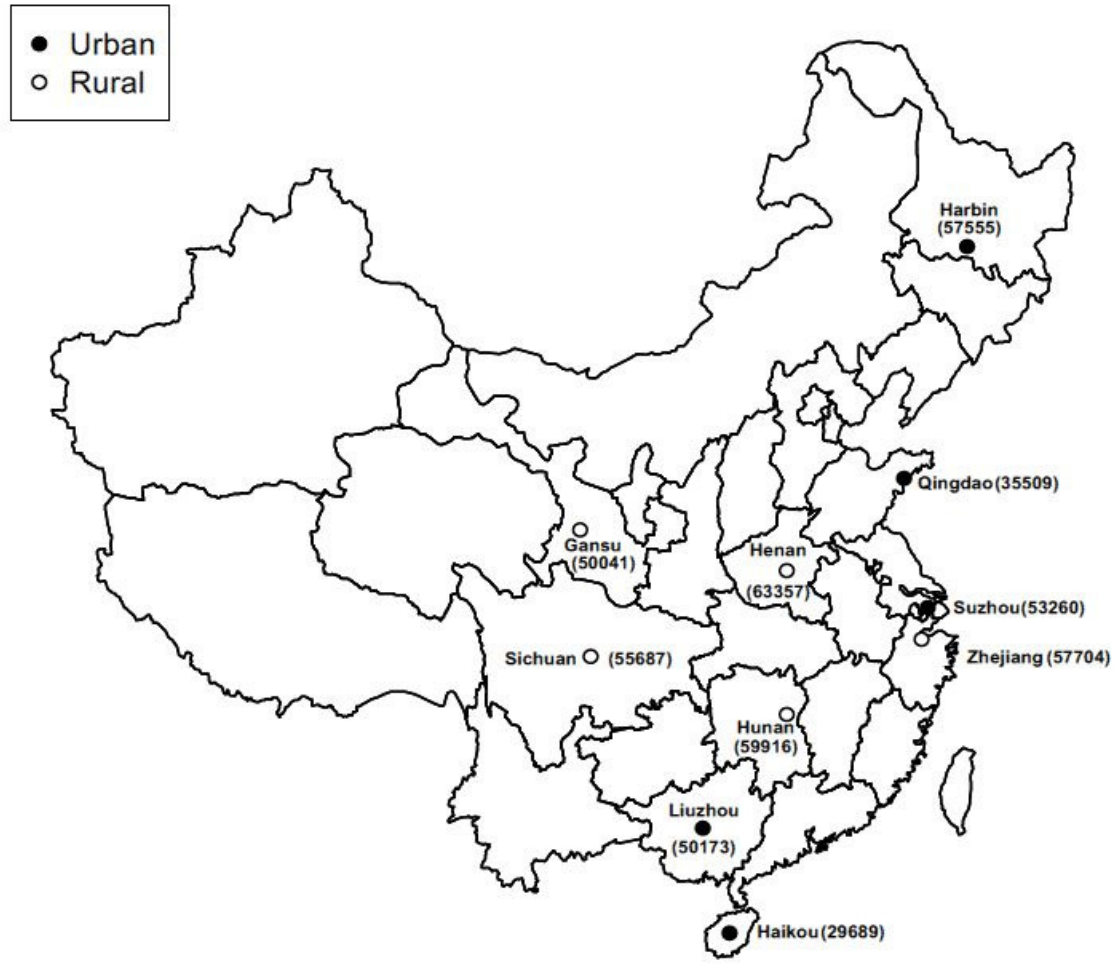


Figure 3.1: Locations of the CKB study sites and number of individuals recruited at the baseline survey (2004-2008)

years) were also recruited into the study to improve local participation. The estimated overall population response rate for the study was approximately 30%, ranging between 26% and 38% in rural, and 16% and 50% urban areas. Overall, 512 891 individuals were recruited during the baseline survey to the CKB prospective study. The study was not designed to be representative of the general adult Chinese population. Indeed, the study was planned to include several particular regions throughout China, with quite different disease and exposure profiles. Consequently, the participants, chiefly defined by exposure or outcome, were selected to give valid causal inferences for the eligible groups they represent (i.e., internal validity); and thus they do not need to be 'representative' of a source

population where they were sampled from.[173, 174] The very large sample size and the diversity of the population covered (e.g., urban and rural, men and women, young and old, and individuals of different socioeconomic status, geographical regions) would not be expected to bias the estimated effect sizes, and the scientific inferences.[174]

3.3 Data collection and procedures

3.3.1 Baseline survey (2004-2008)

During the baseline survey (2004-2008), study assessment centres were set up in local communities (i.e., urban communities or rural villages), with around 15 full-time staff with relevant medical qualifications, and fieldwork experiences. Eligible residents within each site (rural village or urban community) were invited to participate in the CKB baseline study via linkage through their unique national identity cards. All participants provided written informed consent prior to participation, which included permissions for long-term storage of their blood samples for anonymised and non-specified medical research, and access to their medical records.

The baseline assessment took approximately 60-75 minutes to complete the entire data collection process, on a daily recruitment rate of 70-80 per site. In each clinic site, an interviewer-administered, laptop-based electronic questionnaire was used to collect information on demographic and socioeconomic characteristics, lifestyle (e.g., diet, smoking, alcohol, and physical activity), personal and family medical history, and reproductive history (for women). Physical measurements including height, weight, hip and waist circumference, lung function, heart rate, and blood pressure were recorded using standard protocols. [156, 157, 175, 176] Individuals

reporting a history of doctor-diagnosed CHD, stroke, and hypertension, were also asked about their current medications (e.g., statins).

At baseline, a single 10 ml non-fasting blood sample, (except for the Zhejiang region – where participants were asked to fast to allow for specific, locally arranged test) was collected from participants into an ethylenediaminetetraacetic acid (EDTA) vacutainer (BD Hemogard^{TM}, BD, Franklin Lakes, New Jersey, USA), and the time since the last meal was also recorded. Immediate, on-site measurement of plasma glucose concentrations and hepatitis B surface antigen were undertaken. The remaining filled EDTA vacutainers were stored at 0-4℃ in containers with ice packs for a few hours before being transported to the local study laboratory where they were immediately centrifuged and aliquoted into four cryovials (3 plasma and 1 DNA containing buffy coat). These cryovials were stored at -40℃ for 3-4 months before being transported on dry ice to Beijing where they were stored at -80℃. Two cryovials of the plasma from each participant were subsequently transported on dry ice to Oxford where they were placed in long-term storage in liquid nitrogen tanks, with the remaining aliquots were retained at -80℃ in Beijing. Some of the samples in Beijing were then transferred to liquid nitrogen tank in 2014.

After a few weeks of the initial baseline survey in each study site, a random sample of approximately 3% of study participants was invited to participate in a quality control survey within a few weeks of the initial baseline survey. This involved a replication of the baseline survey and collection of some repeat measurements, but blood samples were not collected. Additional quality control was achieved through regular central monitoring of recruitment, collected data and

blood processing time, and through six monthly and annual active site visits by regional, national, and international staff, respectively.

3.3.2 Resurveys

On a five-year basis, an approximately 5% of surviving participants, with the administrative unit (i.e., rural village or urban residential committee) as the basic sampling unit, were invited to participate in resurveys to repeat the collection of identical information to that at the baseline survey together with selected additional information. These resurveys enabled an assessment of "regression dilution bias" in certain risk factors,[177, 178] of consistency of data, and accommodation of additional information. Two resurveys have been completed, with the first resurvey undertaken in 2008, and the second resurvey between 2013 and 2014. In the first resurvey, 24 760 participants were invited, and a response rate of 80% was achieved (n=19 788). Of this, non-fasting blood samples were collected, except Zhejiang, where participants were asked to fast prior to attending the assessments. In the second resurvey, 33 656 were invited with a response rate of 75% (n=25 242). Non-fasting blood samples and urine samples from each participant were collected across all regions in the second resurvey. In addition, all the participants of the second resurvey also had on-site blood lipid measurements (including TC, HDL-C, triglycerides), which were measured using *Mission*® 3-in-1 Cholesterol monitoring system (*ACON* Laboratories, Inc. China), and the LDL-C was calculated automatically via the machine using the Friedewald equation.[179]

3.3.3 Long-term follow up for mortality and morbidity

In the CKB study, study participants were followed up for cause-specific mortality and morbidity by linkage with routinely collected data from multiple-overlapping

previously established sources in China. All 10 study regions are within China's Disease Surveillance Points (DSP) system, which provides mortality statistics for the whole country.[180] The vital status of the CKB participants was monitored using death certificates reported to the regional CDC, where the deaths were coded using the 10th International Classification of Diseases (ICD-10) by trained physicians. In the CKB study, annual active confirmation of vital status was also conducted by reviewing of residential records or by visits to local communities, to minimise the under-reporting of deaths, and to identify the CKB participants who have moved out of the study regions. Any additional deaths which were not recorded in routine procedures, for example, in remote rural areas of China, medical attention was not readily accessible (<5%), the causes were sought by reviewing medical notes or verbal autopsy using a validated instrument.[181]

Information on incidence of several main chronic diseases (i.e., stroke, IHD, cancer, and diabetes) was collected using established disease registries that were available in 8 study regions (Harbin, Suzhou, Liuzhou, Sichuan, Gansu, Henan, Zhejiang, and Hunan). One additional study area (Qingdao) only had access to cancer registry data. However, detailed information on ICD-10 coded hospitalizations, examinations and treatment procedures were obtained through linkage to established national health insurance claim database covering almost all of the CKB participants (~98%). Notably, since the middle 1990s, brain-scanning technologies have been used routinely in China for stroke diagnoses of those who admitted to hospitals, allowing further assessment of stroke pathological types. Medical records of participants with cancer, IHD, and stroke events recorded in any of the three follow-up data sources – the death registries, CDC disease reporting system, and the national health insurance claim database

– were reviewed by clinicians in the CKB national co-ordinating centre (i.e., NCC) and international co-ordinating centre (i.e., ICC), who were blinded to both the baseline and the resurvey data, in order to confirm, or in certain conditions, enable further phenotyping.

In addition to the routinely collected data on prospective cause-specific mortality and morbidity, adjudication exercises were conducted to assess the accuracy of the main disease outcomes of interest within the CKB (e.g., stroke, IHD and cancer), with the intention of extending validation to a wider range of other diseases or health-related outcomes. A portable validation device was used to verify the reported diagnoses of stroke and IHD by checking and recording details of presenting symptoms and signs, key diagnostic test results and discharge letters from the medical records. Subsequently, the information collected from these sources was used to adjudicate stroke pathological types by neurologists and of IHD sub-types by cardiologists, respectively. Hence, all verified stroke events were then classified into the relevant stroke pathological types and all verified IHD events were classified into relevant sub-types. The stroke adjudication exercise is still on-going and expected to complete later 2017 for the disease events occurred before 1st January 2015. The preliminary results have suggested a relatively high level of accuracy of confirmation of reported diagnoses of stroke pathological types in the CKB study (Data were unavailable for the present thesis).

3.3.4 Nested case-control study of stroke

After around 10-years' follow-up, up to a censoring date (1st January 2014), there were around 27 000 brain-imaging confirmed incident IS cases, and 6000 incident ICH cases recorded. Among those cases without reporting statin treatment at

baseline, approximately 5000 incident IS, 5000 incident ICH, and 6000 healthy controls were selected retrospectively in time for the nested case-control study of stroke.

In CKB, the shared controls were selected chiefly to match ICH cases by region, sex, and year of birth. In selecting controls, the ICH cases were ordered according to date of event (i.e., starting with the most recent event, and working backwards to the earliest case). For individuals with ICH events on the same day, rank oldest first. Proceeding in this order, for each ICH case, identify the set of possible controls matching all selection criteria, including matching region, year of birth and sex. If <2 possible controls were identified, expand year of birth selection by ±1 year. Repeat as necessary until ≥2 possible controls were identified. Details on the selection of cases of two major stroke pathological types, and controls will be reported separately in **Chapters 4-5**.

Biochemistry measurements

In the nested case-control study of stroke types, a group of previously identified blood biomarkers (e.g., lipids, inflammatory and kidney function biomarkers, homocysteine, Vitamin D, etc) for vascular disease were assayed using 1 ml baseline plasma samples, and for a representative proportion of individuals collected at the resurvey. Plasma biochemistry measurements were conducted at the CTSU laboratory, which has been accredited (International Organisation for Standardisation 17025:2005) for many of the routine tests. Lipid-related biomarkers were measured on biochemistry analyser Beckman Coulter AU680 (Beckman Coulter Inc., UK), in addition to several other previously established

biomarkers for vascular disease. **Table 3.1** reports a summary of the methods used to measure the lipid-related biomarkers currently available in the CKB study.

In both clinical practice and epidemiological studies, the most common approach to determine LDL-C is the Friedewald equation, which calculates LDL-C from measurements of TC, triglycerides, and HDL-C in a fasting state. Although convenient, the Friedewald calculation[179] has several well-established limitations (e.g., requiring fasting blood samples and also for triglycerides to have concentrations <4.5 mmol/L), which led to the development of methods for direct LDL-C measurement.[182] Approaches of directly measuring LDL-C have progressed from precipitation methods,[183] immuno-separation methods,[184] to the homogeneous assays,[185] of which, direct homogeneous assay has been recommended as the most accurate biochemistry measurement for LDL-C among general populations who have no prior conditions. Among individuals being treated for cardiovascular disease or who had other conditions that might be expected to affect lipoprotein compositions (such as anacetrapib), alternative methods have been recommended (e.g., beta-quantification reference method), due to the potential underestimation of LDL-C via detergent-based methods, or Friedewald equation calculations.[186, 187]

DNA extraction and genotyping

Individuals included in the nested case-control study of stroke were genotyped using an Affymetrix Axiom® 700 000 SNP array, which was similar to the genotyping approach used in the UK Biobank study,[167] facilitating future cross-validation between these studies. The complete GWAS data will be released in later 2017, and thus this thesis used the released SNP data that has been

prioritised to a panel including 384 SNPs, of the entire GWAS assay. SNPs included in the panel were selected on the basis of previous associations with chronic diseases (e.g., stroke, IHD, cancer), intermediate phenotypes (e.g., lipids, blood pressure or BMI), metabolic pathways (e.g., vitamin D, homocysteine), and risk exposures (e.g., smoking and alcohol) with priority given to SNPs: demonstrated associations with the outcomes of interest in Chinese or other East Asian populations, particularly detected in large-scale GWAS; demonstrated associations with the outcomes of interest in other populations, with conserved LD across populations; functional variants (e.g., non-synonymous substitutions); population-specific haplotypes at loci of specific interest; and perceived biological significance.

3.4 Baseline characteristics of the study population

3.4.1 The study population

Demographic and lifestyle characteristics

Overall, 512 891 participants (210 259 men and 302 632 women) were recruited to the study in 2004-2008, with 55.9% from rural areas. Across the 10 study regions, the number of recruited participants ranged from 29 689 (Haikou) to 63 357 (Henan). At baseline, the mean (SD) age was 52.4 (10.9) years in men, and 51.0 (10.4) years in women, directly standardised to the region structure of the CKB study population (**Table 3.2**). Proportions of participants who completed high school or above were around 26.1% in men and 17.5% in women, respectively.

The baseline distributions of most lifestyle risk factors were largely similar between men and women, apart from smoking, and alcohol consumption, where there were striking sex-differences observed. The prevalence of ever-regular smoking in men and women were 74.2% and 3.3%, respectively. Likewise, 33.4% of men and 2.1% of women reported weekly alcohol consumption. Regular consumption (defined as consumption on at least 4-6 days per week) of fresh vegetables was almost universal. Approximately 30% of the study participants reported consumption of fresh fruit on a daily basis or at least 4 to 6 days per week. Women were more likely to consume fresh fruit frequently than men (31.3% versus 23.4%). Around half of participants reported regular consumption of meat or poultry, and around 10% of the participants reported regular consumption of soybean, fish or seafood, dairy products, and fresh eggs. Overall, proportions of participant regularly consuming different food varied between regions, but with relatively limited differences between sexes. The mean (SD) levels of daily total physical activity in men and women were 22.0 (14.2) MET hours /day and 20.4 (11.6) MET hours /day, respectively.

Prevalent diseases

Among the participants recruited in the CKB prospective study, hypertension was the most commonly reported doctor-diagnosed chronic condition (men: 10.8%; women 12.3%), with markedly lower (generally <3%) self-reported prevalence of most other conditions at baseline (**Table 3.2**). The prevalence of self-reported prevalent conditions also varied substantially between regions.

Anthropometric characteristics

Table 3.3 shows the baseline anthropometric characteristics of the study population. Mean (SD) age- and study region-standardised SBP levels were slightly higher in men than in women (132.3 [18.8] versus 130.5 [20.2] mm Hg). BMI was similar between men and women (23.5 [3.1] versus 23.8 [3.3] kg/m^2). Men had a slightly higher mean (SD) waist circumference than women (82.1 [9.2] versus 79.1 [9.7] cm), but lower mean (SD) hip circumference (90.7 [5.9] versus 91.1 [6.3] cm). Men had a lower body fat percentage than women (22.1 [5.9] versus 32.1 [6.9] %). The mean (SD) random blood glucose concentrations in the entire population was 6.1 (2.3) mmol/L, and was similar in men and women (6.0 [2.3] versus 6.2 [2.3] mmol/L).

3.4.2 The resurvey participants

Table 3.4 shows the selected baseline characteristics of the overall CKB study population of 512 891 participants, the first resurvey of 19 788 participants, and the second resurvey of 25 242 participants. The distributions of demographic characteristics, cardiovascular risk factors, and the prevalent self-reported disease were broadly similar across the three study populations, despite slight differences in some characteristics. The proportions of women were slightly higher in the resurvey populations (61% in the first resurvey and 62% in the second resurvey) compared with overall population recruited at baseline (59%). The proportion of individuals reporting prevalent chronic conditions (e.g., CHD, stroke or TIA, diabetes, and hypertension) in the first resurvey population was similar to those observed in the overall study population, while in the second resurvey there were

slightly higher proportions of stroke or TIA (1.7%, 1.6%, and 2.9%, respectively, in the overall study population, the first resurvey, and the second resurvey).

3.5 Summary

The comparatively large-scale of study participants recruited, the prospective study design, and the uniquely large number of stroke cases identified through multiple overlapping data sources with brain imaging information, illustrate the importance of the CKB study as a unique resource for stroke epidemiology. The collection of baseline blood samples, and resurveys among representative subsets, along with the diversity of the exposure information collected, make it a valuable resource to assess the associations of major blood lipids with stroke pathological types, before and after adjustment for various vascular risk factors. The availability of genetic data will not only help identify those genetic variants acting as correlates or determinants of stroke pathological types, but will also facilitate investigation of the causal relevance of the observational associations, providing further insights into prevention and treatment of stroke in China, and elsewhere.

3.6 Tables

Table 3.1: Biochemical methods used for lipid-related biomarkers

Assay	Method
Total cholesterol	Enzymatic colour test
LDL cholesterol	Direct homogenous method
HDL cholesterol	Direct homogenous method
Triglycerides	Enzymatic colour test
Apolipoprotein A1	Immuno-turbidimetric test
Apolipoprotein B100	Immuno-turbidimetric test

LDL=Low-density lipoprotein; HDL=High-density lipoprotein.

Table 3.2: Selected demographic and lifestyle characteristics of the CKB study participants at baseline survey (2004-2008), by sex

Characteristics*	Men (n = 210 259)	Women (n = 302 632)
Age (years)		
30-59	72.4	77.9
60-69	19.7	16.6
70-79	7.8	5.5
Mean (SD)	52.4 (10.9)	51.1 (10.4)
Marital status, %		
Married with spouse	93.3	88.4
Widowed, separated or divorced	5.3	11.4
Never married	1.5	0.2
Highest level of completed education, %		
No formal education	8.3	26.2
Primary school	32.3	31.6
Secondary school	33.2	24.7
High school or above	26.1	17.5
Annual household income (yuan / year), %		
<2500	2.6	3.4
2500-4999	6.2	7.1
5000-9999	17.1	19.3
10 000-19 999	28.4	29.5
20 000-34 999	25.2	24.3
≥35 000	20.3	16.4
Smoking category, %		
Never	14.4	94.8
Occasional	11.4	1.8
Ex-regular	12.9	0.9
Current regular	61.3	2.4
Alcohol consumption, %		
Never	19.9	63.7
Occasional	43.1	33.8
Ex-regular	3.6	0.5
Current regular	33.4	2.1
Regular consumption of food groups, %		
Fresh vegetables	98.2	98.4
Meat / poultry	53.1	43.6
Fresh eggs	25.3	23.7
Fresh fruit	23.4	31.3
Dairy products	10.8	12.6
Soybean	10.6	9.4
Fish / seafood	9.7	8.3
Physical activity (MET hours / day), %		
<10	24.7	23.1
10-14.9	14.1	22.1
≥15	61.2	54.9
Mean (SD)	22.0 (14.2)	20.4 (11.6)

Table continues on next page

Table 3.2: Selected demographic and lifestyle characteristics of the CKB study participants at baseline survey (2004-2008), by sex

Characteristics*	Men (n = 210 259)	Women (n = 302 632)
Self-reported prevalent diseases, %		
Ischaemic heart disease	2.6	3.3
Stroke / Transient ischaemic attack	2.2	1.4
Diabetes	2.9	3.4
Hypertension	10.8	12.3
Chronic respiratory disease	3.3	2.7
Chronic kidney disease	1.3	1.6
Cancer	0.4	0.5

MET=Metabolic equivalent of task. * Standardised to the age and study area structure of the study population, except for age, where direct standardisation was made for region only.

Table 3.3: Selected anthropometric measurements of the CKB study participants at baseline survey (2004-2008), by sex

Characteristics	Men, % (n = 210 259)	Women, % (n = 302 632)
Systolic blood pressure (mmHg)		
<120	26.9	34.8
120-139	43.6	36.5
≥140	29.5	28.6
Mean (SD)	132.3 (18.8)	130.5 (20.2)
Height (cm)		
<155	5.3	55.5
155-159	14.6	28.0
160-164	27.7	12.9
≥165	52.5	3.5
Mean (SD)	165.3 (5.8)	154.1 (5.3)
Body mass index (kg/m^2)*		
<22.0	35.0	31.2
22.0-25.0	33.8	34.4
≥25.0	31.2	34.3
Mean (SD)	23.5 (3.1)	23.8 (3.3)
Waist circumference (cm)		
<80	43.2	55.1
80-89	34.7	31.4
≥90	22.2	13.5
Mean (SD)	82.1 (9.2)	79.1 (6.3)
Hip circumference (cm)		
<88	35.4	32.8
80-89	28.7	30.1
≥90	35.9	37.1
Mean (SD)	90.7 (5.9)	91.1 (6.3)
Body fat percentage (%)		
<15	12.6	0.4
15-24	56.2	15.2
25-34	28.8	51.7
≥35	2.4	32.7
Mean (SD)	22.1 (5.9)	32.1 (6.9)
Random blood glucose (mmol/L)**		
<5.8	61.4	54.6
5.8-7.7	30.1	36.4
7.8-11.0	5.6	5.9
≥11.1	2.9	3.1
Mean (SD)	6.0 (2.3)	6.2 (2.3)

Standardised to the age and study region structure of the study population. *Data missing for two participants; **Among those with baseline random blood glucose measurements (n=504 550).

Table 3.4: Comparisons of the selected baseline characteristics of the resurvey participants with those of the overall cohort population

Characteristics*	Overall population (n = 512 891)	First resurvey (n = 19 788)	Second resurvey (n = 25 242)
Demographic			
Age (years), mean	51.5 (10.7)	51.6 (10.6)	51.0 (10.1)
Female, %	59.0	60.7	61.7
Urban, %	44.1	41.2	43.1
High school education or above, %	21.0	19.5	18.9
Income >20 000 yuan/year, %	42.7	37.1	37.6
Cardiovascular risk factors			
Ever regular smoker, %	32.4	30.9	30.1
Regular drinker, %	14.9	14.5	14.6
Physical activity (MET hours /day), mean	21.1 (13.9)	20.7 (14.2)	21.4 (14.0)
Systolic blood pressure (mmHg), mean	131.1 (21.3)	132.2 (21.8)	131.0 (20.8)
Body mass index (kg/m^2), mean	23.7 (3.4)	23.7 (3.4)	23.7 (3.4)
Random blood glucose (mmol/L), mean	6.1 (2.3)	5.9 (2.3)	5.9 (2.1)
Self-reported prevalent diseases, %			
Ischaemic heart disease	3.1	3.4	2.8
Stroke / Transient ischaemic attack	1.7	1.6	2.9
Diabetes	3.2	3.1	2.6
Hypertension	11.6	11.3	10.4

MET=Metabolic equivalent of task. *Values are directly standardised to the age and region structure of the CKB population unless otherwise stated.

Chapter 4 Major lipids and risk of ischaemic stroke

This chapter first describes the incidence rates of IS by sex, and region, in the overall CKB population, and then estimates the regression dilution ratios for major lipid fractions among individuals with repeat measurements from the baseline survey, the first and second resurveys. The key objectives of this chapter were to examine the shape and strength of the associations of major blood lipids with risk of incident IS in a nested case-control study of IS, after controlling for confounding factors, and correction for regression dilution bias in lipid measurements.

4.1 Introduction

Stroke, including both IS and HS, is the second leading cause of premature death, and the third leading cause of disability worldwide.[1, 3] About 70-80% of all stroke events were attributable to ischaemia, and in 2015, there were about 5.4 million incident IS events, 3.0 million deaths from IS, and 45.2 million DALYs lost due to IS, with the majority of the burden of stroke occurring in LMICs.[6, 41, 45]

Several major risk factors for IS have been reliably characterised,[12] with relatively limited evidence on the shape and strength of the associations of major blood lipids, including LDL-C, HDL-C, and triglycerides, with risk of IS. Nevertheless, previous studies of major blood lipids with IS were chiefly conducted in Western populations,[13, 14] and less is known about the relevance of major blood lipids for risk of IS in other populations, where the rates of IS are high, and the lipid distributions also differ from those in Western populations.

Therefore, in this chapter, the entire 0.5 million CKB prospective study participants were used describe the rates of IS in different settings. The nested case-control

study of IS within the CKB prospective study, including about 5000 incident IS cases and 6000 healthy controls, was used to assess the shape and strength of associations of major blood lipids with IS risk, after adjustment for various vascular risk factors, and correction for regression dilution bias.

4.2 Study design and methods

The description of the incidence rates of IS was conducted in the overall CKB participants with no prior self-reported history of cardiovascular diseases (i.e., CHD, stroke or TIA) at recruitment. The overall study population and design of the CKB prospective study have been described in detail elsewhere,[156, 157] and have been summarised in **Chapter 3**. The analyses of prospective associations of major blood lipids with risk of IS were undertaken in a nested case-control study of stroke. The following sections will mainly report on the study design, methods, and participants included in the nested case-control study of stroke.

4.2.1 The nested case-control study

The nested case-control study of stroke within the CKB study was designed to address several research questions, and initially selected approximately 5000 incident ICH cases, 5000 incident IS cases, and 6000 healthy controls. The present chapter describes the details of the nested case-control study of IS, and relevant details about ICH case selection will be presented in **Chapter 5**.

Cases and controls were originally selected censoring up to 1st January 2014, among the CKB participants who: a) were aged between 30-79 years at the baseline survey (2004-2008); b) did not report a doctor-diagnosed CHD, stroke, TIA, and cancer at baseline; and c) did not report use of statins. Following these

criteria, in the overall CKB study (n=489 762), approximately 25 000 incident IS cases were identified. Of individuals with recorded diagnosis of IS and aged below 71 years, 5486 incident IS cases were randomly selected. The stroke event occurred earlier would be counted, if one individual had both IS and ICH recorded on the same day. The same controls were used with the ICH cases (See **Chapter 5**). Briefly, the controls were identified among individuals who were free of stroke of any type, or unspecified type, MI, or CHD by the censoring date, matching to ICH cases by sex, year of birth, and region.

The disease status (i.e., IS case or control) was re-identified (at the time of writing the thesis) extending the censoring date to 1st January 2016, in order to maximise the follow-up time (chiefly for the controls). To minimise bias from reverse causality, participants were excluded who developed MI, SAH, and stroke of unknown types, prior to any IS events. Few previously identified controls developed IS events, while several IS cases were re-classified as controls, due to the initial linkage errors. **Table 4.1** shows the details of any changes in the disease outcome definitions between 2014 and 2016.

4.2.2 Statistical analyses

4.2.2.1 Selection of baseline variables

Biochemistry measurements of plasma concentrations of lipid-related biomarkers were selected, in addition to other variables which may potentially confound any associations of major blood lipids with risk of IS, suggested by the **Literature Review (Chapter 2)**. The following variables were selected: measurements of TC, LDL-C, HDL-C, triglycerides, Apo A1, Apo B100; age, sex, study region, education, household income, smoking status, alcohol consumption, information on dietary

patterns (i.e., regular consumption of fresh vegetables, meat or poultry, fish or other seafood, egg, soybean, fresh fruit, and dairy products), physical activity (MET hours /day), self-reported health conditions (i.e., self-rated health, diabetes, hypertension, and blood pressure lowering medication); measurements of SBP, BMI, waist circumference, hip circumference, body fat percentage, random blood glucose (RBG), and estimated glomerular filtration rate (eGFR, which was calculated based on the Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI][188]).

4.2.2.2 Participant exclusions

Preliminary analyses were conducted to describe the distribution of selected variables and to identify missing, erroneous or outlying values. Bar charts were produced to describe the distribution of categorical variables, whereas histograms and descriptive statistics (mean, SD, median, minimum and maximum values) were used to describe the distribution of numerical variables.

To avoid a few potentially unreliable measurements unduly affecting the results, few individuals who had lipid measurements beyond the 99th percentile were excluded (**Table 4.2**). For logical considerations, participants with non-HDL-C (i.e., TC minus HDL-C) or remnant cholesterol (i.e., total cholesterol minus HDL-C minus LDL-C) less than 0 (or equal to) were excluded as well. In total, there were 83 records excluded due to the sample extraction failure, and 81 records due to potential unreliable measurements. Of the remainder, one individual had baseline BMI measurement missing, and was substituted with the mean BMI (23.7 kg/m^2) of the overall CKB study population. There were 23 individuals with fat percentage

measurements missing, which were substituted with sex-specific median values estimated among the overall CKB population (Men: 21.7%; Women: 31.7%).

Following these exclusions, 16 541 individuals from the baseline survey, and 1776 from the resurveys remained. Among individuals with baseline lipid measurements, there were 5475 IS cases, 4776 ICH cases, and 6290 controls. Among individuals with resurvey samples, 486 were from the first resurvey (2008), and 1290 were from the second resurvey (2013-2014). Overall, 317 individuals had repeat measurements of major blood lipids at the first resurvey, and 858 individuals had repeat measurements at the second resurvey. Participants with paired measurements at the baseline survey, the first resurvey, or the second resurvey were used to estimate the regression dilution ratios (RDRs) of major blood lipids.

4.2.2.3 Definitions and categorisation of baseline variables

For the descriptive analyses of the baseline characteristics of study participants by lipid concentrations, participants were grouped into quintiles of the lipid measurements, approximately. For the prospective associations of major blood lipids with risk of IS, participants were first grouped into quintiles based on their lipid measurements to assess the shape of any associations. To facilitate the comparisons of the strength of associations of different lipid-related biomarkers with risk of IS, the relative risks were then expressed in proportion to 1 SD difference of lipid measurements. Values of triglycerides were taken natural logarithm (Log_e) transformation to improve normality, when assessing the risk of IS associated with percentage changes in plasma triglyceride concentrations.

Age at baseline was divided into six groups: 30-39, 40-49, 50-59, 60-69, and 70-79 years. Education was categorised according to the highest level of education

completed by individuals: no formal education, primary school, secondary school, and high school or further education (e.g., technical college or university).

For dietary variables, the baseline questionnaire collected information about the frequency of usual consumption of certain foods over the last year: never or rarely, monthly, 1-3 days per week, 4-6 days per week, or daily. Based on the overall food consumption patterns in China, consumption of certain foods were re-grouped as a binary variable: regular and non-regular consumption. Regular consumption was defined as consumption on at least 4-6 days per week of this particular food (i.e., meat or poultry, fish or seafood, eggs, soybean, dairy products, fresh fruit, and fresh vegetable).

For alcohol intake, participants were categorised according to their frequency of their usual consumption over the previous years: never drinkers, occasional drinkers (including monthly drinker), ex-regular drinkers (including reduced drinker), or currently regular drinkers. For smoking, participants were categorised as never smokers, occasional smokers, ex-regular smokers, or current smokers.

For baseline RBG, participants were categorised into five groups: ≤4.7, 4.8-5.7, 5.8-7.7, 7.8-11.0, and 11.1+ mmol/L. Consistent with the diagnostic criteria for diabetes recommended by the WHO,[189] participants were classified as diabetes at baseline if they reported a history of physician-diagnosed diabetes in the baseline survey questionnaire, or their RBG test during the baseline survey was ≥11.1 mmol/L, or their fasting blood glucose (taken the following day in those with a raised RBG test [set as ≥7.8 mmol/L in this study]) was ≥7.0 mmol/L.[190]

For baseline SBP, participants were grouped into five groups: ≤124, 125-144, 145-164, and 165+ mm Hg. Prevalent hypertension was defined as self-reported

doctor-diagnosed hypertension. Blood pressure-lowering medication use was defined as use of any standard blood pressure-lowering medication (excluding traditional Chinese medicine), including: angiotensin-converting-enzyme (ACE) inhibitors, beta-blockers, calcium-channel blockers and diuretics.

BMI was calculated by dividing weight in kilograms by the square of their height in metres. All measures of adiposity (BMI, waist circumference, hip circumference, and fat percentage) were categorised into quintiles, appropriately.

Physical activity was expressed in MET hours performed per day, which expresses a person's average level of physical activity over a period of time by combining the intensity and duration of different physical activities into a single measure. MET values are estimates of energy expenditure relative to an individual's rate of energy expenditure at rest (their resting metabolic rate), and as such, the value is independent of a participant's weight. In CKB, participants answered questions on the type and duration of their physical activity at work and leisure over the year prior to the survey. Each activity was assigned a MET value previously devised from physical activity compendium,[191] and this was multiplied by the hours per day on average that the activity was performed. A participant's total MET hours per day was the sum of the MET hours per day of each activity, relating to household, commuting, non-sedentary leisure, and occupational. Participants were categorised into five levels of total physical activity: 0-<10, 10-<20, 20-<30, 30-<40, and 40+ MET hours /day.

For chronic kidney function, eGFR has been widely considered as a better overall index than others (e.g., creatinine concentration) for examining associations with disease. In the present study, individuals with creatinine concentrations greater

than 176 umol/L (beyond the 99th percentile) were excluded. Of the remainder, eGFR was calculated using the CKD-EPI equation.[188]

4.2.2.4 Incidence rates of ischaemic stroke in the CKB participants

Incidence rates of IS were estimated in the overall CKB study population from recruitment until the censoring date of 1st January 2016. The analyses were conducted on 489 762 participants with no prior history of physician-diagnosed cardiovascular disease (i.e., CHD, stroke, or TIA) at the baseline survey. Within the 4.4 million person-years of follow-up, 32 869 incident IS cases were identified. Crude incidence rates were calculated as the ratio of number of first-ever IS cases to the total person-years in that age-at-risk group, and were reported, separately, by sex and region. The overall incidence rates were estimated by combining the unweighted average of incidence rates in each component age-at-risk group.

4.2.2.5 Baseline characteristics of participants in the nested case-control study

The distribution of baseline characteristics of study participants were described using means or proportions by disease status (i.e., IS cases and controls), directly standardised by sex, age (10-year age group), and region of the study population, where appropriate. For triglycerides, median values (inter-quartile range [IQR]) were calculated, rather than using mean (SD) values, due to the right-skewed distribution.

Distribution of baseline characteristics of study participants by lipid measurements were estimated on 11 765 individuals, directly standardised by sex, age (10-year age group), and region of the study population, i.e., weighted by the number of

participants in each stratum. Smoking status and alcohol consumption were re-categorised as binary variables in this section: never versus ever smokers (including occasional, ex-regular, and current smokers), and never versus ever drinkers (including occasional, ex-regular, and current smokers), and were reported, separately, by sex.

4.2.2.6 Cross-sectional correlations of lipid-related biomarkers with other vascular risk factors

Partial Pearson correlation coefficients were used to estimate the correlation between major blood lipid fractions and other continuous covariates on 16 541 individuals, adjusting for disease status (i.e., IS, ICH, and control), age at baseline (continuous), sex, and study region. Correlations of major blood lipid fractions with eGFR and RBG were estimated among a smaller subset of participants with available data on eGFR and RBG (with the number of participants reported in the relevant sections). Equivalent correlational analyses by disease status (i.e., IS cases, ICH cases and controls) showed similar results.

4.2.2.7 Regression dilution ratios of major blood lipids

In prospective epidemiological studies, disease rates during follow-up are typically analysed with respect to the values of risk factors (e.g., blood pressure, LDL-C) measured at an initial baseline survey. This tends to, however, underestimate the real associations of disease risk with the “usual” levels of such risk factors during a particular exposure period (although exceptions are possible),[177] because of the “regression dilution bias”.[192] This is caused by the combined effects of measurement errors, short-term biological variability (including both transient fluctuations and any diurnal or seasonal variation), and longer-term within-person

differences in risk factor levels (which may occur for several reasons, including changes in lifestyle, treatment or disease status). The magnitude of the regression dilution bias (i.e., RDR) associated with a particular exposure period can be estimated either non-parametrically or parametrically using pairs of measurements, which are made in the same individual on different occasions separated by an appropriate time interval (i.e., the length of time from baseline to the midpoint of the exposure period of interest).[177, 192]

The magnitude of the RDR of each lipid fraction in this study was estimated using both non-parametric (i.e., MacMahon-Peto range ratio method) and parametric methods (i.e., self-correlation coefficient). The RDRs between the baseline survey versus the first and the second resurveys, respectively, were estimated separately (because fewer individuals had a re-measurement in the first resurvey than that in the second resurvey). The regression lines, of the type $y=at+b$ (where t indicated time interval [years] between paired measurements, and y denoted the predicted RDR), weighted by the number of participants who contributed data from the baseline survey and the resurveys), were fitted to the non-parametric estimates.

Non-parametric estimates (MacMahon-Peto method)

With the non-parametric method of estimating the RDRs, the baseline measurements were divided into quintiles, and individuals were allocated to a group throughout all the resurveys according to which fifth their baseline measurement fell. The means of the measurements within each of five groups were calculated separately for the baseline, and the first and second resurveys. The absolute differences between top and bottom quintiles were calculated at each resurvey, and this difference was called the “inter-quintile” range for that

survey (e.g., baseline or resurveys). The ratio of the inter-quintile range at the resurvey to that at the baseline was calculated for each lipid fraction within each resurvey and then plotted against their average follow-up time to resurvey after the baseline survey.

To estimate the usual values some years after the baseline survey in each of the five categories of the lipid fraction, the means of the measured values in each of those five categories were shrunk towards the overall mean by the RDR, and the usual value in each of the quintiles is taken to be *M+*RDR*(m-M)* (where M and m denote, respectively, the overall mean and the measured mean of each baseline quintile). Because of this shrinkage, graphs of risk of disease associated with usual lipid values are, appropriately, about 1/RDR times steeper than uncorrected graphs of risk of disease associated with measured lipid values.[67, 177]

Parametric estimates (Self-correlation coefficient)

Pearson's correlation was used to estimate the self-correlation coefficients, ρ_{self}, between the baseline and re-measurements of each lipid fraction, separately for each resurvey. The standard error of the correlation coefficient was calculated as $(1- \rho_{self}^2)/\sqrt{n}$.[178]

The parametric approach should provide similar estimates of the RDRs to those obtained by the non-parametric method when the baseline measurements and re-measurements have similar variance. Estimates based on a self-correlation (using all the data) will have less random variation than those based on a ratio of ranges (using only data from the top and bottom quintiles). Nevertheless, since the non-parametric method requires no assumptions, it may be more reliable to estimate the regression dilution bias over longer periods of follow-up.[177]

In this thesis, RDRs estimated using both methods for each lipid fraction were reported for comparison, and the non-parametric estimates were used to correct the regression dilution bias in the subsequent prospective association analyses. The regression coefficient relating risk to usual levels of a lipid fraction during the exposure period can therefore be estimated as 1/RDR times the "uncorrected" regression coefficient (i.e., log HR), and then taking the exponential. Standard errors for the "corrected" HRs were obtained by dividing the standard errors of the "uncorrected" regression coefficients by the RDR.

4.2.2.8 Prospective associations of major blood lipids with ischaemic stroke

Prospective analyses were conducted on 11 765 participants, following the exclusions described in the previous section. Cox proportional regression was used to estimate HRs and their corresponding 95% CIs for associations between measured lipid values and IS risk, stratified by age-at-risk (within whatever 5-year age-at-risk group was being considered), sex, and study region, and adjusted extensively for other measured risk factors, where appropriate. The Cox model gives a useful statistic for the average HRs over the follow-up period, even if the proportional hazards assumption is violated.[193] Sensitivity analyses using logistic regression gave similar results to the estimates arising from Cox regression (Data not shown).

Shape of the associations

The *shape* of the associations was obtained by plotting the HRs of IS risk in quintile groups defined by the baseline lipid measurements against the "usual" lipid measurements during the exposure period. The resulting HRs were reported as "floating absolute risk",[194] which does not alter their values, but ascribes a

group-specific variance to each of the log risk in that one category. Hence, each HR, including that for the reference group, is associated with a group-specific CI that reflects the amount of data only in that one particular category.[195, 196] The 95% group-specific CI for HR is (HR/T, HR×T), where T=exp (1.96√v) and v is the variance of the log risk. If the reference group with HR=1 and another group with HR=R have, respectively, group-specific CIs (a, b) and (x, y), then the CI for R that allows for the variation in both of the groups (i.e., the co-variance between group) is (√[xy/k], √[xyk]), where log (k) is given by √($\log^2$ [y/x]+$\log^2$ [b/a]); since k>y/x, this CI is wider than (x, y). To determine the slope of an association characterised above, a linear regression line was fitted through the log HRs weighted by the reciprocal of the variance in that particular group.

The prospective associations were investigated using Cox regression in four models, with progressive degrees of adjustments: **(a) Basic adjustment**: associations were stratified by 5-year age-at-risk, region, and sex; and adjusted for educational level (no formal education, primary school, secondary school, and high school or above), smoking (never, occasional, ex-regular, or current), alcohol consumption (never, ex-regular, occasional, or current), physical activity (five groups: <10, 10-<20, 20-<30, 30-<40, or 40+ MET hours /day), and baseline diabetes status (yes/no); **(b) SBP adjustment:** associations were further adjusted for baseline SBP, in addition to the adjustment described in (a); **(c) Mutual adjustment for one lipid fraction**: associations were further adjusted for one lipid fraction (not the exposure lipid fraction), in addition to that mentioned in (b); **(d) Mutual adjustment for another lipid fraction**: associations were further adjusted for another lipid fraction, in addition to those included in (c). For example, when assessing the association of LDL-C with risk of IS, the association analyses were

first adjusted for HDL-C (c), followed by the further adjustment for triglycerides (d), in addition to other vascular risk factors included in (b) SBP adjustment.

Strength of the associations

The *strength* of the associations was estimated, after the confirmation of the log-linear relationship of lipid measurements with IS risk. The associations of usual lipid values (in SD units or unit relating to clinical practice) with risk of incident IS were assessed using Cox proportional regression. These analyses were initially stratified by 5-year age-at-risk, region, and sex, and were followed by progressive inclusion of the following covariates into the regression model: education, smoking status, alcohol consumption, physical activity, baseline diabetes, SBP, and other lipid fractions (in SD units). Wald χ^2 statistics and their corresponding *p* values, were used to assess evidence of association after the sequential inclusion of additional covariates. The primary association analyses were not adjusted for adiposity, since adiposity acts as a determinant of several vascular risk factors, and it is therefore of little independent value in the multivariate regression analyses. However, sensitivity analyses were reported comparing the risk estimates from models with and without adjustment for adiposity measurements.

Subgroup analyses of the associations

The *consistency* of the associations of each lipid fraction (in SD unit) with risk of IS was compared across different population subgroups. All the analyses were stratified by age-at-risk (5-year), sex, study region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and SBP,

except the one of interest (i.e., stratified by). Given the differences in baseline SBP between cases and controls included in the nested case-control study, baseline SBP was also adjusted for when comparing the proportional relevance of lipid measurements for risk of IS across levels of baseline SBP. A chi-square test of heterogeneity or trend was used to assess the evidence of differences or trends in the strength of association in different population subgroups.

Sensitivity analyses

Based on the **Literature Review** (**Chapter 2**), and the initial exploratory analyses conducted, it has been suggested that SBP may have a major impact on the relevance of lipid fractions to IS risk, as indicated by the differences in χ^2 statistics before and after adjustment for SBP. Adjusting for measured SBP at baseline, rather than the long-term "usual" SBP, may, therefore, have underestimated the effect of SBP had on the associations of lipids with risk of IS. As a sensitivity analysis, usual SBP (with RDR of 0.66), rather than baseline SBP, was corrected, for comparison, assuming the lipids were measured without errors (i.e., not corrected for usual lipids). The correction was achieved by comparing the differences in the log HR between Cox regressions with (β_1) and without (β_2) adjusting for baseline SBP, and dividing the differences ($\Delta\beta=\beta_1-\beta_2$) by 0.66. The derived difference denotes the variance accounted by the usual SBP (i.e., $\Delta\beta/0.66$). The derived difference was added to the regression coefficient without adjusting for baseline SBP (β_2), and taking exponential. Likewise, the 95% CIs around the HRs was estimated by dividing the differences in standard errors between models by 0.66, and then adding to the one without adjustment for baseline SBP.

While it would be ideal to correct for the regression dilution bias of lipids and other covariates included in the multivariable analyses, but, as relatively few individuals had repeat measurements, the present study had insufficient statistical power to reliably assess the correlation of the measurement errors between blood lipids and SBP, and would otherwise introduce more uncertainties. Corrections for multivariable regression dilution bias are not reported in the thesis.

Statistical analyses were conducted in SAS version 9.3 (SAS Institute Inc., Cary, North Carolina, USA). Plots were produced using R version 3.2. Cox regression was conducted using a macro previously developed by colleagues in CTSU. The log-linear and forest plots were produced using an R package developed within CTSU.

4.3 Results

4.3.1 Incidence rates of ischaemic stroke in CKB participants

Description of the incidence rates of IS was conducted on 489 762 individuals with no prior history of cardiovascular disease (i.e., CHD, stroke, or TIA) at baseline. The number recruited from each of the ten study regions varied from 29 028 in Haikou to 60 357 in Henan. Among the individuals included in the analyses, 59% were women, and the mean baseline age was 52 years for men and was 51 years women, respectively. Overall, 49% of the participants were educated to middle school or above (men 58% and women 43%; urban 68% and rural 35%). Within a median of 8.8 years' follow-up, 32 869 incident IS cases were identified. The overall incidence rates of IS at age 30-90 years was 761 cases per 100 000 person-years (standardised by taking the unweighted average of the component age range incidence rates in both men and women) (**Table 4.3)**.

The incidence rates of IS increased with age, ranging from 53 cases per 100 000 person-years among individuals aged 30-40 years to 3590 cases per 100 000 person-years among those aged 80-90 years. The incidence rates of IS approximately doubled with each decade of increasing age in both men and women. In each age group, IS incidence rates were moderately higher in men than that in women, with more extreme sex-differences at older ages (**Table 4.3**). Excluding few individuals aged older 75 years, the incidence rate of IS in this Chinese population aged at 30-75 years was 658 cases per 100 000 person-years, with 736 in men and 606 in women, respectively.

There was considerable regional variation in IS incidence rates recorded in this study. The annual incidence rates at ages 30-90 years ranged from 330 cases in Sichuan to 2008 cases in Harbin per 100 000 person-years. In both men and women, the incidence rates tended to be higher in Northern (1081 cases per 100 000 person-years) than Southern (556) geographical regions, and in urban (937) than in rural areas (630). Consistent with the overall pattern, within each study area, the crude annual incidence rates of IS were moderately higher in men than in women.

4.3.2 Baseline characteristics of the nested case-control study participants

Descriptive analyses were conducted on 11 765 participants, including 5475 IS cases and 6290 controls (**Table 4.4**). In this sample population, directly standardised to the sex and region distribution of the study population included, IS cases were slightly younger than the controls (54.2 [SD: 10.3] versus 56.5 [12.2] years), and had a greater proportion of women than controls (52.9% versus

48.2%). The proportion of those living in urban areas was higher in cases than in controls (46.2% versus 21.1%).

Overall, between IS cases and controls, the patterns of smoking, alcohol consumption, and consumption of certain foods, were similar, but substantial sex-differences were observed for several characteristics. One-third of the study participants were current-smokers, with this varied from 62.7% in men versus 3.0% in women. About 16.1% of the participants drank alcohol at least once per week on average, with this varied from 30.5% in men versus 2.1% in women. Almost all participants consumed fresh vegetables daily (98%), and around 20% consumed fresh fruit on at least four days per week. Around 39% of the participants consumed meat or poultry regularly, and 25% and 5% of the individuals consumed eggs, fish or other seafood, at least 4-6 days per week, respectively. Around 10% of the participants regularly consumed dairy products of any type, and soybeans relevant food.

At baseline, 13.9% of the IS cases rated their overall health status as poor, compared to that of 9.2% in controls. The prevalence of diabetes at baseline in cases – including both a self-reported history of diabetes and those diagnosed from the blood glucose test performed during the baseline survey – was more than double that in controls (10.7% versus 5.1%), as was the proportion with self-reported hypertension (22.4% versus 10.1%).

At baseline, the overall mean (SD) BMI was slightly higher in IS cases and controls (24.1 [4.0] versus 23.3 [3.6] kg/m^2) (**Table 4.5**). Likewise, the mean of waist circumferences, hip circumference, and body fat percentage were all moderately higher in IS cases than in controls. Baseline SBP was, on average,

10 mm Hg higher in cases than controls (143 versus 133 mm Hg). Accounting for the age, sex, and study region, 24.8% of IS cases had a baseline SBP of 160 mm Hg or greater, compared to that of 10.8% in controls. Mean [SD] levels of physical activity varied little between IS cases and controls (18.2 [13.2] versus 19.1 [12.7] MET hours /day). Among a subset of participants with baseline measurements of random blood glucose (5437 cases and 6218 controls), the mean RBG was slightly higher in cases than that in controls (6.6 versus 6.0 mmol/L). After excluding few individuals with unreliable measurements of creatinine (>176 umol/L), the mean [SD] eGFR were similar between cases and controls (99.6 [13.1] versus 100.1 [10.2] ml/min/1.73m^2).

Irrespective of the disease status, the lipid-related biomarkers were approximately normally distributed, with the exception of triglycerides, where a right-skewed distribution was observed. Overall, mean values of lipid-related biomarkers were largely similar in each lipid fractions between IS cases and controls (**Table 4.6**).

4.3.3 Cross-sectional correlations of major lipids with other vascular factors

There were no significant differences by disease status in the correlations of major blood lipids with baseline characteristics, and, the correlation results were thus reported on all participants (n=16 541), with small differences in the numbers of participants for different biomarkers due to either missing or unreliable data.

Inter-correlations between lipoprotein-related biomarkers

Table 4.7 reports the Pearson correlation coefficients within lipid-related biomarkers, in addition to their correlations with baseline SBP, BMI, eGFR, and RBG. The evidence of a significant correlation in this table was defined as the

corresponding *p* value of 0.001 or below, after allowing for multiple testing. After accounting for age at baseline, sex, study regions and disease status, there were strong correlations within groups of lipid-related biomarkers (i.e., TC, LDL-C and Apo B100; HDL-C and Apo A1). Non-HDL-C and remnant cholesterol were calculated from values of the other lipid fractions, rather than directly measured, and their correlations were described separately in the end of the section.

TC, as a measurement of the overall cholesterol content transported in both LDL and HDL particles, was suggestive of a much stronger correlation with LDL-C than with HDL-C, with Pearson coefficients of 0.87 and 0.30, respectively. Each LDL particle contains one molecule of Apo B100, whereas there are no constant proportions between HDL particles and Apo A1. As expected, LDL-C was strongly correlated with Apo B100 (r=0.92), which was slightly greater than the magnitude of the correlation between HDL-C and Apo A1 (r=0.84). There were weak positive, and negative correlations of HDL-C with LDL-C (r=0.09), and with Apo B100 (r=-0.05), respectively. The diverging pattern of these correlations could reflect the inverse correlation between HDL-C and triglycerides (see below), or due to the play of chance. Similar patterns of the correlations were observed within pairs of Apo A1 with LDL-C (r=0.05), and Apo B100 (r=-0.02). There was a moderate inverse correlation between measured plasma concentrations of triglycerides (Log_e) and HDL-C (r=-0.40), and a weaker correlation was observed between plasma concentrations of triglycerides and Apo A1 (r=-0.12).

Non-HDL-C (i.e., TC minus HDL-C) was most strongly correlated with TC (r=0.95), Apo B100 (r=0.93), LDL-C (r=0.89), and had a modest correlation with triglycerides (r=0.47). In contrast, there was no statistically significant correlation

between non-HDL-C and HDL-C (p=0.063), and a weak positive correlation between non-HDL-C and Apo A1 (r=0.06). Remnant cholesterol represents the cholesterol content in triglyceride-rich lipoproteins, including approximately VLDLs and chylomicrons in non-fasting status, and VLDLs alone in fasting status. Remnant cholesterol was calculated as the plasma cholesterol content exclusive of that in LDL and HDL particles. Remnant cholesterol, was most strongly correlated with triglycerides (r=0.73), and moderately, weakly correlated with Apo B100 (r=0.48), and LDL-C (r=0.24), respectively. The correlations between remnant cholesterol and other lipid fractions should be interpreted with caution, as the plasma remnant cholesterol concentrations in the present and previous studies were calculated from other lipid measurements, and thus more likely to be subject to measurement errors.

Overall characteristics of lipid distributions

After adjustment for age, sex, study region, and disease status, the mean (SD) plasma concentrations of TC were 4.6 (0.9) mmol/L, and LDL-C were 2.4 (0.6) mmol/L, and HDL-C were 1.2 (0.3) mmol/L. The median (IQR) of triglycerides were 1.6 (1.3) mmol/L. The mean (SD) plasma concentrations of Apo A1 were 133.8 (21.5) mg/dl and Apo B100 were 84.5 (19.1) mg/dl (**Table 4.8.1**). There was modest variation in plasma lipid concentrations across the study regions, with slightly higher estimates in urban than in rural areas.

Overall, mean plasma concentrations of TC, LDL-C, and triglycerides tended to increase with age, peaking at age 50-60 years, while HDL-C measurements showed little age relevance (**Figure 4.1**). Due to the limited power of these sex-specific analyses, confidence intervals of sex-specific mean values overlapped

substantially between men and women. Overall, women appeared to have lower mean concentrations of TC, LDL-C, and triglycerides before the peaking (at age 50-60 years), while the opposite was true beyond this age group. Women had higher plasma concentrations HDL-C and Apo A1 throughout the age range than men (**Figure 4.2**).

Table 4.8.2-9 shows the baseline characteristics of study participants by lipid measurements. Overall, individuals with diabetes at baseline had higher mean plasma concentrations of TC, LDL-C and triglycerides, but lower HDL-C. Higher levels of physical activity appeared to be associated with an improved lipid profile: lower mean concentrations of TC, LDL-C, triglycerides, and higher HDL-C, however, these associations were only modest. Apart from the higher proportions of ever drinkers in individuals with higher HDL-C, and Apo A1, there was little evidence of any effect of smoking, or alcohol consumption on lipid values. Of all the dietary patterns recorded, individuals who consumed animal products regularly (i.e., meat or poultry, fish or other seafood, and dairy products) were more likely to have higher plasma concentrations of all lipid fractions, i.e., higher TC, LDL-C, triglycerides, and HDL-C. The relevance of dietary factors to lipid profiles in the present study may be less reliable than those suggested in others, if at all, given the substantially larger variation in dietary patterns between study regions (inter-regional variation) compared with those observed within each region (intra-regional variation), and, the comparatively small number of participants with lipid measurements included in the present analysis. In CKB, the dietary information at baseline was assessed with a relatively simple and qualitative food-frequency questionnaire. Arguably, more detailed quantitative questionnaires (e.g., the Food Frequency Questionnaire[197, 198]) can improve the dietary information for each

participant, but are not feasible, in large-scale population studies. Furthermore, in the present population, where the participants were relatively unselected in each region, the inter-region dietary variation tended to be larger than the intra-region variation. Hence, a self-administered quantitative questionnaire might be less reliable, simply because it is subject to more measurement errors than the real between-person variation would be, in each individual region.

Total cholesterol, low-density lipoprotein cholesterol, Apolipoprotein B100

Participants with higher LDL-C were slightly older, more likely to be women, to live in urban areas, and had lower levels of physical activity (p_{trend} <0.0001). There was no evidence of any relevance of education levels to LDL-C (p_{trend} =0.062), nor of household income (p_{trend} =0.55) (**Table 4.8.2**). There were no differences in LDL-C concentrations by smoking status or alcohol consumption. For the consumption of certain foods, after the adjustment for age, sex, region and disease status, individuals who regularly consumed meat or poultry tended to have higher mean plasma concentrations of LDL-C. A positive trend was observed in the consumption of fish or other seafood, and dairy products with LDL-C values, even though the proportions of regular consumption of these foods were contemporarily less common in this Chinese population than in Western populations. There were no differences in the distribution of egg or fresh fruit consumption by LDL-C concentrations.

There was strong evidence for a higher proportion of either diabetes or hypertension diagnoses at baseline among individuals with higher LDL-C (p_{trend} <0.0001), although no such difference was observed among individuals with self-rated poor health (p_{trend} =0.087) (**Table 4.8.3**). Adjusting for age at

baseline, sex, and study regions, LDL-C was weakly positively correlated with SBP (r=0.13), BMI (r=0.17), and RBG (r=0.08), and negatively (i.e., inversely) correlated with eGFR (r=-0.13) (**Table 4.7**).

The distributions of baseline characteristics among individuals with different values of TC and Apo B100 were consistent with those described for LDL-C. However, in men, the proportion of ever drinkers tended to be higher among individuals with higher levels of TC, and this trend was not observed in women (**Table 4.8.2**). This discrepancy can partly be explained by the slightly higher correlations of TC with HDL-C than that of LDL-C with HDL-C, where HDL-C was more likely to differ among individuals with different levels of alcohol consumption (**Table 4.8.4**), or caused due to the play of chance.

High-density lipoprotein cholesterol, Apolipoprotein A1

Table 4.8.3 presents the distributions of selected baseline characteristics among individuals with different plasma HDL-C concentrations. Overall, participants with higher HDL-C concentrations were more likely to be women (p_{trend} <0.0001), and of rural residence (p_{trend} =0.0087). In both men and women, ever-drinkers appeared to have higher plasma concentrations of HDL-C, and there was a positive correlation between levels of physical activity and HDL-C concentrations (p_{trend} <0.0001). Proportions of regular consumption of meat or poultry, egg, and dairy products appeared higher in individuals who had higher HDL-C. There was strong evidence that individuals having physician-diagnosed diabetes, or hypertension at baseline tended to have lower HDL-C (p_{trend} <0.0001). HDL-C concentrations were not correlated with self-rated health status. A moderate inverse correlation was observed between baseline HDL-C and BMI (r=-0.30), and

much weaker inverse correlations were observed with SBP (r=-0.05), and RBG (r =-0.10) (**Table 4.7**). Distributions of baseline characteristics by different Apo A1 concentrations were consistent with those of HDL-C (**Table 4.7** and **Table 4.8.6**).

Triglycerides

At baseline, individuals with higher mean plasma triglyceride concentrations were more likely to be women, and had lower levels of physical activity (p_{trend} <0.0001). There was a higher prevalence of both diabetes and hypertension among individuals who had higher plasma triglyceride concentrations. Triglyceride measurements were moderately positive correlated with BMI (r=0.36), and were weakly correlated with SBP (r=0.16), eGFR (r=-0.10), and RBG (r=0.19).

Non-HDL cholesterol and remnant cholesterol

As derived variables, the distributions of baseline characteristics across quintiles of non-HDL-C were similar to those for LDL-C, and Apo B100. Similar patterns were also observed between remnant cholesterol and triglycerides (**Table 4.7** and **Table 4.8.8-9**).

4.3.4 Regression dilution ratios of lipid-related biomarkers

Table 4.9 reports the mean values of lipid fractions at baseline and the first resurvey (approximately 2.7 years of follow-up since the baseline survey), and the second resurvey (8.0 years later after the baseline survey), from a subset of the study participants with paired measurements. **Figure 4.3** shows the serial convergence pattern of the ranges of major blood lipid biomarkers. Although the magnitude of the serial shrinkage differs somewhat between different lipid fractions, consistent patterns were observed. The inter-quintile range (i.e., the

difference between mean values of two extreme groups) attenuated markedly between the baseline and the first resurvey, followed by a much more modest shrinkage occurring thereafter (i.e., between the first and the second resurveys).

Estimates of the RDR from MacMahon-Peto's range ratio method and self-correlation method were reported to permit comparisons of consistency for each lipid fraction (**Table 4.9**). In general, the non-parametric and parametric methods yielded fairly similar results. Both methods indicated a general trend for more extreme RDRs with increasing time intervals between measurements. Based on the estimates from non-parametric methods, the RDRs for major blood lipid fractions, except triglycerides, declined by about 1% for each year of subsequent follow-up, whereas for triglycerides a slightly larger estimate was observed (3%), probably reflecting the larger within-person variation of plasma triglyceride concentrations. The RDRs for each lipid fraction was calculated for the approximate mid-point of exposure period (i.e., approximately at 3.7 years of follow-up after the baseline survey) using sequential surveys (**Table 4.9**). In the present study, the RDRs of major lipid fractions were all around 2/3, as would be expected from previous large prospective studies with similar follow-up time interval.[13, 14]

4.3.5 Prospective associations of major lipids with ischaemic stroke

Given SBP is one of the strongest risk factor of IS, and the substantial effect that SBP had on the relevance of major lipids for IS, the main results described and discussed in this thesis were on the SBP adjusted results, unless otherwise stated.

4.3.5.1 Low-density lipoprotein cholesterol, Apolipoprotein B100

Overall, throughout the range examined (i.e., 1.7-3.2 mmol/L), there was a log-linear positive relationship (i.e., one in which the percentage difference in risk remains constant across the range of exposures examined) of usual LDL-C with risk of IS (HR 1.11; 95% CI: 1.06-1.16 per SD higher), which translates to an adjusted HR of 1.17 (95% CI: 1.10-1.25) for each 1 mmol/L higher LDL-C.

The shape and the strength of the associations of usual LDL-C with risk of IS are demonstrated in **Figure 4.4** and **Figure 4.5**, respectively, after adjustment for various vascular risk factors recorded at baseline. Compared to individuals within the bottom quintile of usual LDL-C (mean: 1.7 mmol/L), the adjusted HRs of IS incidence were 1.08 (95% CI: 1.02-1.16), 1.12 (95% CI: 1.06-1.19), 1.19 (95% CI: 1.12-1.26), and 1.24 (95% CI: 1.17-1.31), for each consecutive higher quintile of usual LDL-C concentrations of approximately 2.1 mmol/L, 2.4 mmol/L, 2.6 mmol/L, and 3.1 mmol/L, respectively. **Figure 4.5** shows the proportional excess risk of IS associated with higher LDL-C concentrations remained significant, although was attenuated, after progressively adjustment for potential confounders, including the mutual adjustment for measurements of HDL-C and triglycerides.

The observed association of usual LDL-C with risk of IS was chiefly attenuated after adjustment for baseline SBP by approximately 20-25% (**Figure 4.5**). Given the evidence of weak correlations of LDL-C with HDL-C and triglycerides (**Table 4.7**), a mutual adjustment for HDL-C and triglycerides had little effect on the proportional relevance of LDL-C for risk of IS. Adjustment for HDL-C slightly increased the informativeness of LDL-C, albeit the confidence intervals were largely overlapping before and after the adjustment for HDL-C (**Figure 4.5**).

Apart from the moderate heterogeneity demonstrated in association of usual LDL-C with risk of IS between study region (rural versus urban), sex, and smoking status, there was no substantial heterogeneity by subgroups of various vascular risk factors (**Figure 4.6-8**). The HRs for risk of IS associated with each SD (0.69 mmol/L) higher LDL-C appeared to be higher in rural (HR 1.18; 95% CI: 1.11-1.25) than in urban areas (HR 1.05; 95% CI: 0.99-1.12). The association of LDL-C with fatal IS (n=226; HR 1.41; 95% CI: 1.15-1.73 per SD higher) seemed to be more extreme than non-fatal events (n=5249; HR 1.11; 95% CI: 1.07-1.16), and the proportion of fatal IS was slightly higher in rural than urban areas (5.5% versus 2.5%). Further adjustment for inflammatory biomarkers (e.g., fibrinogen) did not alter the regional difference of associations of LDL-C with risk of IS. Since in the nested case-control study of incident stroke types within CKB, the controls were selected chiefly to match ICH cases by study region, sex, and year of birth, around 78% of the controls were from rural areas. In contrast, around half of the IS cases were from urban areas, and thus, in urban areas, the imbalance between cases and controls (ratio of case to control was around 2) might contribute to this regional difference, if any.

The relative risk estimates were more extreme in men (HR 1.17; 95% CI: 1.10-1.25) than in women (HR 1.07; 95% CI: 1.01-1.13), and in ever-smokers (HR 1.20; 95% CI: 1.12-1.27) than in never-smokers (HR 1.05; 95% CI: 0.99-1.11). In the present nested case-control study of IS, proportions of ever-smokers (including ex-regular and current-smoker) in males and females were around 87% and 6%, respectively. Within either males or females, there was no evidence of any interaction between smoking and LDL-C, although this might be due the insufficient statistical power. Also, there was no statistically significant interaction

between LDL-C and sex when the analyses were performed separately by smoking status. Given the modest differences, and the number of the tests undertaken, such differences could reflect the play of chance. There was no evidence that the associations of LDL-C with IS varied by levels of eGFR ($p_{heterogeneity}$ =0.56), or HDL-C ($p_{heterogeneity}$ =0.20).

Consistent with the relevance of LDL-C for risk of IS, there was a log-linear positive relationship of baseline Apo B100 with risk of IS (**Figure 4.9**), with each 30 mg/dl higher plasma Apo B100 concentrations associated with a multivariable-adjusted HR of 1.11 (95% CI: 1.05-1.17) for risk of IS. As with LDL-C, the strength of the association was chiefly attenuated by adjusting for baseline SBP, and unaltered with further adjustment for Apo A1 and triglycerides. Adjusting for baseline SBP had a slightly larger effect on the association of Apo B100 with risk IS than that for LDL-C, with Wald χ^2 reduced by two-thirds and a half, respectively.

It has been suggested risk of atherosclerosis (evidence chiefly from coronary events) among individuals with discordant LDL-related measures, may be under- or over-estimated using LDL-C alone. Discordance was defined as LDL-C ≥median and the alternative measure <median, or vice versa,[199] or difference between percentiles greater than 5%.[200] In the present CKB study (n=11 765), 28.9% had discordant levels between Apo B100 and LDL-C, similar with that estimated in another study (~20%).[199] Compared to participants with both Apo B100 and LDL-C being lower than their median values, the higher risk of IS was only apparent among individuals with both high concentrations of Apo B100 and LDL-C (**Table 4.11**). Analyses in CKB suggested in general population measurements LDL-C and Apo B100 demonstrated similar risk prediction for IS.

4.3.5.2 High-density lipoprotein cholesterol, Apolipoprotein A1

Overall, in this population with relatively low LDL-C concentrations compared with those typically observed in Western populations, throughout the range of HDL-C assessed (i.e., 0.90-1.70 mmol/L), HDL-C was still inversely associated with risk of IS (**Figure 4.11**). The association of usual HDL-C for risk of IS was chiefly attenuated by adjustment for baseline SBP, and varied little with progressive adjustment for education, smoking, alcohol consumption, physical activity, and baseline diabetes (**Figure 4.12**). Mutual adjustment for baseline plasma concentrations of triglycerides, and LDL-C had little effect on the association of HDL-C with risk of IS.

There was no evidence of significant heterogeneity regarding the inverse association of usual HDL-C concentrations with risk of IS, across demographic factors (e.g., sex, study area, and education levels), lifestyle factors (e.g., smoking, alcohol consumption, and dietary patterns), and other factors (e.g., SBP and RBG) (**Figure 4.13-15**).

Figure 4.16 shows the relative risks of participants in each quintile of usual Apo A1 concentrations, compared to those within the top quintile. The strength of the associations of usual Apo A1 with risk of IS was equivalent to those observed for usual HDL-C (**Figure 4.17**). The shape and strength of the association of usual Apo A1 with risk of IS was unaltered by further adjustment for measurements of Apo B100 and triglycerides.

4.3.5.3 Triglycerides

There was weak evidence of a positive association of usual plasma triglyceride concentrations with risk of IS (HR 1.05; 95% CI: 1.00-1.10 per SD [0.58] higher triglycerides [Log_e]) after adjustment for baseline diabetes, SBP, and other established vascular risk factors (i.e., age, sex, study region, education, smoking, alcohol consumption, and physical activity) (**Figure 4.18**).

Figure 4.19 shows the results of associations of usual plasma triglyceride concentrations with risk of IS after sequential adjustment for various risk factors. When the associations were stratified by age-at-risk, sex, study region, education, smoking, alcohol, and physical activity, there was strong evidence that each 30% higher usual plasma triglyceride concentrations were associated with 6% (95% CI: 4-9%) higher risk of IS. Further adjustment for baseline diabetes moderately attenuated the proportional excess risks of IS (HR 1.05; 95% CI: 1.03–1.07). The association of plasma triglyceride concentrations with risk of IS appeared to be wholly or largely accounted by the adjustment for baseline SBP (HR 1.02; 95% CI: 1.00-1.04). The association was attenuated to null after further adjustment for LDL-C and HDL-C (HR 1.00; 95% CI: 0.95-1.05). There was no evidence of any substantial heterogeneity in the relevance of triglycerides for IS risk by levels of any of the baseline vascular risk factors examined (**Figure 4.20-22**).

4.3.5.4 Comparison of associations of lipid-related biomarkers with ischaemic stroke

Overall, the associations of LDL-C with risk of IS were slightly stronger than that for Apo B100, with HRs of 1.12 (95% CI: 1.07-1.16) and 1.07 (95% CI: 1.03-1.11), respectively for 1 SD higher. HDL-C and Apo A1 also showed similar inverse

associations with risk of IS, and the HRs were, respectively, 0.93 (95% CI: 0.89-0.97) and 0.94 (95% CI: 0.90-0.98). There was a weak positive association of triglycerides with risk of IS (HR 1.05; 95% CI: 1.00-1.10 per SD [0.58] higher triglycerides [Log_e]), which was weaker than that for LDL-C (**Table 4.10**). The associations of these lipid-related biomarkers with risk of IS were unaltered after further adjustment for BMI (**Table 4.10**).

Figure 4.23 compares the effects of adjustment for measured (or usual) SBP on the associations of baseline lipid-related biomarkers with risk of IS. As previously described, the associations of lipid-related biomarkers with risk of IS were chiefly attenuated after the adjustment for baseline SBP. The adjustment for usual SBP attenuated the associations further. Among the lipid fractions studied in this thesis, the largest attenuation by SBP was observed in the association of plasma triglyceride concentrations with risk of IS, followed by that in LDL-C, and Apo B100, whereas the relevance of HDL-C, and Apo A1 to IS was almost unaltered by the adjustment for SBP.

Baseline non-HDL-C (i.e., TC minus LDL-C) was positively associated with risk of incident IS (HR 1.06; 95% CI: 1.03-1.09 per 1 SD higher), after stratified by age-at-risk, sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, diabetic status, and baseline SBP, comparable with estimates for LDL-C, and Apo B100. There was weak evidence of a positive relationship between TC and IS risk, compared to those for LDL-C and Apo B100. Remnant cholesterol was, as previously described, defined as the remnant cholesterol apart from cholesterol content inclusive of HDL, and LDL particles, and as such, reflected chiefly the cholesterol content in those triglyceride-rich lipoproteins, as

suggested by the Pearson correlation coefficient (0.75) between remnant cholesterol and triglycerides (**Table 4.7**). Adjusted HRs of remnant cholesterol with risk of IS appeared to be weaker than those for triglycerides, after adjustment for various baseline vascular risk factors (**Figure 4.23**).

4.4 Discussion

4.4.1 Incidence rates of ischaemic stroke in CKB participants

In this large-scale general population-based prospective study of 489 762 participants with no prior history of CHD or stroke or TIA at recruitment, 32 869 incident IS events were observed during nearly 4.3 million person-years of follow-up. The annual IS incidence rate at age 30-89 years in this population was 734 per 100 000 person-years. The overall annual IS incidence rates were higher in men (863) than in women (692), in urban (937) than in rural (630) areas, and in northern (1081) than in southern (529) regions in China.

As already described in **Chapter 1**, reliable, contemporary estimates of IS incidence rates in China are limited. In Sino-MONICA-Beijing study (1984-2004) of men and women aged 25 to 74 years, around 14 500 stroke events (first-ever and recurrent events) were observed, and it was estimated that throughout the study period, the incidence rates of first-ever IS have increased by 8.7% annually on average.[51, 57] It has been noted that there was a significant increase in use of CT/MRI over the two decades' monitoring period, from around 4% in 1984 to 84% in 1994, and kept in a high use of around 97% in the second decade, which can potentially explain the dramatic increase in IS incidence rates around 1994. On average, in 2004, the annual incidence rate of IS estimated from Sino-MONICA-Beijing project was 213 cases per 100 000 person-years, among those aged 25-

74 years.[57] The estimated annual incidence rates were higher in men than that in women over the study period, although was not statistically significant.[57]

The IS incidence rates estimated in the Sino-MONICA-Beijing project were higher than those reported in high quality studies conducted in Western populations at around the same time,[201] of which, the rates were around half those reported in the Sino-MONICA-Beijing project in 2004, after standardising for differences in age-structure (i.e., Chinese population in 2000). Around the same time period, the rates of IS in the UK, as estimated in the Oxford Vascular Study in 2004,[202] was around 141 cases per 100 000 person-years, albeit proportions of the older participants in this study were greater than the Sino-MONICA and many other previous stroke epidemiological studies.[203]

In 2013, a door-to-door national survey (the NESS-China study) was conducted involving 480 687 individuals aged over 20 years (mean: 46.6; SD: 16.3) across 31 provinces in mainland China.[58] This survey was based on the National Disease Surveillance Points system, including 64 urban and 93 rural areas. A questionnaire-based primary screening of a history of stroke among 4500 residents (approximately 1500 households) within each survey sites was conducted by local CDC physicians. Incidence rates were calculated as the stroke events newly-occurred a year prior to the survey. Stroke pathological types were corroborated by findings from CT/MRI/autopsy in 83% of incident, and 90% prevalent cases. Overall, 69.6% of the incident, and 77.8% of the prevalent stroke cases were attributable to ischaemia, whereas 23.8% and 15.8%, respectively, were attributable to ICH.[58] Between 2012 and 2013, around 1643 incident stroke events were identified, and the crude incidence of stroke was 345.1 events

(95%CI: 328.4-361.7) per 100 000 person-years, with age standardised to the China census population 2010.[204] The incidence rates of total stroke or stroke pathological types increased with age, and were higher in men than that in women, albeit the sex-difference was not statistically significant. Of this, crude incidence rates were highest for IS (240.3), and ICH (82.1), and SAH (15.3), but, the estimates were based on a relatively limited number of stroke events by type.

In the NESS-China study, the mean age of the study population (46 years) was substantially younger than that in the CKB prospective study (59 years). The incidence rates of IS aged 40-49 years in CKB were 173.0, which was slightly lower than that in the NESS-China study of 240.3 (95%CI: 226.4-254.2). Incidence rates are generally lower in prospective epidemiological studies than estimates from the general population, partly due to that less healthy individuals are less likely to participate in such studies. In the NESS-China study, it is unknown whether individuals with prior history of cardiovascular diseases were excluded in the preliminary screening and the subsequent analyses, and whether those recurrent strokes occurred within survey period were also counted as incident stroke (i.e., first-ever and recurrent stroke). The household-based survey approach could help improve the sample size, but family members may share similar genetic or non-genetic risk factors for stroke, and consequently has overestimated the stroke incidence rates in a general population. In addition, a door-to-door approach was applied in the NESS-China study, and such active screening of stroke cases could also accounted for the slightly higher estimates in the NESS-China than other studies in China. A door-to-door assessment will, however, be resource-intensive in large-scale cohort studies (such as the CKB study). In CKB, stroke event data were obtained from three sources: the CDC

disease reporting system, the national health insurance system and death registries (see **Chapter 3** for details). Obtaining outcome data from multiple, pre-existing data sources, as in the CKB, not only can enable relatively cost-efficient identification of large numbers of disease events over long follow-up periods, but also can reduce susceptibility to biases in one database, and undercounting of stroke events.[43]

Both the NESS-China and the CKB studies have reported higher incidence rates of IS than those previously reported in the Sino-MONICA-Beijing project. Some increase in the incidence of IS since the Sino-MONICA-Beijing study should be expected given the epidemiological transitions occurring in China. Although the CKB study was not designed to be nationally representative, given the scale and the diversity of the CKB study population, in addition to a minimal loss to follow-up, this would not reduce its generalisability to the Chinese adult population. More relevant, the increasing incidence rates of IS with age, and higher incidence rates observed, in men than in women, in urban than in rural areas, and in Northern than Southern regions in China, consistent with previous results of studies conducted in China,[51, 52, 57, 205] and elsewhere.[6, 8, 39]

Substantial geographical variation in stroke incidence has been reported in previous studies as in the CKB study. In the NESS-China study, the highest annual incidence was in Northeast (365 events per 100 000 person-years, respectively), then Central areas (326), and the lowest incidence was in Southwest (154). Even within the UK, mortality due to stroke varies by nearly 50%, with fairly low rates in Oxfordshire.[206] However, nearly all this variation can be accounted for by differences in the frequency of major risk factors.[206, 207] The

North-South gradient in stroke incidence in China has been attributed to higher levels of blood pressure in northern regions,[208] however, a previous study conducted in the late 1980's across 67 rural counties in China found the geographic variation in stroke rates in China was poorly explained by variation in blood pressure, and in CKB, blood pressure was not found to be particularly higher in CKB in the Northern than Southern site.

4.4.2 Trends of changes in major lipids

Data on lipid distributions at population levels in China are limited. In the 1989 Ecological Survey across 69 Chinese counties, on average, mean concentrations of TC were 3.8 mmol/L, HDL-C were 1.2 mmol/L, non-HDL-C were 2.6 mmol/L, triglycerides were 1.1 mmol/L, and Apo A1 were 125 mg/dl, Apo B100 were 60 mg/dl, respectively.[160] Around the same time, TC was slightly higher in Shanghai (4.2 mmol/L).[75] In 2002, according to a national survey, mean concentrations of TC, LDL-C, HDL-C, and triglycerides in Chinese population were 4.5 mmol/L, 2.4 mmol/L, 2.3 mmol/L, and 1.1 mmol/L, respectively.[209] Estimates in the present study (with baseline plasma samples collected in 2004-2008) were, on average, consistent with previous reports in China, higher in urban (5.0 [SD: 1.0] mmol/L) than in rural (5.0 [SD: 1.0] versus 4.5 [0.9] mmol/L) areas.

During the mid-1980s to mid-1990s, the WHO MONICA Project monitored classic risk factors for CHD in 38 populations from 21 countries, showing the mean concentrations of TC in China were around 4.5 mmol/L, compared to that of 5.8 mmol/L in Western populations.[210] Between 1984 and 1999, among individuals included in the Sino-MONICA-Beijing projects,[211] five surveys were conducted to monitor the longitudinal changes in cardiovascular risk factors, and demonstrated

that mean TC concentrations in Beijing have increased from 4.2 mmol/L to 5.3 mmol/L.[212] Of this, mean TC were higher in urban than in rural areas, whereas greater increasing rates were observed in rural than in urban areas.

Globally, between 1980 and 2008, mean concentrations of TC declined in the HICs including Australasia, North America, and Western Europe, and in Central and Eastern Europe, whilst, mean TC increased in east and Southeast Asia and Pacific by 0.08 mmol/L per decade.[161] Despite these diverging trends, mean TC in 2008 was highest in the HICs (mean: ~5.2 mmol/L), of which, countries with the highest TC were mostly in Western Europe (~5.5 mmol/L).[161, 213-216] TC has a noticeable age dependence, peaking at around 50-60 years, as has been observed in the present and previous studies.[94, 217 161] When the data were available, at the population levels, TC and LDL-C were correlated,[161] and thus, trends in TC may broadly represent the global patterns of LDL-C.

4.4.3 Prospective associations of major lipids with ischaemic stroke

4.4.3.1 Low-density lipoprotein cholesterol

A few previous observational prospective studies, in different populations, with reasonably large numbers of IS cases, have provided evidence of a log-linear positive association of higher concentrations of LDL-C (or, non-HDL-C, and TC, approximately) with risk of IS. In the PSC study of 2134 IS deaths, the slightly higher proportional excess risks of IS mortality associated with lower TC concentrations was restricted among individuals with baseline SBP of 145 mmHg or below, above which, there was no evidence of any association between usual TC and mortality from IS.[13] The PSC results were altered marginally after combining the estimates of the MRFIT study.[13, 108] One possible explanation of the

absence of an overall association of usual TC with IS risk in PSC may reflect the potential misclassification of stroke pathological types, particularly in individual studies established before 1990, when the CT/MRI scanning was less widely used. By contrast, the ERFC study of 2534 incident IS cases (no Chinese populations were included), reported that each 1 mmol/L higher usual non-HDL-C was associated with a 12% (95% CI: 7-19%) higher risk of IS, independent of HDL-C, and triglyceride concentrations.[14] The APCSC study in both Asian and Australasian populations, involving a much smaller number of IS cases (n=749), has suggested a more extreme proportional excess risk associated with per 1 mmol/L higher usual TC (HR 1.25; 95% CI: 1.13-1.40), but the estimate was compatible with that in the ERFC and the present studies.

Compared to individual prospective studies, meta-analyses of IPD data has increased statistical power by increasing the number of cases, but they may be limited by the extent to which they could examine the effects of potential confounders have on the associations of interest. From the available evidence, three individual prospective studies with relatively large numbers of IS cases (>1000), i.e., the AMORIS study,[116] the KNHS study,[111] and the Kailuan study,[112] have reported largely comparable estimates of the association of baseline LDL-C with IS risk. The AMORIS study has included 4470 incident IS, identified from the Swedish national hospital discharge registry, and the national death registry. This study reported the associations between any lipoprotein-related biomarkers with stroke pathological types, separately, by sex and by severity (fatal or nonfatal), with no overall estimates. The risk estimates of the associations of LDL-C with IS were comparable in men and women, and in fatal and nonfatal IS cases, albeit the association was only significant for nonfatal events. All the group-specific

estimates were combined for this thesis using inverse-variance weighted fix-effect method, and it was estimated, in AMORIS, the relative risk estimates of baseline LDL-C with risk of IS were slightly lower (HR of 1.11; 95% CI: 1.07-1.14 per 1 mmol/L higher). The AMORIS study however, did not collect detailed data on baseline SBP, smoking, and diabetes, and thus, the estimates would have been substantially susceptible to residual confounding from these established vascular risk factors.

The KNHS study of an employee-based Korean population, with more than 6000 IS cases, demonstrated that IS was strongly and positively associated with baseline TC (HR 1.20; 95% CI: 1.16-1.24 per 1 mmol/L higher baseline TC), but the log-linear relationship appeared shallow below 4.1 mmol/L. The proportional excess IS risk in the KNHS study relative to per 1 mmol/L higher baseline TC was slightly larger than those in other studies, due potentially to the relatively younger and healthier nature of the study population. The Kailuan study conducted in a Chinese population, involving 1156 IS cases, each 1 mmol/L higher baseline non-HDL-C was associated with a 20% (95% CI: 10-33%) higher risk of IS, after accounting for various vascular risk factors, including HDL-C and triglycerides. The relatively shorter follow-up period (4-years) compared with other previous studies may have partially explained the slightly larger estimate in this than other studies. In aggregate, directly-measured LDL-C was not available in any of the three studies and none of these have corrected for regression dilution bias. Hence, estimates in these studies are likely to subject to residual confounding and random errors of a single measurement at the initial survey.

The present study, in addition to previous studies in different populations, has confirmed the log-linear positive association of usual LDL-C with risk of IS, extending the lower range further down to around 1.7 mmol/L in a general population. The strength of the association estimated in the CKB study with 5475 incident IS cases was consistent with those observed in Western populations, for example, the ERFC (non-HDL-C range: 3.3-5.7 mmol/L), and the AMORIS (range of LDL-C was not reported) studies,[116] but were slightly less extreme than that in the APCSC,[117] the KNHS,[111] and the Kailuan studies.[112] Both the KNHS and the Kailuan studies adjusted for hypertension status as a binary variable, rather than blood pressure as a continuous trait. Given the substantial effect of blood pressure on the association of LDL-C with risk of IS, as suggested in the PSC, the ERFC and the present studies, inadequate adjustment for blood pressure may have overestimated the relevance of LDL-C to IS risk. This may also suggest the predominant role of higher blood pressure in stroke risk overall, or by type, if any.

Previous studies have found the proportional relevance of LDL-C for risk of IS was substantially weaker than that to IHD (4-fold), [13, 14, 111] for example, in the ERFC study, the adjusted HRs were 1.12 (95% CI: 1.07-1.18) for IS and 1.49 (95% CI: 1.41-1.57) for IHD, which differs substantially from the comparable beneficial effects by lowering LDL-C with statin treatment on risk reduction of IS (RR 0.80; 95%CI: 0.75-0.86) and IHD (RR 0.76; 95% CI: 0.73-0.79).[15, 18] In the present study, a direct comparison of the strength of associations of LDL-C with IS and IHD is not currently available, but the strength of the association with IS appeared consistent with that in the statin trials with a relatively short-term follow-up period. Due to the heterogeneous nature of IS pathological mechanisms, compared to that in CHD, the relevance of LDL-C to different subtypes of IS may differ somewhat.

Adjudication exercises of IS subtypes have not fully complete, and hence investigation of the proportional relevance of LDL-C to IS subtypes is not accessible in the thesis. Some have suggested that lacunar infarction was more predominant in Chinese populations than Western populations,[11] which could have contributed to the more modest effect of LDL-C on IS than for IHD, at least in Chinese populations, however, evidence on the differences in IS subtype distribution across ethnicities has been inconclusive.[10]

In the present study, the association of usual LDL-C with incident IS risk appeared to be stronger in rural than in urban areas. The present study did not include those who reported statin or other lipid-lowering therapies at baseline, and therefore, the regional differences cannot be explained by difference in the use of lipid treatment. This could probably reflect more severity of IS cases (fatal versus non-fatal), provided the slightly higher proportion of fatal IS in rural than in urban, or inadequate management of certain conditions (e.g., diabetes) in rural than in urban areas, [218] or higher prevalence of other risk factors (e.g., smoking) in rural areas,[195] which is associated or correlated with higher LDL-C. Interpretation of the differences in populations subgroups should be cautious, as the number of IS cases in each stratum was relatively small, albeit the IS incidence is more common in urban than in rural areas and the absolute excess difference would probably be larger in urban than in rural areas.

4.4.3.2 High-density lipoprotein cholesterol

In this population with comparatively lower LDL-C, HDL-C was inversely associated with risk of IS (HR 0.93; 95% CI: 0.89-0.97 per SD [0.29 mmol/L] higher). The inverse association of usual HDL-C with risk of IS remained

significant after further adjustment for LDL-C, and triglycerides. The proportional differences in IS associated with usual HDL-C were largely independent of various vascular risk factors, and thus has confirmed the predictive value of HDL-C measurements in IS risk in this Chinese population.

Reliable observational evidence in regards to the associations of HDL-C with IS was limited, and chiefly from the ERFC,[14] and the AMORIS studies.[116] There was weak evidence, in the ERFC study, of an inverse association of usual HDL-C with IS risk (HR 0.89; 95% CI: 0.82-0.97 per 1 SD [0.39 mmol/L] higher HDL-C). The risk of IS, plotted on a log-scale, relative to usual HDL-C appeared to be shallower beyond 1.3 mmol/L.[14] In the AMORIS study, there was an inverse association overall, irrespective of sex and fatal status (HR 0.88; 95% CI: 0.85-0.90 per SD [0.39 mmol/L] higher HDL-C; the overall HR was calculated for comparison in the thesis from the published results). IS cases in the AMORIS study were identified from those routinely collected registries (i.e., hospital discharge registries and national death registries). Information on certain major vascular risk factors were not accessible (e.g., smoking, high blood pressure, and diabetes), and consequently, the associations in AMORIS study may be susceptible to residual confounding from such vascular risk factors. The shape and the strength of association of usual HDL-C with risk of IS observed in the present study were comparable with those in the ERFC and the AMORIS studies. In particular, the present study, involving a large number of IS cases from a general Chinese population, with contemporarily low LDL-C concentrations, demonstrated that plasma concentrations of HDL-C was inversely associated with risk of IS.

4.4.3.3 Triglycerides

In the CKB study, there was a marginal positive association of triglycerides with risk of IS (HR 1.05; 95% CI: 1.00-1.10 per SD [0.58] higher triglycerides [Log_e]). There is a lack of reliable evidence on the associations of triglycerides with risk of IS, with most of the evidence obtained from the ERFC, and the CCHS studies.[83, 107] The ERFC study demonstrated a moderate positive association of usual triglycerides with risk of IS (HR 1.12; 95% CI: 1.05-1.19 per SD [0.52] higher triglycerides [Log_e]), adjusting for age, sex, SBP, smoking, BMI, and diabetes. The proportional excess risk observed in the present study was less extreme than that in the ERFC study, albeit these were still statistically comparable. In both the ERFC and CKB studies, the association of triglycerides with IS was chiefly attenuated by the adjustment for SBP, with Wald χ^2 reduced by around 40% in both studies. In the ERFC study, however, the association of triglycerides with CHD was altered marginally by adjusting for SBP, probably reflecting the much more predominant effect of higher SBP on IS risk versus that on CHD.[14] The APCSC study with a much smaller number of IS events reported a marginally positive log-linear association of usual triglycerides with IS (HR 1.35; 95% CI: 1.00-1.83 per SD higher triglycerides [Log_e], and the SD was not reported).[122]

When adjusting for age, and sex only, a combined analysis of the CCHS and the CGPS, which had a similar number of IS cases to those recorded in the ERFC study reported similar estimates of the relative risks, but up to a much higher triglyceride concentrations (6.6 mmol/L) than that in the ERFC study (2.7 mmol/L).[83] In the ERFC study, of around 83% were Caucasians, the median concentration of triglycerides was 1.4 mmol/L, which was similar to that in the

Chinese population (1.6 mmol/L). The more extreme triglycerides values observed in the combination of the CCHS and the CGPS, both of which are general population-based studies, may limited its generalisability to other populations. Fuller adjustment for other confounders, performed in the CCHS study alone, provided consistent results with the combined results. In the CCHS study, the adjustment for SBP had little effect on the association of triglycerides with IS risk, in contrast with findings in both the ERFC,[14] and the present CKB studies.

Notably, in the ERFC study, plasma triglyceride concentrations appeared to provide little incremental predictive value in CHD, in addition to non-HDL-C, and HDL-C, as evidenced by the fact that the association of triglycerides with IHD was attenuated to null after adjustment for non-HDL-C and HDL-C,[14, 124] while no adjustments for non-HDL-C were made in the CCHS study.[219] Consistent attenuation patterns for IS after adjustment for cholesterol fractions were also observed in both the ERFC and CKB studies. The risk estimates of triglycerides for risk of IS, in both the ERFC and CKB, were chiefly affected by the adjustment for SBP, and it may be uninformative to assess its predictive value in IS risk by a further adjustment for other lipid fractions. In the AMORIS study, plasma triglyceride concentrations were positively associated with risk of IS (HR 1.17; 95% CI: 1.14-1.21 per SD). The AMORIS study was unable to adequately adjust for diabetes, and blood pressure. In the present CKB study, similar results were observed without adjusting for diabetes, or systolic blood pressure (HR 1.14; 95% CI: 1.08-1.20 per SD [0.58] higher triglycerides [Log_e]).

Triglycerides per se may be less likely to cause atherosclerosis than LDL-C, but may be an indicator for the triglyceride-rich lipoproteins, including VLDLs, IDLs,

and chylomicrons,[83] as suggested by genetic studies.[124] No large-scale RCTs have examined the effect of reducing triglycerides on cardiovascular disease risk. A meta-regression analysis of the effect of triglycerides-lowering in fibrate trials suggested that a 0.1 mmol/L reduction in triglycerides resulted in a 5% (95% CI: 1-10%) lower risk of coronary disease, with no effects on stroke.[148] Fibrates, however, can also lower LDL-C, and the observed risk reduction in coronary events was well consistent with the LDL-C reduction.[15] Substantial uncertainties persist about the relevance of lowering triglycerides for risk of ischaemic vascular disease, and more accurately, with what relevant treatment. To date, there is no consensus in guidelines about the relevance of raised triglycerides for prevention of cardiovascular disease, although there is agreement that very high triglyceride concentrations need to be treated to prevent acute pancreatitis.[220-223]

Remnant cholesterol is the cholesterol content in the triglyceride-rich lipoproteins, i.e., VLDLs, and chylomicrons, and thus, remnant cholesterol is strongly correlated with triglycerides concentrations. A direct assay that measures all remnants during the lipoprotein metabolism process has not yet been developed, and in previous studies, remnant cholesterol has been calculated as TC minus HDL-C minus LDL-C. Remnant cholesterol has been postulated to relate to risk of IHD,[224, 225] whilst no relevant evidence has been reported in regards to risk in IS. In the CKB study, there was no evidence of any association of remnant cholesterol with risk of IS.

4.4.4 Strengths and limitations

In estimating the IS incidence rates, the use of multiple overlapping surveillance systems – the mortality, disease surveillance system, and the health insurance systems – in identifying stroke events, would have substantially reduced the

undercounting of stroke events, selection and recall bias, in any individual data source. Furthermore, the remarkably high rates of use of CT/MRI scanning for diagnosis of stroke pathological types in CKB study allowed more accurate estimates of the epidemiology of stroke pathological types, rather than total stroke.

The present study included participants from a single population, which reduced the potential heterogeneity between different populations. In addition, plasma concentrations of LDL-C in CKB were assayed using direct methods (See **Chapter 3**), while in most of previous studies, LDL-C was calculated using the Friedewald equation.[179] Compared to the direct assay methods, the calculated LDL-C may be subject to the accumulation of measurement errors from other lipid fractions, and limited by the assumptions of the formula (e.g., requires a fasting plasma sample and triglyceride concentrations less than 4.5 mmol/L). Furthermore, in the present study, the diversity of the data collected at baseline, together with the resurvey assays of the lipid measurements have permitted a more comprehensive adjustment for potential confounding factors than previous studies, and correction for regression dilution bias of the lipid measurements.

Whilst the adjustment for regression dilution bias of major blood lipids is the strength of the present analyses, the RDR estimates observed in the present CKB study may still be susceptible to some imprecision, due to the relatively limited number of participants with re-measurements. Despite this, RDR estimates in the CKB study were largely comparable with those observed in previous studies of similar follow-up duration. Furthermore, in regression analyses, confounding variables should ideally be adjusted for regression dilution bias,[226, 227] but is not feasible with the non-parametric method. Several approaches have been

proposed to calibrate the measurement errors in covariates, for example, regression method,[228, 229] and component-of-variance method,[230, 231] the latter of which can include both systematic and random measurement errors in confounders. Such methods require a subset study (e.g., a reproducibility study or validation study) with repeat measurements of all covariates. The relatively limited number of individuals in the present study with repeat measurements of both lipids and other cofounder, particularly SBP, precluded a reliable estimation of the RDRs after adjustment for measurement errors in SBP in addition to lipid measurements.

Furthermore, although the present study has been one of the largest prospective studies of LDL-C with IS cases, the present study has only limited power for some subgroup analyses, for example, age-at-risk, blood pressure, and subtypes of IS. In addition to the possible impact of inaccuracies in recording of IS outcomes, as described above, the data sources used in the CKB may not capture milder, non-fatal IS cases. Due to the relatively short follow-up period, it was not feasible to routinely exclude IS cases occurring in the first few years of follow-up. Therefore, further studies with larger numbers of IS cases, and information on IS subtypes, would enable more thorough characterisations of the associations of major blood lipids with risk of IS.

4.5 Summary

In summary, the incidence rates of IS were 761 cases per 100 000 person-years, higher than previous studies in China and elsewhere, but showed consistent patterns with age, sex, and geographical regions. Throughout the range of usual LDL-C studied (i.e., 1.7-3.2 mmol/L), there was a log-linear positive association of LDL-C with risk of IS, consistent with risk estimates in Western population. There

was a marginal positive association of triglycerides with risk of IS, but was weaker than that for LDL-C. By contrast, HDL-C was inversely associated with risk of IS, independent of LDL-C, and other vascular risk factors. In addition, the associations of Apo B100 and Apo A1 with risk of IS were comparable with estimates for LDL-C and HDL-C, respectively.

Overall, the association of major blood lipids with risk of IS were chiefly attenuated after adjustment for SBP, consistent with the ERFC study (where the evidence was available), providing further emphases on the control and treatment of blood pressure for prevention of IS.

4.6 Tables

Table 4.1: Comparison of the disease status censored in 2014 and in 2016

Disease	**Disease status (2014)**						
status (2016)	IS case	ICH case	Control	MI case	SAH case	Fatal IHD	**Total**
IS cases	5206	56	225	6	2	5	5500
ICH cases	63	4681	44	1	16	1	4806
Control	184	115	5971	29	4	5	6308
Total	5453	4852	6240	36	22	11	16 614

IS=Ischaemic stroke; ICH=Intracerebral haemorrhage; MI=Myocardial infarction; SAH=Subarachnoid haemorrhage; IHD=Ischaemic heart disease;

Table 4.2: Exclusion criteria of lipid measurements in the present analyses

Lipid measurements	Exclusion criteria (mmol/L)*
TC	≥ 10.0
LDL-C	≥ 6.5
HDL-C	≥ 3.5
Triglycerides	≥ 10.0

TC=Total cholesterol; LDL-C=Low-density lipoprotein cholesterol; HDL-C=High-density lipoprotein cholesterol.

*Individuals with lipid concentrations greater than the above cut-offs were excluded.

Table 4.3: Number and crude incidence rate of ischaemic stroke, by sex and area (among 489 762 participants)*

	Men			Women			Total		
	Person-years at risk ($x10^3$)	No. of ischaemic stroke	Crude rate per 100 000 person-years	Person-years at risk ($x10^3$)	No. of ischaemic stroke	Crude rate per 100 000 person-years	Person-years at risk ($x10^3$)	No. of ischaemic stroke	Crude rate per 100 000 person-years
Age at risk (years)									
30-<40	81.25	55	68	133.77	58	43	215.01	113	53
40-<45	217.79	294	135	355.85	332	93	573.65	626	109
45-<50	266.17	710	267	432.05	867	201	698.22	1577	226
50-<55	265.24	1284	484	414.50	1689	407	679.74	2973	437
55-<60	281.23	2009	714	426.99	2847	667	708.21	4856	686
60-<65	232.41	2377	1023	335.25	3188	951	567.67	5565	980
65-<70	170.47	2469	1448	224.13	3029	1351	394.60	5498	1393
70-<75	131.78	2916	2213	158.69	3020	1903	290.46	5936	2044
75-<80	73.61	2269	3082	85.31	2208	2588	158.92	4477	2817
80-<90	15.55	599	3852	19.22	649	3377	34.76	1248	3590
All (30-<90)	*1735.5*	*14 982*	*863*	*2585.76*	*17 887*	*692*	*4321.24*	*32 869*	*761*
Study area									
Qingdao (U)	136.99	567	414	166.04	626	377	303.03	1193	394
Harbin (U)	163.20	3623	2220	253.39	4744	1872	416.58	8367	2008
Haikou (U)	87.26	1097	1257	156.96	1793	1142	244.22	2890	1183
Suzhou (U)	191.56	763	398	276.30	913	330	467.86	1676	358
Liuzhou (U)	150.54	1509	1002	250.91	1550	618	401.46	3059	762
Sichuan (R)	185.19	713	385	307.74	915	297	492.94	1628	330
Gansu (R)	164.06	986	601	270.27	1310	485	434.33	2296	529
Henan (R)	226.85	3018	1330	302.17	3318	1098	529.02	6336	1198
Zhejiang (R)	211.33	1005	476	312.24	1008	323	523.57	2013	384
Hunan (R)	218.52	1701	778	289.71	1710	590	508.23	3411	671
All	*1735.5*	*14 982*	*863*	*2585.73*	*17 887*	*692*	*4321.24*	*32 869*	*761*

* Excluding participants who had prior history of coronary heart disease or stroke or transient ischaemic attack at baseline; U=Urban; R=Rural.

Table 4.4: Selected baseline characteristics of the nested case-control study participants*

Characteristics	Ischaemic stroke cases (n = 5475)	Controls (n = 6290)
Age (years), %		
30-39	9.8	8.1
40-49	19.3	20.8
50-59	37.2	28.1
60-69	31.1	27.4
70-79	2.7	15.7
Mean (SD)	54.2 (10.3)	56.5 (12.2)
Sex (Women), %	53.5	48.1
Area (Urban), %	44.9	21.1
Highest level of education, %		
No formal education	22.1	21.8
Primary school	36.3	36.6
Secondary school	24.4	23.9
High school or above	17.2	17.7
Smoking category, %		
Never-smoker	53.6	55.3
Occasional smoker	5.8	6.5
Ever-regular	7.4	7.3
Current smoker	33.3	31.1
Alcohol consumption, %		
Never-drinker	44.8	42.7
Occasional drinker	6.4	4.7
Ever-regular	32.8	36.8
Current drinker	16.1	15.8
Regular consumption of certain foods, %		
Fresh vegetables	97.9	97.9
Meat or poultry	39.5	38.7
Eggs	25.2	26.9
Fresh fruit	21.1	22.9
Dairy products	10.6	10.7
Soybean	10.3	9.9
Fish or seafood	5.2	5.2
Self-rated health, %		
Poor	13.9	9.2
Fair	47.0	44.8
Good	23.7	28.2
Excellent	15.3	17.7
**Diabetes, % ** **	10.7	5.1
Hypertension, %	22.4	10.1

*Directly standardised to the baseline age (10-year), sex, and study region structure of the CKB study population, where appropriate. **Included doctor-diagnosed and screen-detected diabetes at baseline, see above Section (4.2.2.3) for definition of screen-detected diabetes.

Table 4.5: Selected baseline physical measurements of the nested case-control study participants*

Characteristics	Ischaemic stroke cases, % (n = 5475)	Controls, % (n = 6290)
Body mass index (kg/m^2)		
<20	11.5	15.7
20-<22	18.9	20.7
22-<24	21.7	23.9
24-<26	20.3	19.9
26+	27.6	19.9
Mean (SD)	24.1 (4.0)	23.3 (3.6)
Waist circumference (cm)		
<72	15.4	21.5
72-<78	20.1	23.8
78-<83	19.7	19.5
83-<89	19.5	17.9
89+	25.4	17.2
Mean (SD)	82.2 (11.4)	79.7 (10.1)
Hip circumference (cm)		
<84	14.6	18.5
84-<88	17.7	22.1
88-<92	23.9	22.0
92-<96	18.8	19.0
96+	24.9	18.5
Mean (SD)	91.3 (7.3)	90.0 (6.9)
Body fat percentage (%)		
<18	14.7	18.5
18-<24	20.8	22.9
24-<30	25.1	25.3
30-<35	17.8	16.6
35+	21.6	16.7
Mean (SD)	27.8 (8.0)	26.4 (7.3)
Systolic blood pressure (mmHg)		
<120	16.9	27.1
120-<135	24.4	31.2
135-<145	15.5	15.6
145-<160	18.3	15.3
160+	24.8	10.8
Mean (SD)	143 (31)	133 (22)
Random blood glucose (mmol/L)**		
<4.8	16.7	20.7
4.8-<5.8	34.0	38.4
5.8-<7.8	34.8	32.1
7.8-<11.1	8.0	6.0
11.1+	6.6	2.9
Mean (SD)	6.6 (3.6)	6.0 (3.0)

Table continues on next page.

Table 4.5: Selected baseline physical measurements of the nested case-control study participants*

Characteristics	Ischaemic stroke cases, % (n = 5475)	Controls, % (n = 6290)
Estimated glomerular filtration rate (ml/min/1.73m^2)* **		
<90	19.7	18.7
90-<100	26.7	27.2
100-<110	30.4	30.8
110-<120	17.5	18.1
120+	5.6	5.2
Mean (SD)	99.6 (13.1)	100.1 (10.2)
Total physical activity (MET hours/day), %		
<10	30.9	28.9
10-<20	34.3	34.3
20-<30	16.0	16.4
30-<40	11.3	12.0
40+	7.4	8.5
Mean (SD)	18.2 (13.2)	19.1 (12.7)

MET=Metabolic equivalent of task; *Directly standardised to the baseline age (10-year), sex, and study region structure of the CKB study population, where appropriate; **Among individuals with measurements of random blood glucose at baseline (5439 IS cases versus 6231 controls); ***Among individuals with creatinine measurements lower than 175 umol/L (5469 IS cases versus 6277 controls).

Table 4.6: Distribution of lipid-related biomarkers of the baseline samples among the nested case-control study participants*

Characteristics	Ischaemic stroke cases, % (n = 5475)	Controls, % (n = 6290)
Total cholesterol (mmol/L)		
<4.0	22.3	25.7
4.0-<4.5	20.0	22.7
4.5-<5.0	21.5	19.9
5.0-<5.5	16.6	15.2
5.5+	19.6	16.5
Mean (SD)	4.7 (1.1)	4.6 (1.0)
LDL cholesterol (mmol/L)		
<1.8	16.8	21.5
1.8-<2.3	27.0	29.8
2.3-<2.9	32.6	29.4
2.9-<3.4	15.0	12.8
3.4+	8.7	6.6
Mean (SD)	2.4 (0.8)	2.3 (0.7)
HDL cholesterol (mmol/L)		
<0.8	3.2	2.5
0.8-<1.0	18.9	15.5
1.0-<1.2	30.9	29.7
1.2-<1.4	23.8	25.7
1.4+	23.2	26.6
Mean (SD)	1.2 (0.3)	1.2 (0.3)
Triglycerides (mmol/L)		
<1.0	18.1	21.4
1.0-<1.5	25.0	27.2
1.5-<2.0	18.3	17.9
2.0-<3.0	20.5	19.3
3.0+	18.1	14.2
Median (IQR)	1.7 (1.4)	1.5 (1.2)
Apolipoprotein A1 (mg/dl)		
<110	11.8	10.6
110-<130	36.9	35.3
130-<140	17.4	18.5
140-<160	22.1	24.6
160+	11.8	11.0
Mean (SD)	133.3 (26.3)	134.1 (22.6)
Apolipoprotein B100 (mg/dl)		
<60	9.5	12.3
60-<80	29.9	32.9
80-<100	36.6	35.2
100-<120	18.8	15.7
120+	5.2	3.9
Mean (SD)	86.2 (23.9)	83.2 (21.7)

IQR=Interquartile range; *Directly standardised to the baseline age (10-year), sex, and study region structure of the CKB study population, where appropriate.

Table 4.7: Cross-sectional correlations of major blood lipids with blood pressure, adiposity, and kidney function, among the nested case-control study participants (n=16 541)*

Biomarkers	No. of individuals	TC	LDL-C	ApoB100	HDL-C	ApoA1	Log_e TG	non-HDL-C	Rem-C	SBP	BMI	eGFR	RBG
TC	16 541		0.87	0.87	0.30	0.32	0.33	0.95	0.58	0.09	0.17	-0.11	0.07
LDL-C	16 541	0.87		0.92	0.09	0.05	0.15	0.89	0.25	0.06	0.20	-0.08	0.05
Apo B100	16 541	0.87	0.92		-0.05	-0.02	0.34	0.93	0.48	0.10	0.27	-0.11	0.09
HDL-C	16 541	0.30	0.09	-0.05		0.84	-0.38	-0.01**	-0.17	-0.03	-0.29	0.04	-0.09
Apo A1	16 541	0.32	0.05	-0.02	0.84		-0.12	0.06	0.04	0.03	-0.18	0.02	-0.05
Log_e TG	16 541	0.33	0.15	0.34	-0.38	-0.12		0.47	0.75	0.13	0.34	-0.09	0.18
Non-HDL-C	16 541	0.95	0.89	0.93	-0.01**	0.06	0.47		0.67	0.10	0.27	-0.12	0.10
Rem-C	16 541	0.58	0.25	0.48	-0.17	0.04	0.75	0.67		0.11	0.24	-0.13	0.14
SBP	16 541	0.09	0.06	0.10	-0.03	0.03	0.13	0.10	0.11		0.23	-0.06	0.07
BMI	16 541	0.17	0.20	0.27	-0.29	-0.18	0.34	0.27	0.24	0.23		-0.08	0.09
eGFR ***	16 284	-0.11	-0.08	-0.11	0.04	0.02	-0.09	-0.12	-0.13	-0.06	-0.08		-0.03
RBG ****	16 340	0.07	0.05	0.09	-0.09	-0.05	0.18	0.10	0.14	0.07	0.09	-0.03	

TC=Total cholesterol; LDL-C=Low-density lipoprotein cholesterol; Apo=Apolipoprotein; Rem-C=Remnant cholesterol; HDL-C=High-density lipoprotein; TG=Triglycerides; SBP=Systolic blood pressure; BMI=Body mass index; eGFR=estimated glomerular filtration rate; RBG=Random blood glucose. * Pearson correlation coefficients adjusted for disease status, age at baseline, sex, and study region; ** Not statistically significant at levels of p <0.001; *** Among individuals with baseline creatinine measurement less than 176 umol/L (n=16 284); **** Among individuals with baseline measurements of random blood glucose (16 340).

Table 4.8.1: Mean (SD) baseline values of the lipid biomarkers among the nested case-control study participants (n=11 765)

	TC, mmol/L	LDL-C, mmol/L	HDL-C, mmol/L	Triglycerides, log_e	Apo A1, mg/dl	Apo B100, mg/dl
Sex						
Men	4.49 (0.88)	2.30 (0.65)	1.21 (0.29)	0.45 (0.59)	131.35 (23.04)	81.47 (19.48)
Women	4.80 (0.91)	2.43 (0.66)	1.27 (0.28)	0.53 (0.56)	137.33 (21.01)	85.74 (19.72)
Study area						
Qingdao (U)	5.32 (1.13)	2.84 (0.82)	1.27 (0.35)	0.52 (0.65)	135.91 (24.75)	98.67 (23.81)
Harbin (U)	4.88 (1.14)	2.54 (0.85)	1.20 (0.36)	0.53 (0.81)	129.7 (25.98)	90.02 (25.41)
Haikou (U)	5.39 (1.14)	3.02 (0.86)	1.26 (0.30)	0.46 (0.64)	135.12 (22.89)	99.72 (23.75)
Suzhou (U)	4.61 (0.92)	2.35 (0.72)	1.27 (0.32)	0.37 (0.63)	136.78 (22.77)	82.68 (21.60)
Liuzhou (U)	4.96 (0.97)	2.67 (0.74)	1.21 (0.28)	0.53 (0.61)	131.45 (21.11)	91.80 (20.56)
Urban areas	*4.93 (1.02)*	*2.60 (0.77)*	*1.23 (0.30)*	*0.49 (0.65)*	*132.78 (22.74)*	*90.56 (22.33)*
Sichuan (R)	4.65 (1.39)	2.31 (1.01)	1.40 (0.53)	0.40 (0.80)	148.14 (36.75)	79.22 (29.14)
Gansu (R)	4.25 (0.93)	2.08 (0.64)	1.21 (0.28)	0.52 (0.59)	130.84 (22.61)	75.82 (20.33)
Henan (R)	4.50 (0.89)	2.34 (0.64)	1.15 (0.27)	0.45 (0.59)	126.69 (21.16)	82.83 (19.53)
Zhejiang (R)	4.59 (0.95)	2.30 (0.69)	1.31 (0.32)	0.38 (0.57)	142.12 (23.94)	82.15 (21.38)
Hunan (R)	4.69 (0.95)	2.34 (0.69)	1.26 (0.31)	0.57 (0.60)	137.35 (24.61)	82.96 (20.38)
Rural areas	*4.51 (0.90)*	*2.26 (0.64)*	*1.24 (0.31)*	*0.48 (0.57)*	*134.87 (23.66)*	*80.48 (19.52)*
Overall	*4.64 (0.87)*	*2.36 (0.64)*	*1.24 (0.28)*	*0.48 (0.56)*	*134.4 (21.79)*	*83.45 (19.22)*

SD=Standard deviation; U=Urban areas; R=Rural areas; TC=Total cholesterol; LDL-C=Low-density lipoprotein cholesterol; HDL-C=High-density lipoprotein cholesterol; Apo=Apolipoprotein; Directly standardised to the baseline age (10-year), sex, region, and disease status structure of the study population included, where appropriate. Values are mean (SD) unless otherwise stated.

Table 4.8.2: Selected baseline characteristics of the nested case-control study participants by levels of total cholesterol (n=11 765)

Characteristics	Total cholesterol (mmol/L)					P for trend
	≤3.84	3.85-4.33	4.34-4.79	4.80-5.37	≥5.38	
Number of participants	2332	2393	2323	2352	2365	
Demographic factors						
Age, Mean (SD), years	53.8 (11.6)	54.9 (10.5)	55.3 (10.4)	56.2 (10.4)	56.8 (12.2)	<.0001
Women, %	39.0	46.2	51.0	55.4	62.4	<.0001
Urban, %	18.2	26.7	31.2	38.3	45.8	<.0001
High school or above, %	15.6	16.8	17.6	18.1	17.7	0.034
Household income >20 000 yuan/year, %	30.6	32.4	31.2	32.1	32.0	0.46
Lifestyle factors						
Male ever smoker, %	87.4	86.9	86.0	88.3	86.8	0.88
Female ever smoker, %	7.0	5.9	5.3	6.0	6.3	0.77
Male ever drinker, %	74.7	75.8	76.2	78.4	79.0	0.0035
Female ever drinker, %	37.6	36.4	37.2	37.0	35.8	0.45
Physical activity, Mean (SD), MET hours/day	19.7 (12.4)	18.9 (11.5)	19 (11.8)	18.2 (13)	17.6 (15.1)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	34.2	39.3	38.1	41.2	42.4	<.0001
Fish or other seafood	3.6	4.6	5.2	4.7	5.8	0.0028
Egg	25.2	26.6	25.9	27.2	27.8	0.078
Fresh fruit	22.6	21.8	21.7	22.4	21.2	0.50
Dairy products	10.2	10.1	10.9	11.3	12.5	0.0050
Health conditions, %						
Diabetes	6.8	6.2	6.8	8.7	11.5	<.0001
Hypertension	12.9	15.1	14.4	16.6	18.0	<.0001
Poor health	11.5	10.7	10.3	10.8	12.0	0.61

Values are mean (SD) unless otherwise stated. All values are directly standardised to the baseline age (10-year group), sex, and study region structure of the study population included, where appropriate.

Table 4.8.3: Selected baseline characteristics of the nested case-control study participants by levels of LDL cholesterol (n=11 765)

Characteristics	LDL cholesterol (mmol/L)					P for trend
	≤1.77	1.78-2.13	2.14-2.47	2.48-2.90	≥2.91	
Number of participants	2359	2330	2385	2353	2338	
Demographic factors						
Age, Mean (SD), years	53.5 (11)	54.9 (10.9)	55.5 (10.2)	56.4 (10.2)	56.9 (12.5)	<.0001
Women, %	44.4	48.6	51.4	52.2	57.9	<.0001
Urban, %	19.4	25.5	30.5	36.8	48.8	<.0001
High school or above, %	16.5	17.2	16.9	18.5	18.0	0.067
Household income >20 000 yuan/year, %	31.8	32.0	32.0	31.9	32.2	0.84
Lifestyle factors						
Male ever smoker, %	87.9	88.0	87.0	86.5	86.1	0.13
Female ever smoker, %	6.5	5.8	6.1	5.4	5.9	0.48
Male ever drinker, %	77.8	75.9	74.2	76.7	78.6	0.75
Female ever drinker, %	38.5	37.3	37.2	36.5	35.6	0.11
Physical activity, Mean (SD), MET hours/day	19.8 (13.5)	18.7 (11.5)	19.0 (11.4)	18.4 (12.7)	17.6 (13.8)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	35.7	38.4	39.2	40.3	41.8	<.0001
Fish or other seafood	4.1	5.0	4.4	5.1	6.3	0.0003
Egg	26.3	26.3	26.9	25.7	27.0	0.92
Fresh fruit	20.9	21.9	20.9	23.2	21.7	0.25
Dairy products	9.4	10.1	10.1	11.9	12.5	<.0001
Health conditions, %						
Diabetes	8.1	7.2	6.9	8.0	10.5	0.0005
Hypertension	14.2	14.9	15.4	16.1	18.3	0.0002
Poor health	12.0	11.6	10.0	10.8	11.6	0.47

Conventions as in Table 4.8.2

Table 4.8.4: Selected baseline characteristics of the nested case-control study participants by levels of HDL cholesterol (n=11 765)

Characteristics	HDL cholesterol (mmol/L)					P for trend
	≤0.90	0.91-1.10	1.11-1.30	1.31-1.60	≥1.61	
Number of participants	1157	3174	3275	2959	1200	
Demographic factors						
Age, Mean (SD), years	55.5 (12.3)	55.5 (10.6)	55.5 (10.1)	55.7 (10.4)	55.7 (11.3)	0.38
Women, %	34.9	45.7	53.3	57.9	61.1	<.0001
Urban, %	32.9	32.2	31.0	33.3	32.3	0.78
High school or above, %	19.0	18.7	17.1	15.4	18.3	0.0034
Household income >20 000 yuan/year, %	33.8	32.0	32.0	31.5	32.6	0.44
Lifestyle factors						
Male ever smoker, %	87.2	86.9	85.5	87.3	89.7	0.17
Female ever smoker, %	7.6	5.7	6.6	5.0	6.5	0.36
Male ever drinker, %	72.3	73.9	76.8	80.0	82.3	<.0001
Female ever drinker, %	34.4	35.9	36.5	38.2	40.1	0.0072
Physical activity, Mean (SD), MET hours/day	17.2 (11)	18.2 (11.3)	19.1 (11.6)	19.3 (12.7)	20.1 (13.3)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	40.0	38.9	37.2	41.1	41.2	0.043
Fish or other seafood	5.1	5.5	5.3	5.1	5.1	0.52
Egg	27.5	24.3	26.2	26.9	32.2	0.0005
Fresh fruit	20.6	22.3	22.5	22.0	21.1	0.92
Dairy products	9.5	10.3	10.6	11.9	14.3	<.0001
Health conditions, %						
Diabetes	12.1	9.9	6.8	6.0	5.7	<.0001
Hypertension	19.3	18.2	15.3	13.1	8.6	<.0001
Poor health	11.6	11.4	10.8	11.1	12.0	0.94

Conventions as in Table 4.8.2

Table 4.8.5: Selected baseline characteristics of the nested case-control study participants by levels of triglycerides (n=11 765)

Characteristics	Triglycerides (mmol/L)					P for trend
	≤0.98	0.99-1.35	1.36-1.82	1.83-2.62	≥2.63	
Number of participants	2357	2323	2360	2378	2347	
Demographic factors						
Age, Mean (SD), years	54.9 (11)	55.7 (10.3)	55.9 (10.2)	55.7 (10.2)	55.2 (10.1)	0.50
Women, %	43.8	49.3	52.7	56.1	52.6	<.0001
Urban, %	33.8	30.5	32.9	33.3	33.1	0.67
High school or above, %	16.5	17.1	17.5	18.0	18.7	0.018
Household income >20 000 yuan/year, %	31.5	31.1	31.2	31.9	33.9	0.031
Lifestyle factors						
Male ever smoker, %	86.3	86.2	87.1	86.3	86.6	0.78
Female ever smoker, %	5.2	5.3	5.2	6.9	6.0	0.10
Male ever drinker, %	75.4	75.4	76.7	76.9	80.0	0.004
Female ever drinker, %	38.9	38.6	37.8	36.0	35.5	0.017
Physical activity, Mean (SD), MET hours/day	20 (12.6)	19.1 (11.6)	18.6 (11.6)	18.5 (12.3)	17.3 (11.9)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	40.2	38.3	38.1	40.5	40.8	0.20
Fish or other seafood	4.6	5.7	4.4	5.1	5.8	0.23
Egg	28.2	25.6	26.9	26.1	26.1	0.19
Fresh fruit	21.9	20.7	22.4	22.7	21.9	0.39
Dairy products	11.7	11.1	10.7	11.5	11.6	0.91
Health conditions, %						
Diabetes	4.3	5.6	6.5	9.3	13.5	<.0001
Hypertension	9.4	13.7	15.4	17.6	20.3	<.0001
Poor health	11.7	11.9	11.7	10.1	10.9	0.097

Conventions as in Table 4.8.2

Table 4.8.6: Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein A1 (n=11 765)

Characteristics	Apolipoprotein A1 (mg/dl)					P for trend
	≤116.2	116.3-127.1	127.2-137.9	138.0-151.5	≥151.6	
Number of participants	2348	2367	2350	2342	2358	
Demographic factors						
Age, Mean (SD), years	54 (12.5)	55 (10.7)	55.0 (10.5)	56.4 (10)	57.4 (9.8)	<.0001
Women, %	33.9	48.1	54.3	59.1	60.8	<.0001
Urban, %	34.1	33.4	31.6	33.3	30.2	0.014
High school or above, %	16.4	19.4	16.9	16.8	16.1	0.13
Household income >20 000 yuan/year, %	31.1	32.3	30.9	33.0	32.0	0.37
Lifestyle factors						
Male ever smoker, %	87.3	85.8	86.0	86.4	90.1	0.095
Female ever smoker, %	6.3	6.4	5.7	5.4	5.8	0.42
Male ever drinker, %	72.2	73.4	75.2	80.1	82.8	<.0001
Female ever drinker, %	34.2	34.6	36.1	38.4	40.1	<.0001
Physical activity, Mean (SD), MET hours/day	17.8 (11.9)	18.5 (11.6)	19.0 (12.0)	19.2 (12.9)	19.3 (14.4)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	36.9	37.0	38.6	41.0	42.3	<.0001
Fish or other seafood	4.9	5.1	4.7	5.7	5.7	0.072
Egg	24.6	24.4	25.6	27.5	31.3	<.0001
Fresh fruit	21.1	22.6	22.8	21.9	19.4	0.12
Dairy products	9.7	10.7	10.8	12.4	11.5	0.005
Health conditions, %						
Diabetes	9.1	8.6	7.1	7.9	7.7	0.047
Hypertension	17.7	17.1	15.0	14.7	12.7	<.0001
Poor health	11.4	12.1	10.5	10.7	11.4	0.41

Conventions as in Table 4.8.2

Table 4.8.7: Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein B100 (n=11 765)

Characteristics	Apolipoprotein B100 (mg/dl)					P for trend
	≤65.4	65.5-77.5	77.6-88.2	88.3-100.3	≥100.4	
Number of participants	2349	2368	2332	2370	2346	
Demographic factors						
Age, Mean (SD), years	53.6 (12.2)	54.8 (10.9)	55.6 (10.1)	56.2 (10.5)	56.9 (11.5)	<.0001
Women, %	44.4	48.4	50.9	55.2	57.4	<.0001
Urban, %	18.9	24.7	32.4	36.7	47.4	<.0001
High school or above, %	16.8	18.0	16.3	19.0	17.3	0.33
Household income >20 000 yuan/year, %	32.1	32.0	31.0	32.0	34.0	0.13
Lifestyle factors						
Male ever smoker, %	87.2	87.7	87.4	85.5	87.4	0.54
Female ever smoker, %	5.9	6.2	5.5	6.2	5.6	0.84
Male ever drinker, %	76.7	75.1	75.4	78.0	76.5	0.57
Female ever drinker, %	39.7	37.2	37.3	36.2	35.1	0.017
Physical activity, Mean (SD), MET hours/day	19.8 (12.5)	19.2 (11.5)	19 (11.8)	18.1 (12)	17.2 (14.6)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	37.7	38.0	38.5	41.2	42.3	<.0001
Fish or other seafood	3.8	4.3	5.1	4.9	6.2	<.0001
Egg	26.1	26.5	26.6	27.0	25.6	0.99
Fresh fruit	22.8	21.8	20.9	22.6	21.6	0.62
Dairy products	10.7	9.4	10.9	12.2	11.6	0.013
Health conditions, %						
Diabetes	5.7	4.9	7.0	9.0	12.3	<.0001
Hypertension	11.7	12.8	15.9	16.9	19.4	<.0001
Poor health	11.5	11.5	10.4	10.9	11.2	0.61

Conventions as in Table 4.8.2

Table 4.8.8: Selected baseline characteristics of the nested case-control study participants by levels of non-HDL cholesterol (n=11 765)

Characteristics	non-HDL cholesterol (mmol/L)					P for trend
	≤2.62	2.63–3.10	3.11–3.54	3.55–4.12	≥4.13	
Number of participants	2343	2366	2338	2362	2356	
Demographic factors						
Age, Mean (SD), years	53.4 (11.8)	54.9 (10.6)	55.7 (10.2)	56.1 (10.6)	56.7 (12.1)	<.0001
Women, %	41.6	46.4	51.9	56.0	59.2	<.0001
Urban, %	17.9	25.8	31.3	37.5	47.6	<.0001
High school or above, %	17.3	16.4	16.9	19.6	17.4	0.072
Household income >20 000 yuan/year, %	33.4	31.8	31.0	32.0	33.4	0.72
Lifestyle factors						
Male ever smoker, %	87.1	86.6	88.0	85.9	86.3	0.57
Female ever smoker, %	6.2	6.1	5.8	5.7	6.0	0.71
Male ever drinker, %	77.2	74.9	76.0	76.1	78.9	0.43
Female ever drinker, %	40.3	36.7	36.6	35.7	35.1	0.016
Physical activity, Mean (SD), MET hours/day	20 (13.1)	19.3 (11.6)	18.7 (11.8)	18.3 (12.2)	17.1 (16.9)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	36.4	37.1	39.8	40.7	42.4	<.0001
Fish or other seafood	3.8	4.8	4.9	4.8	6.2	0.0001
Egg	26.5	25.6	27.4	26.9	26.3	0.67
Fresh fruit	23.4	21.1	23.1	21.1	22.3	0.41
Dairy products	10.5	9.4	11.1	11.9	12.0	0.0045
Health conditions, %						
Diabetes	6.2	5.1	6.6	8.5	12.6	<.0001
Hypertension	12.9	13.7	14.0	17.1	19.8	<.0001
Poor health	10.7	10.9	10.2	10.4	12.2	0.23

Conventions as in Table 4.8.2

Table 4.8.9: Selected baseline characteristics of the nested case-control study participants by levels of remnant cholesterol (n=11 765)

Characteristics	Remnant cholesterol (mmol/L)					P for trend
	≤0.71	0.72–0.87	0.88–1.04	1.05–1.30	≥1.31	
Number of participants	2374	2346	2334	2368	2343	
Demographic factors						
Age, Mean (SD), years	55 (10.7)	55.4 (10.6)	55.2 (10.3)	55.5 (10.6)	55.8 (11)	0.020
Women, %	39.1	46.4	52.8	58.9	59.9	<.0001
Urban, %	25.5	29.8	32.8	35.4	39.6	<.0001
High school or above, %	16.8	15.9	16.4	19.9	18.3	0.0008
Household income >20 000 yuan/year, %	33.0	30.6	29.8	32.3	34.8	0.025
Lifestyle factors						
Male ever smoker, %	86.6	86.2	87.1	86.2	85.4	0.59
Female ever smoker, %	5.0	5.6	5.2	6.7	6.2	0.086
Male ever drinker, %	75.5	76.2	76.0	78.1	77.2	0.17
Female ever drinker, %	39.9	37.3	36.6	37.0	34.7	0.012
Physical activity, Mean (SD), MET hours/day	19.9 (13.6)	19.1 (11.7)	19 (11.6)	18.5 (12.6)	17 (13.4)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	37.3	39.1	39.3	40.8	41.2	0.0013
Fish or other seafood	4.0	5.3	4.8	5.3	5.6	0.032
Egg	25.0	26.9	27.1	27.8	27.2	0.087
Fresh fruit	23.2	21.7	22.0	21.7	21.3	0.14
Dairy products	11.0	10.8	10.6	11.1	11.7	0.35
Health conditions, %						
Diabetes	5.3	5.1	7.3	8.9	13.1	<.0001
Hypertension	11.5	13.3	14.4	17.5	20.4	<.0001
Poor health	10.9	10.8	11.4	11.3	12.0	0.19

Conventions as in Table 4.8.2

Table 4.9: Non-parametric and parametric estimates of regression dilution for lipoprotein-related biomarkers in the China Kadoorie Biobank study

	No. of individuals	Mean (SD)	Time-interval between paired measurement	MacMahon -Peto estimates	Self-correlation estimates (SE)*
Total cholesterol (mmol/L)					
Baseline	16 541	4.64 (0.95)			
First resurvey	317	4.58 (0.87)	2.74	0.71	0.73 (0.03)
Second resurvey	858	4.86 (1.00)	8.00	0.66	0.64 (0.02)
Middle time follow-up			3.70	0.70	0.71
HDL cholesterol (mmol/L)					
Baseline	16 541	1.24 (0.29)			
First resurvey	317	1.23 (0.28)	2.74	0.72	0.74 (0.03)
Second resurvey	858	1.26 (0.29)	8.00	0.71	0.71 (0.02)
Middle time follow-up			3.70	0.72	0.72
LDL cholesterol (mmol/L)					
Baseline	16 541	2.36 (0.69)			
First resurvey	317	2.32 (0.66)	2.74	0.70	0.74 (0.03)
Second resurvey	858	2.53 (0.73)	8.00	0.67	0.63 (0.02)
Middle time follow-up			3.70	0.68	0.73
Non-HDL cholesterol (mmol/L)					
Baseline	16 541	3.40 (0.90)			
First resurvey	317	3.35 (0.84)	2.74	0.75	0.76 (0.02)
Second resurvey	858	3.60 (0.95)	8.00	0.68	0.66 (0.02)
Middle time follow-up			3.70	0.75	0.74
Log_e triglycerides					
Baseline	16 541	0.48 (0.58)			
First resurvey	317	0.50 (0.60)	2.74	0.65	0.57 (0.04)
Second resurvey	858	0.49 (0.58)	8.00	0.49	0.51 (0.03)
Middle time follow-up			3.70	0.62	0.56
Remnant cholesterol (mmol/L)					
Baseline	16 541	1.04 (0.42)			
First resurvey	317	1.03 (0.43)	2.74	0.66	0.60 (0.04)
Second resurvey	858	1.07 (0.43)	8.00	0.60	0.53 (0.02)
Middle time follow-up			3.70	0.65	0.60
Apolipoprotein A1 (mg/dl)					
Baseline	16 541	134.40 (23.22)			
First resurvey	317	133.70 (20.96)	2.74	0.63	0.62 (0.03)
Second resurvey	858	137.47 (22.16)	8.00	0.67	0.61 (0.02)
Middle time follow-up			3.70	0.65	0.60
Apolipoprotein B100 (mg/dl)					
Baseline	16 541	83.45 (20.76)			
First resurvey	317	84.00 (19.74)	2.74	0.77	0.78 (0.02)
Second resurvey	858	88.94 (21.44)	8.00	0.74	0.69 (0.02)
Middle time follow-up			3.70	0.75	0.76

SD=Standard Deviation; SE=Standard Error; The three surveys were conducted in 2004-2008, 2008-2009, and 2013-2014, respectively. *Pearson product-moment correlation coefficient (r), and SE was calculated as $(1-r^2)/\sqrt{n}$.

Table 4.10: Comparison of hazard ratios (HR) of major lipid-related biomarkers with ischaemic stroke, before and after adjustment for body mass index (BMI)*

Lipoprotein biomarkers (SD)	HR (95% CI)	
	Without BMI	With BMI
LDL cholesterol (0.68 mmol/L)	1.12 (1.07-1.16)	1.10 (1.06-1.15)
Triglycerides (Log_e) 0.58	1.05 (1.00-1.10)	1.03 (0.98-1.08)
HDL cholesterol (0.29 mmol/L)	0.93 (0.89-0.97)	0.94 (0.90-0.99)
Apolipoprotein B100 (21 mg/dl)	1.07 (1.03-1.11)	1.06 (1.02-1.10)
Apolipoprotein A1 (22 mg/dl)	0.93 (0.89-0.98)	0.94 (0.90-0.99)

SD=Standard deviation; CI=Confidence intervals; *Analyses were stratified by age-at-risk (5-year), sex, and study region, and adjusted for education, smoking, alcohol consumption, physical activity, diabetes, and baseline systolic blood pressure (with/-out BMI). Hazards ratios (HRs) were presented for differences in SD units of the measured value of each lipid-related biomarker. Values for triglycerides are natural logarithm of triglyceride measurements.

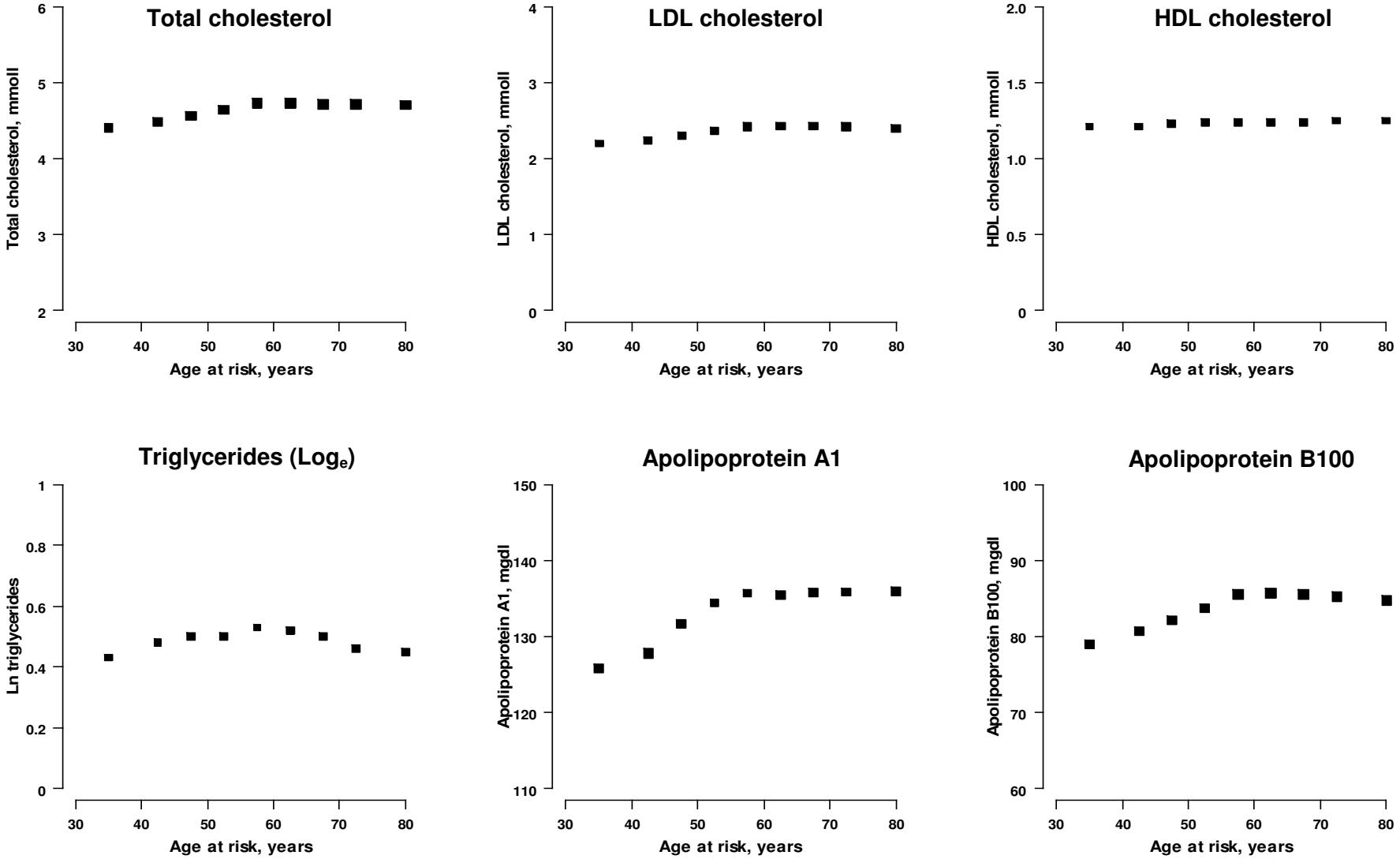
Table 4.11: Discordance analyses of LDL cholesterol and Apolipoprotein B100 in Women's Health Study (WHS) and China Kadoorie Biobank study (CKB)

	WHS*	CKB
LDL-C < Median and Apo B100 < Median**	Reference	1.00 (0.95-1.05)
LDL-C < Median and Apo B100 > Median	2.98 (2.41-3.68)	1.00 (0.89-1.16)
LDL-C > Median and Apo B100 < Median	0.31 (0.23-0.41)	1.08 (0.98-1.19)
LDL-C > Median and Apo B100 > Median	Reference	1.13 (1.09-1.18)

LDL-C=Low-density lipoprotein cholesterol; Apo B100=Apolipoprotein B100. * Median of LDL-C and Apo B100 were 2.4 mmol/L and 84.5 mg/dL, respectively. ** Women's Health Study included 28 345 women aged 45 years or older, with no prior history of self-reported cardiovascular disease and cancer at recruitment, and coronary events were obtained from self-reported follow-up questionnaires.

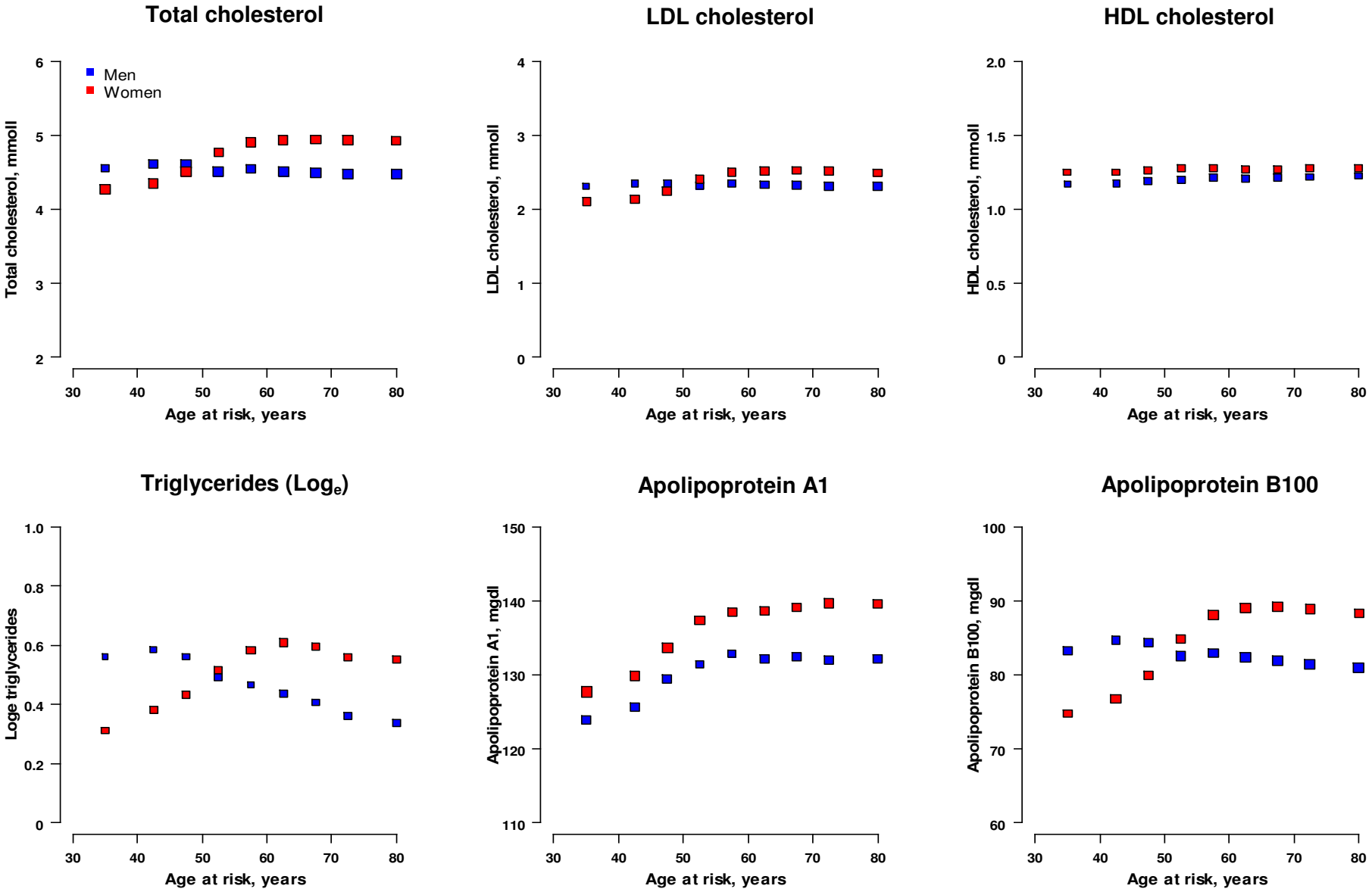
4.7 Figures

Figure 4.1: Mean values of major lipid-related biomarkers by age groups*



Linear regression was used to estimate the group-specific means, adjusted for disease status, regions, and sex. Size of square was in proportional to the number of participants included in that group.

Figure 4.2: Mean values of major lipid-related biomarkers by age groups and sex



Linear regression was used to estimate the group-specific means, adjusted for disease status, and region. Size of square was in proportional to the number of participants included in that group.

Figure 4.3: Serial shrinkages of ranges for major lipids and apolipoproteins among individuals with paired measurements

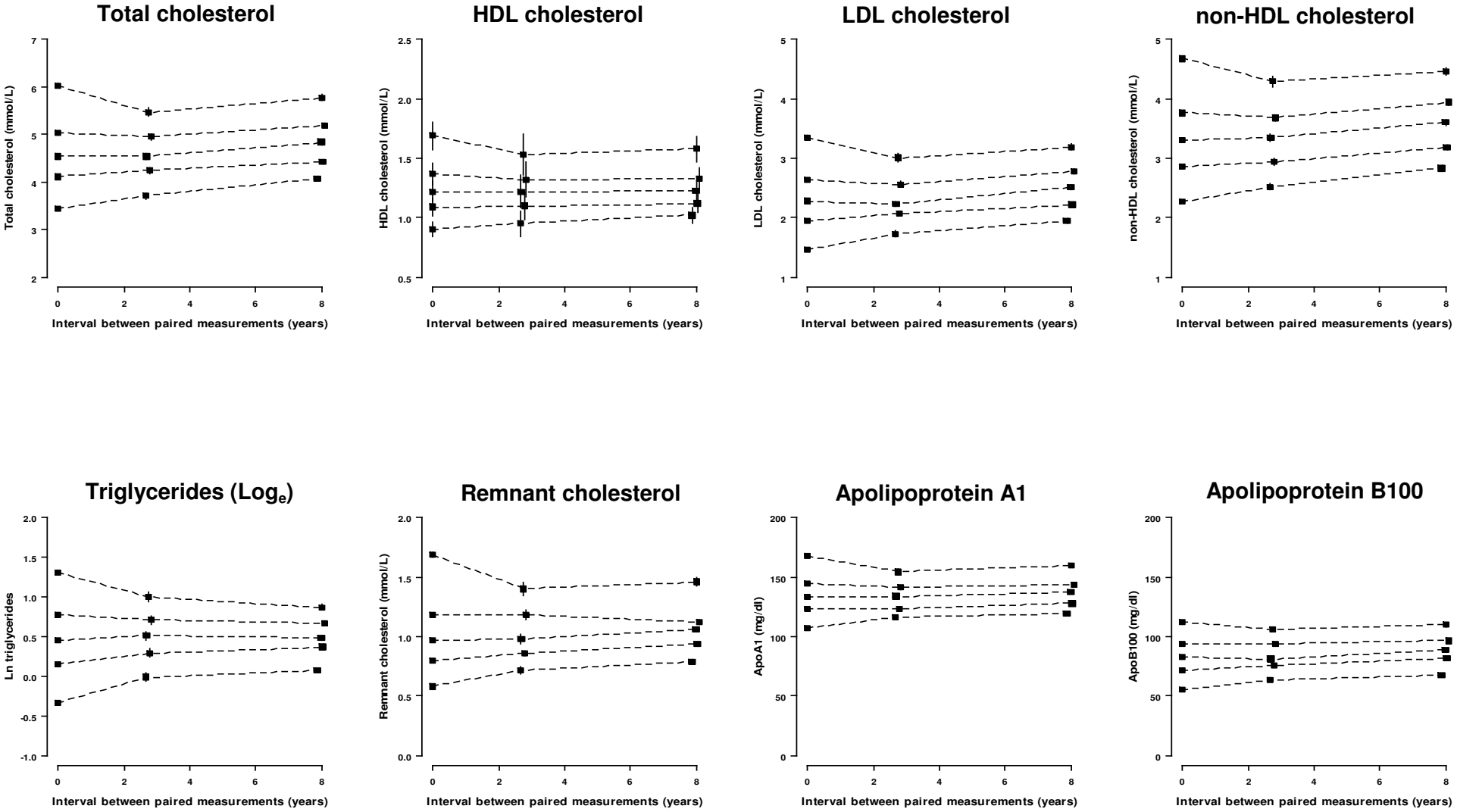
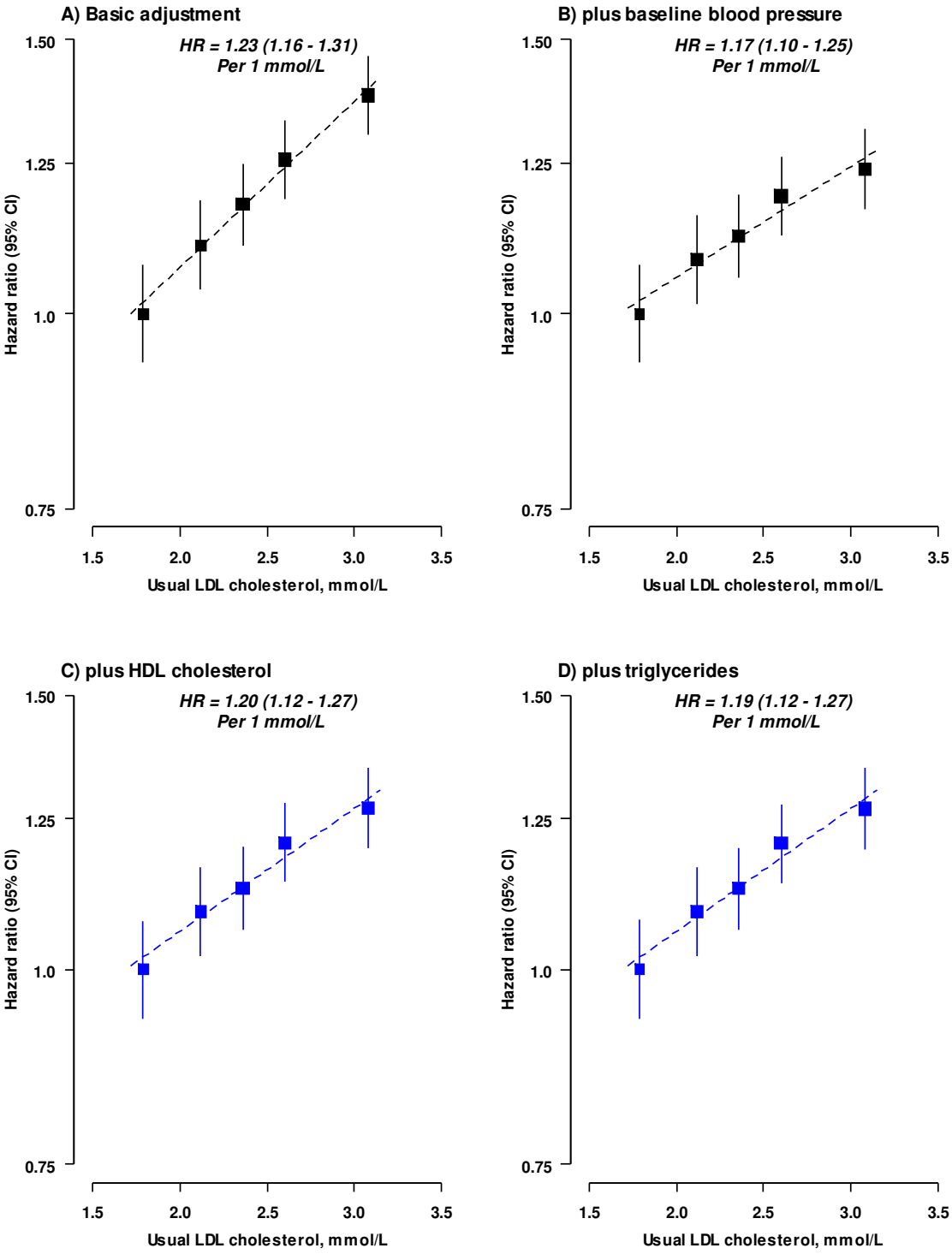
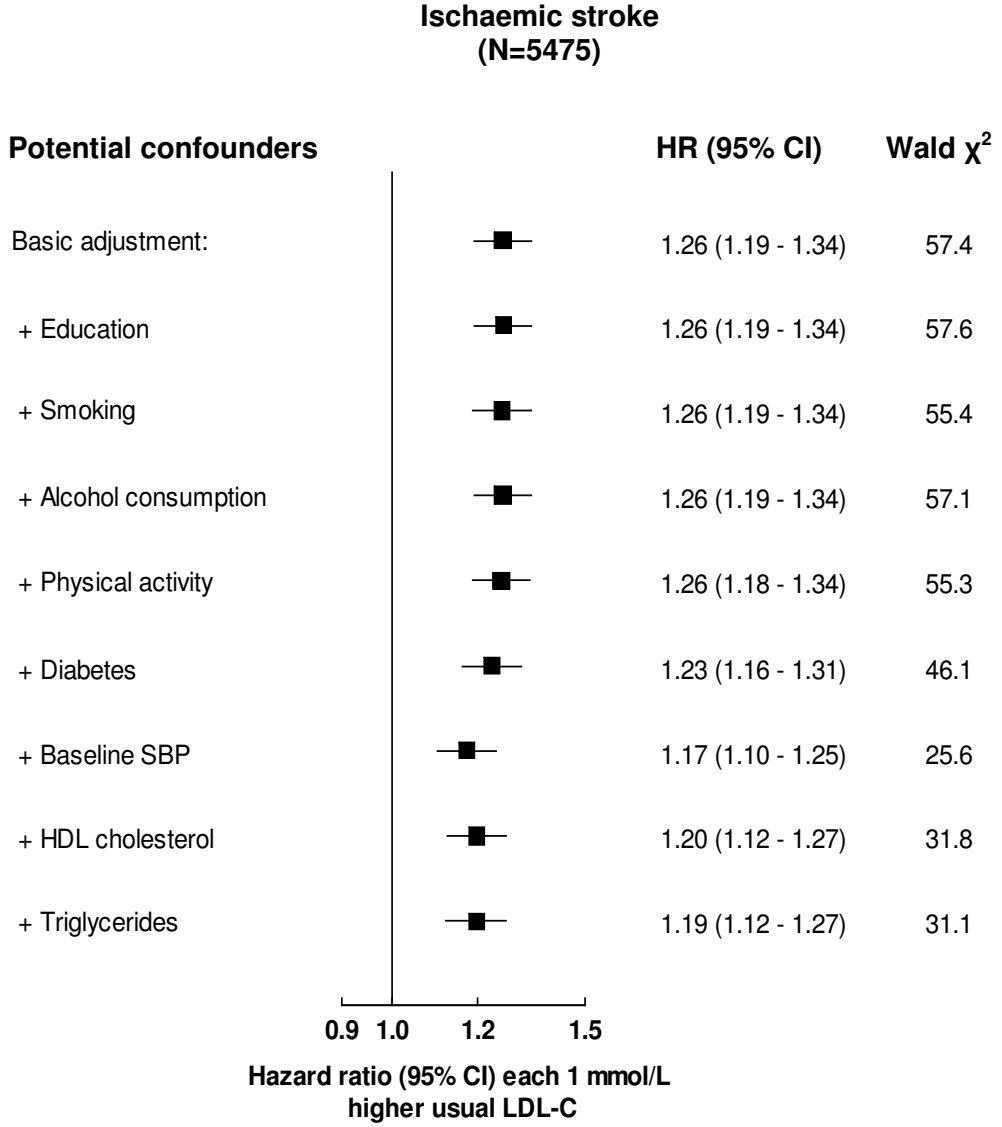


Figure 4.4: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol



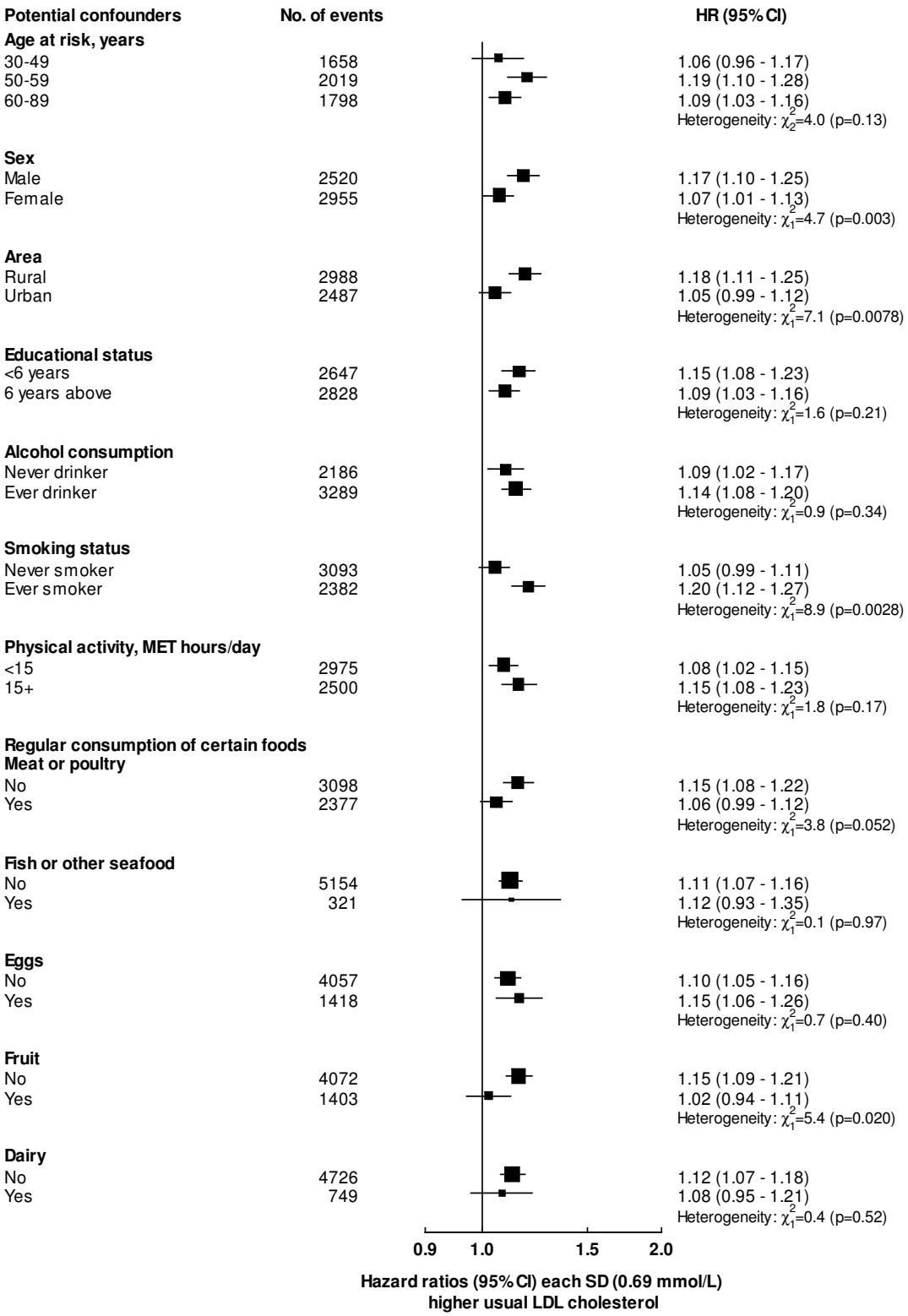
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for HDL cholesterol in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual LDL cholesterol concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 4.5: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol, progressively adjusted for potential confounding factors



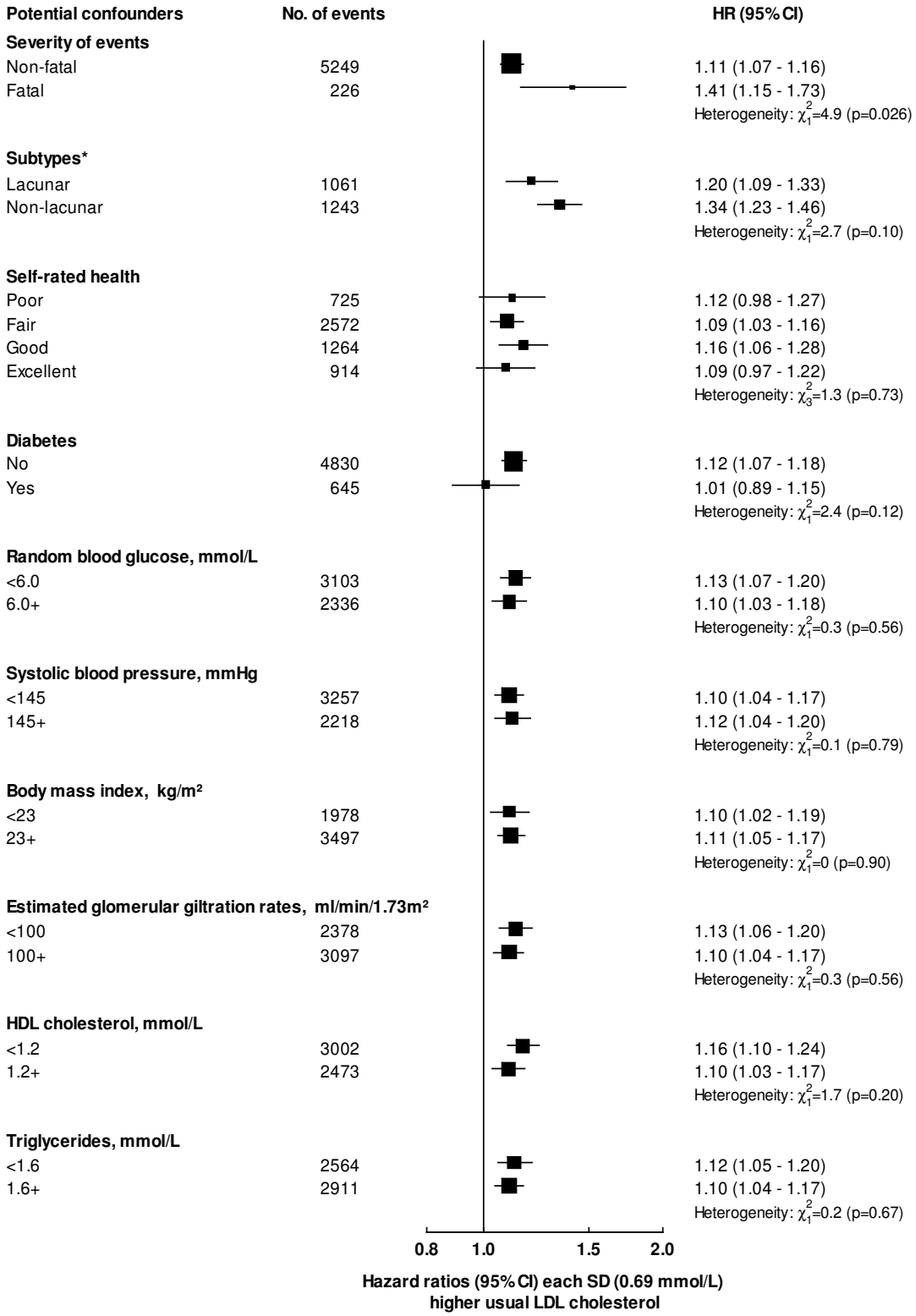
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual LDL cholesterol, per 1 mmol/L higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 4.6: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol in population subgroups of baseline vascular risk factors (I)



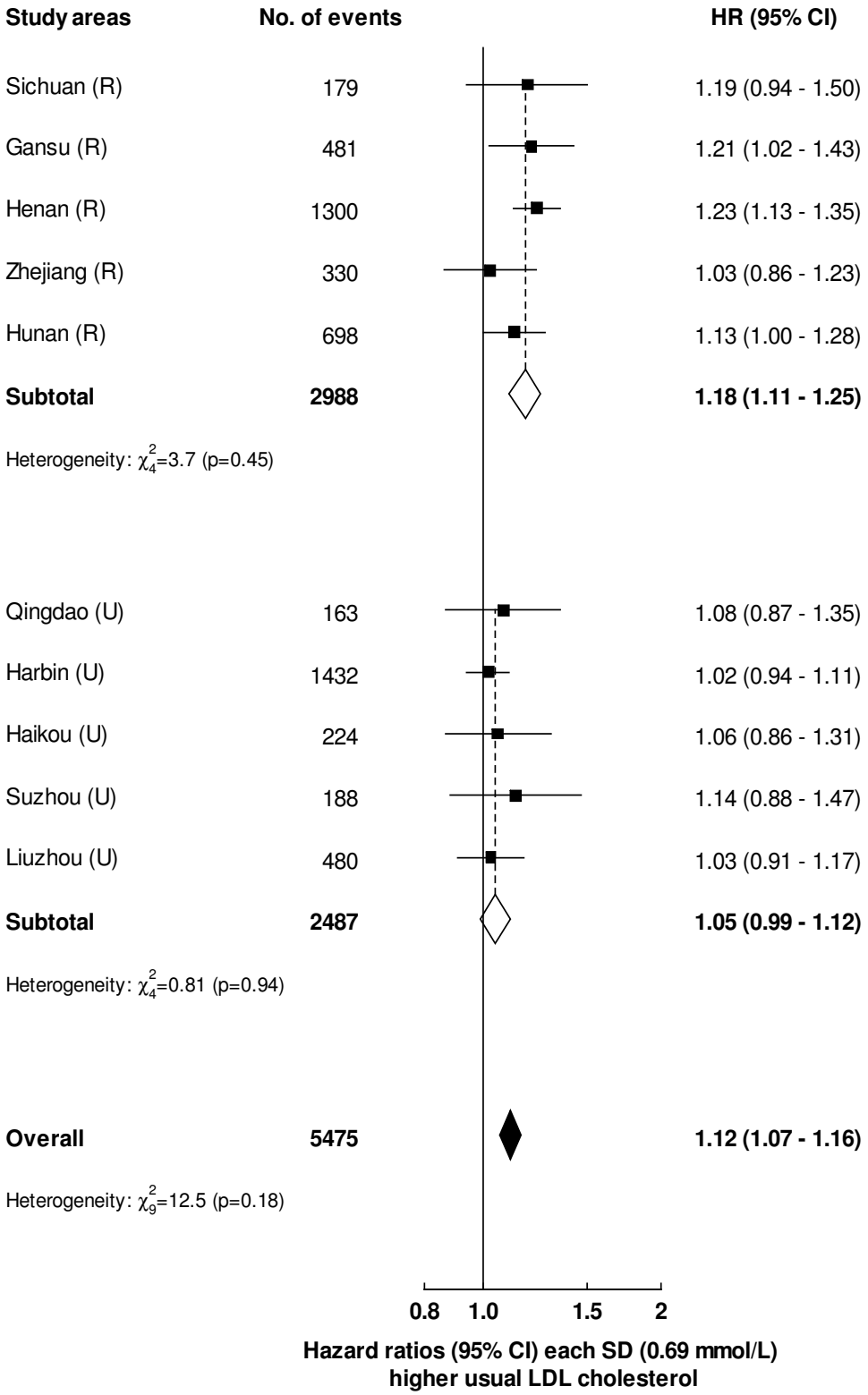
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 4.7: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol in population subgroups of baseline vascular risk factors (II)



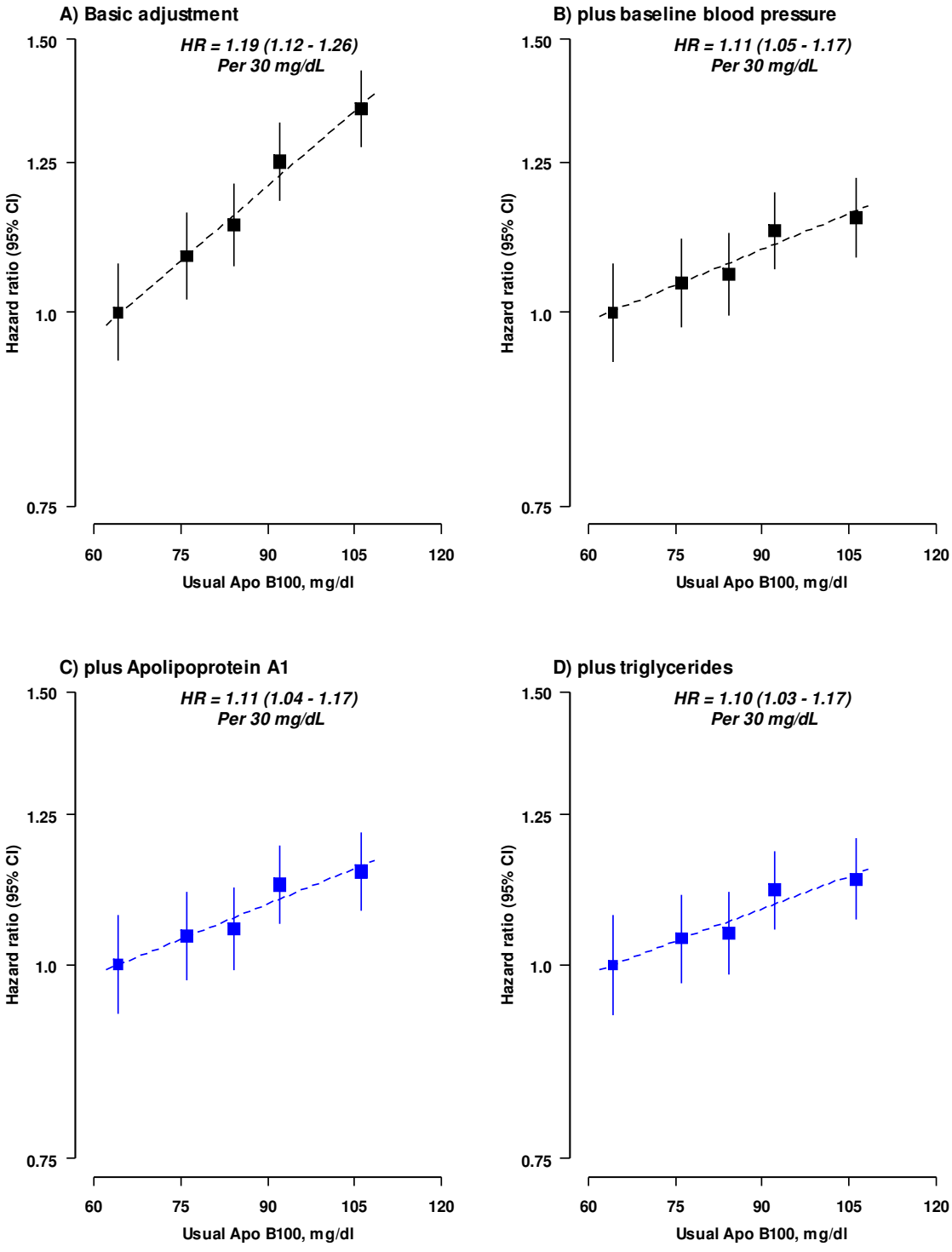
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.*Among those adjudicated ischaemic stroke cases (n=2304).

Figure 4.8: Adjusted hazards ratios (HRs) for risk of ischaemic stroke by usual LDL cholesterol, by region



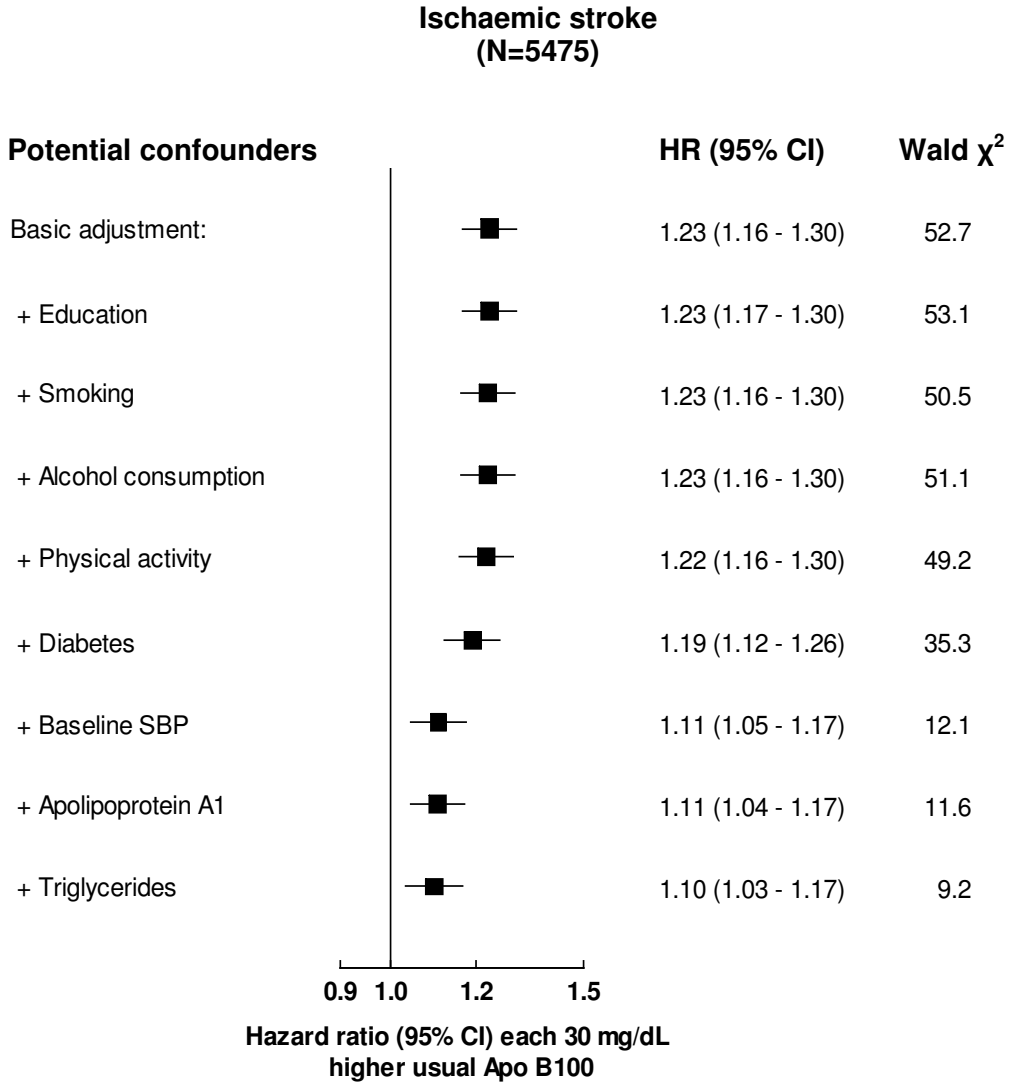
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 4.9: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein B100 (Apo B100)



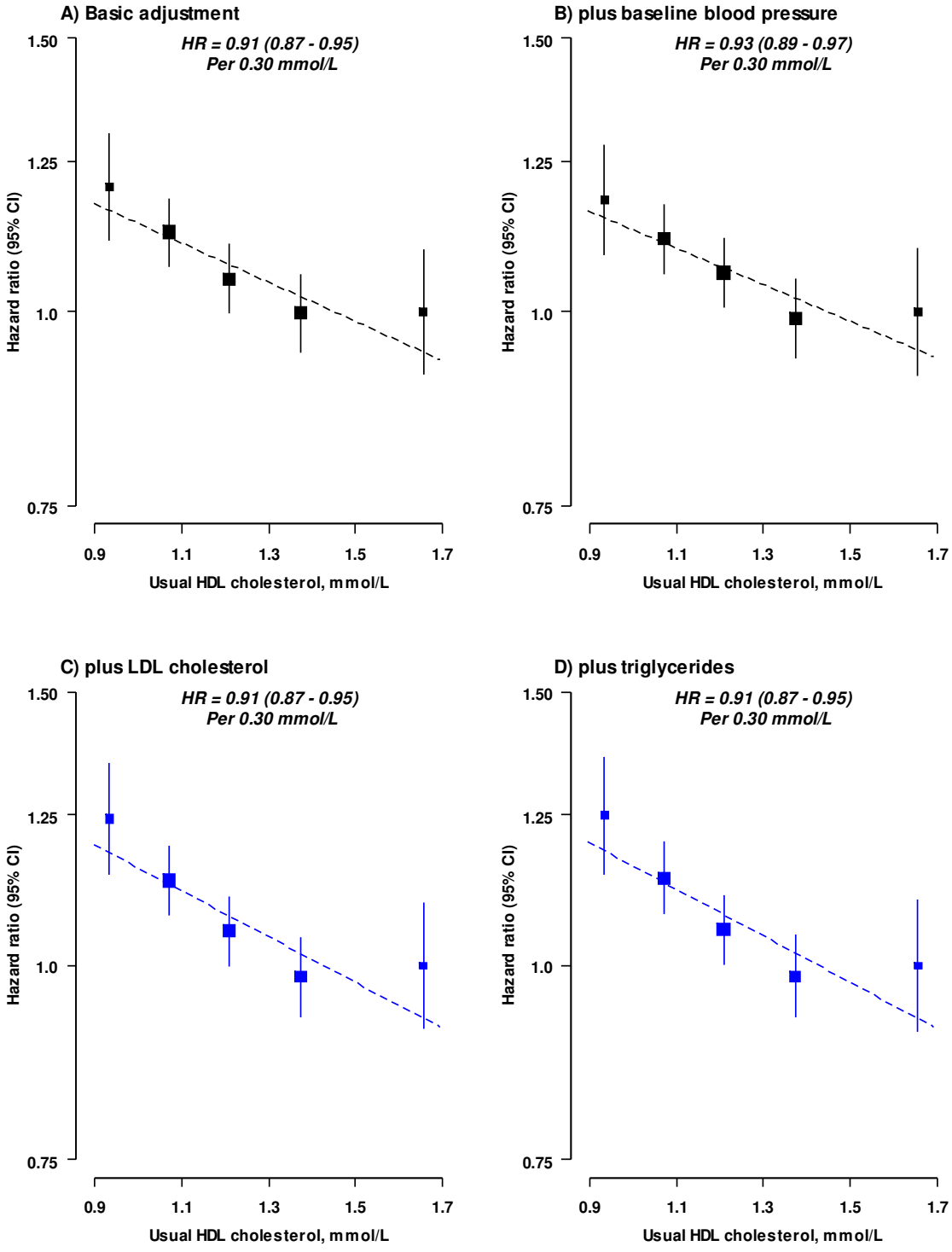
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for Apolipoprotein A1 in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual Apo B100 concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 4.10: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein B100 (Apo B100), progressively adjusted for potential confounding factors



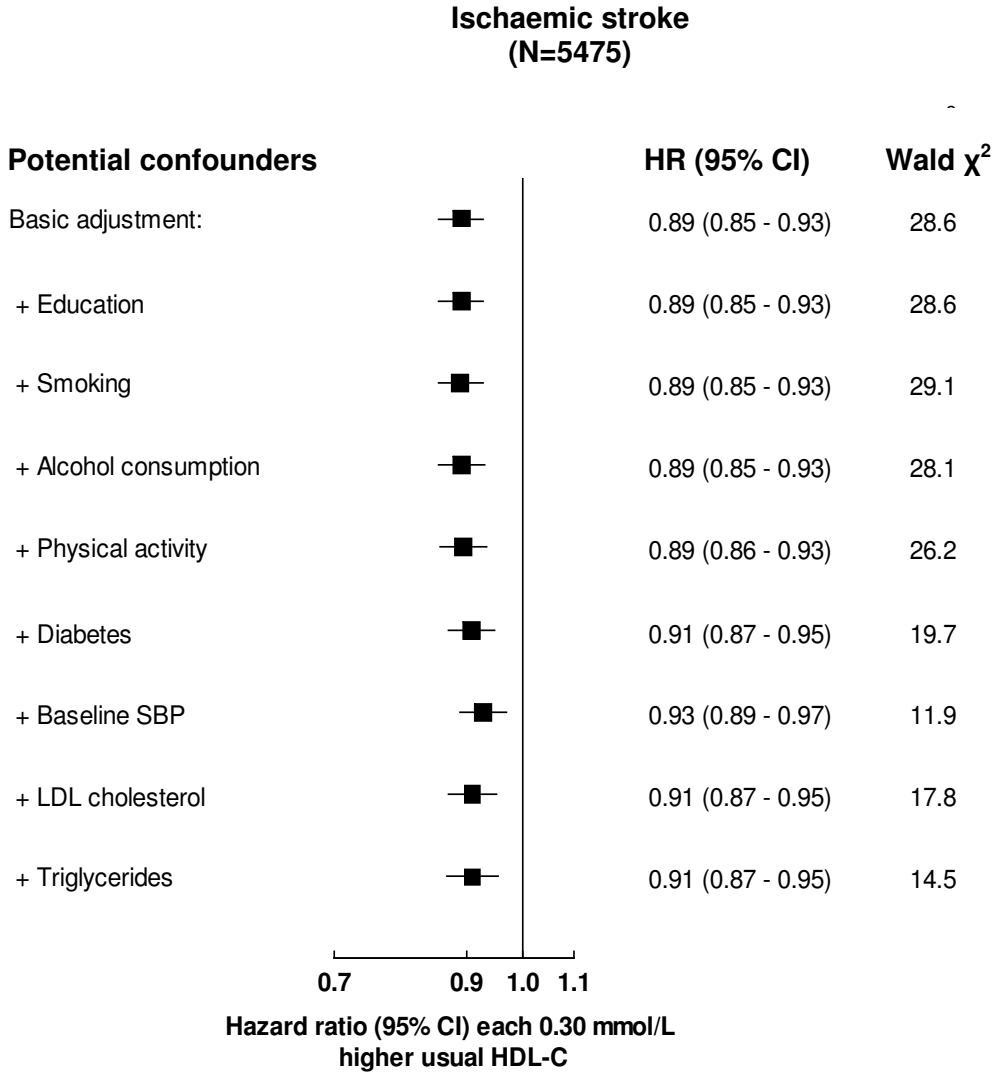
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual Apo B100, per 1 SD higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 4.11: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol (The panel with rectangle indicated the main results)



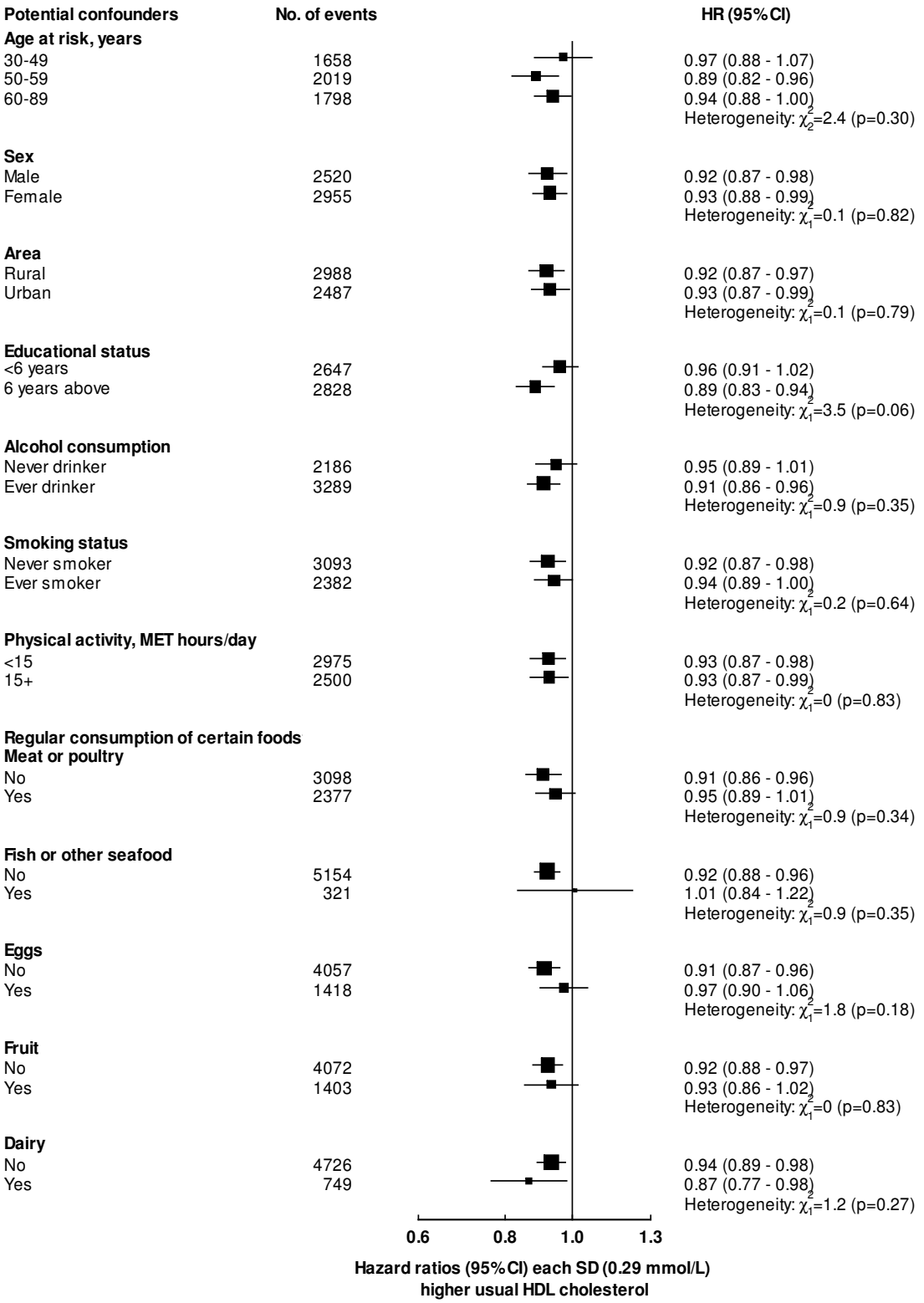
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for LDL cholesterol in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual HDL cholesterol concentrations in each category. Each square has an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 4.12: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol, progressively adjusted for potential confounding factors



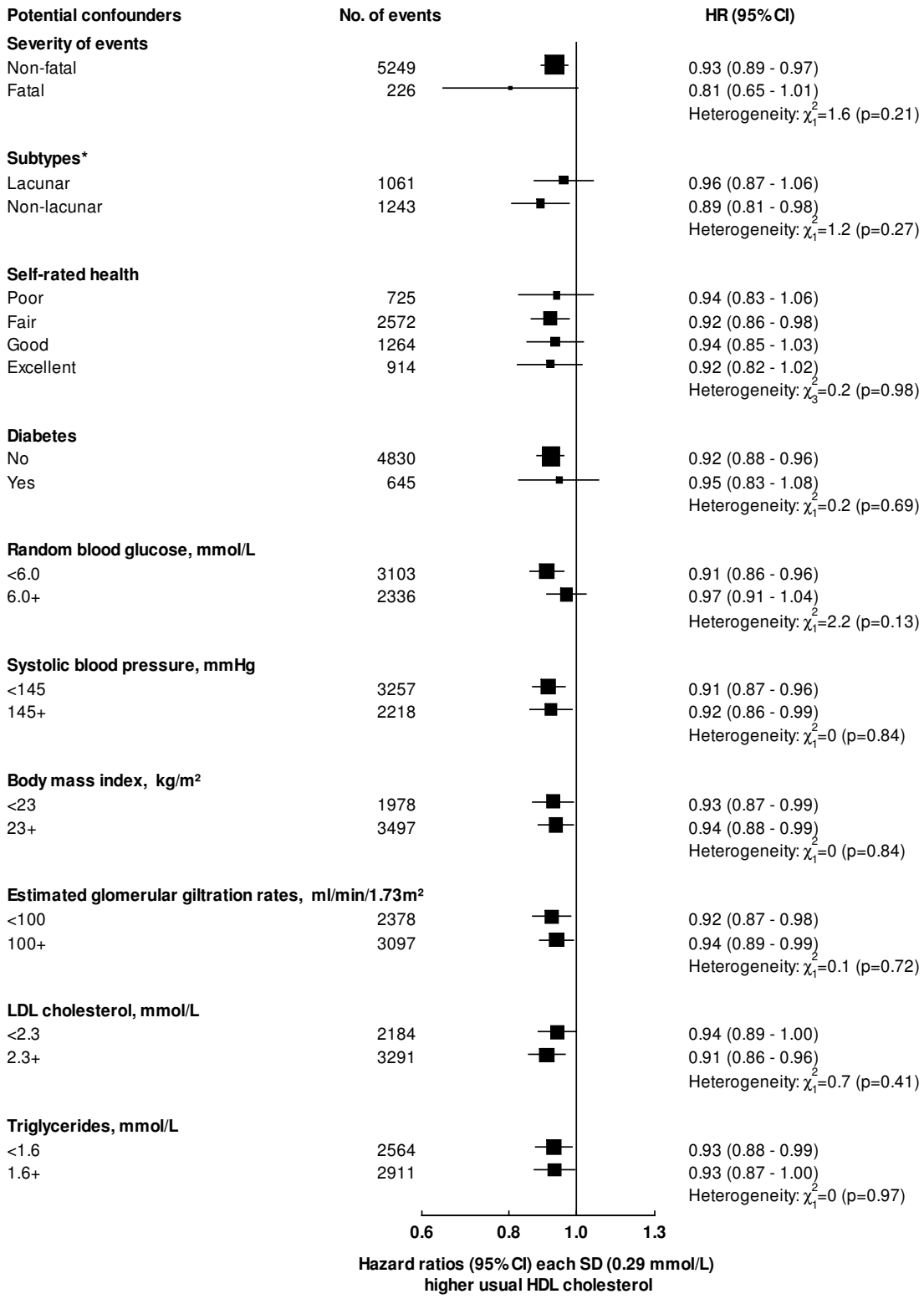
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual HDL cholesterol, per 0.3 mmol/L higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 4.13: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol in population subgroups of baseline vascular risk factors (I)



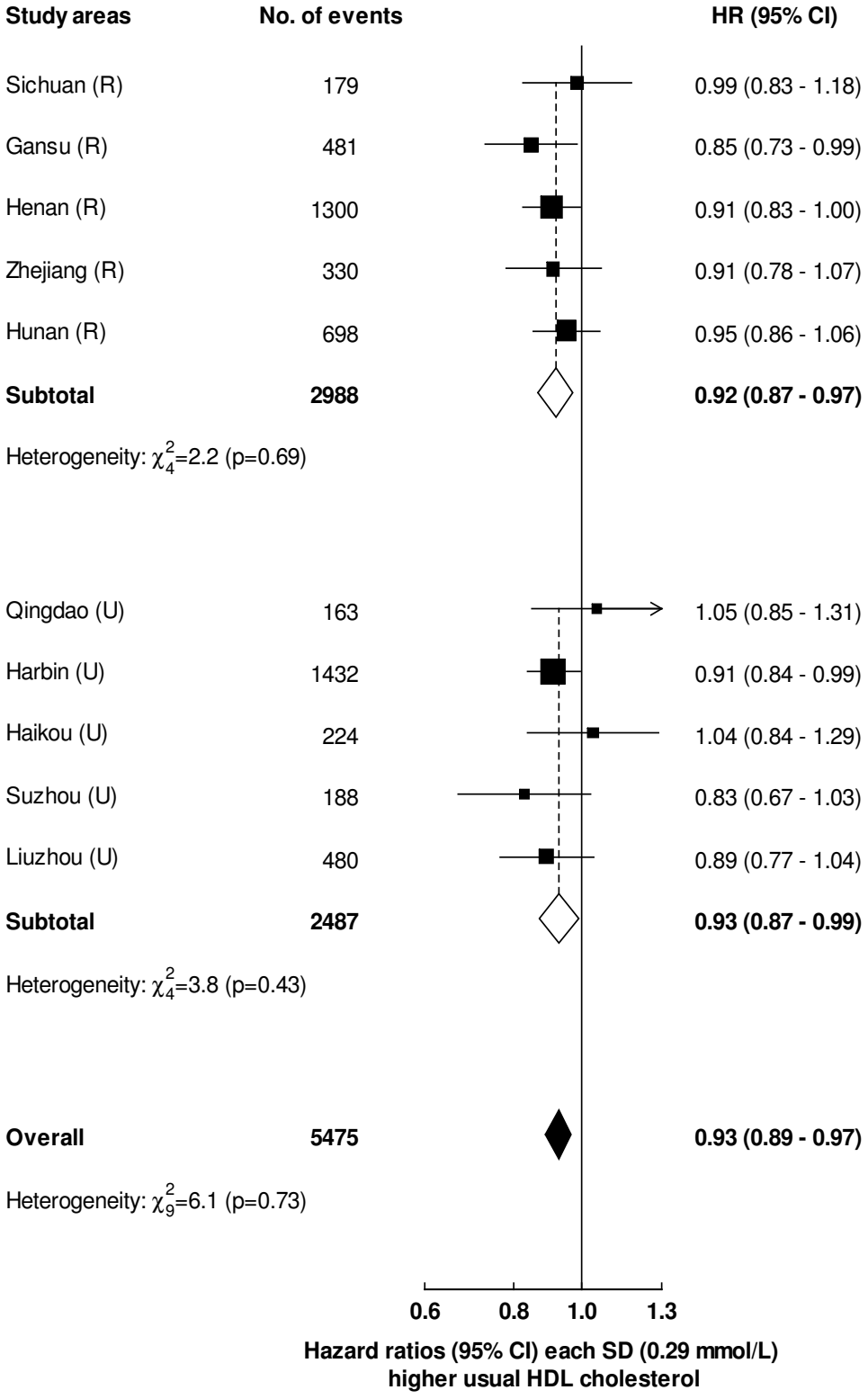
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual HDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 4.14: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol in population subgroups of baseline vascular risk factors (II)



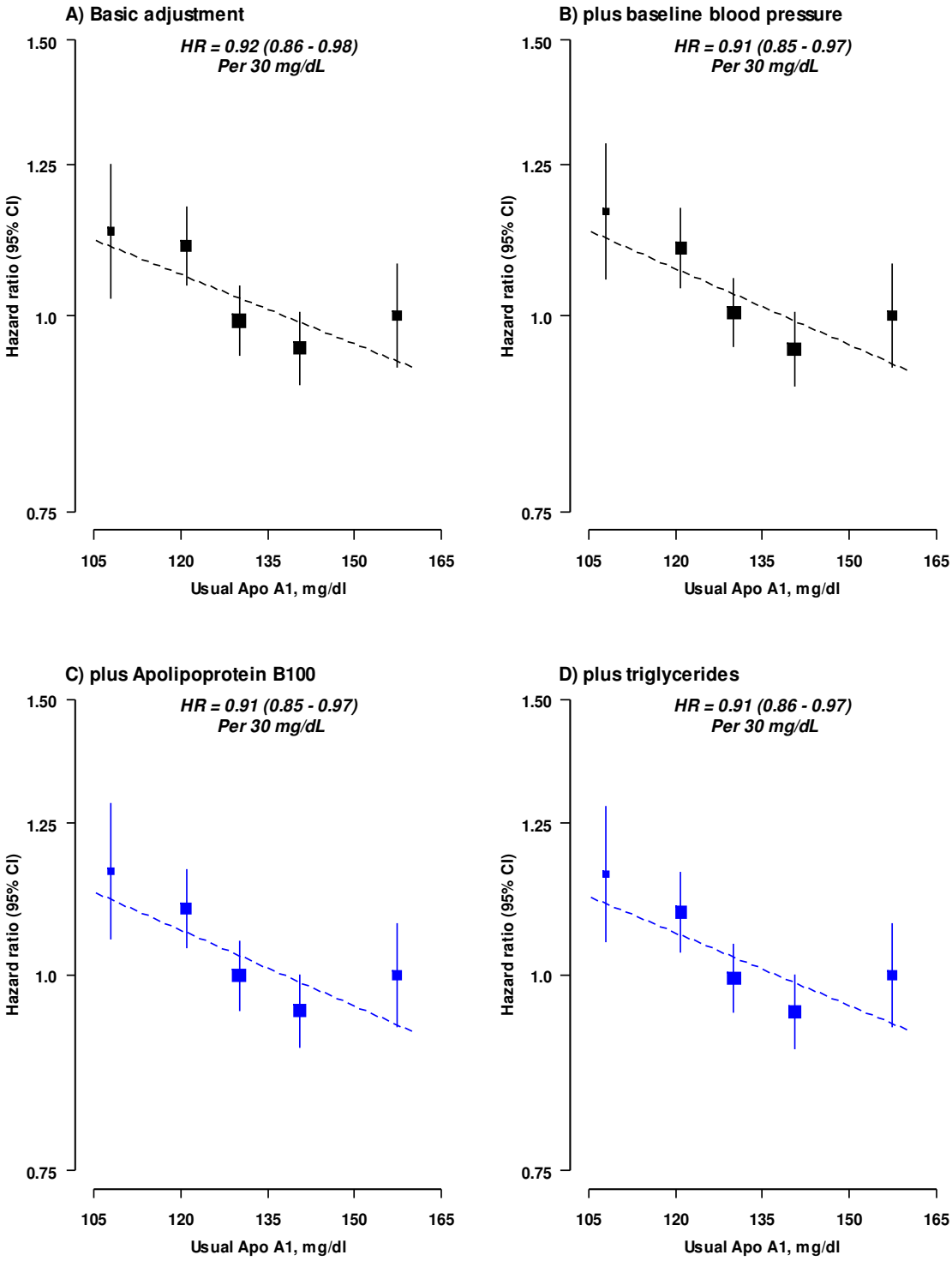
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against HDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.*Among those adjudicated ischaemic stroke cases (n=2304).

Figure 4.15: Adjusted hazards ratios (HRs) for risk of ischaemic stroke by usual HDL cholesterol, by study region



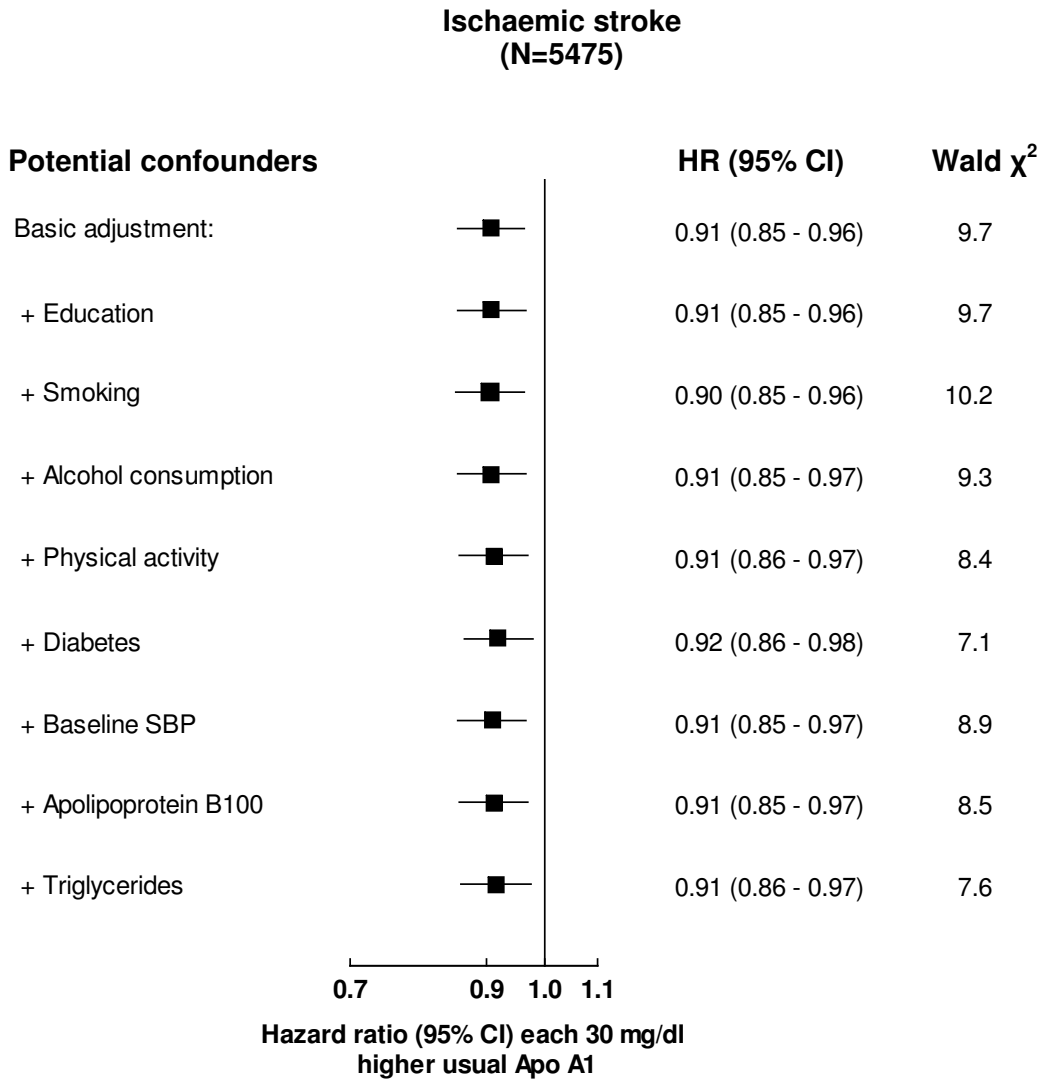
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual HDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 4.16: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein A1 (Apo A1)



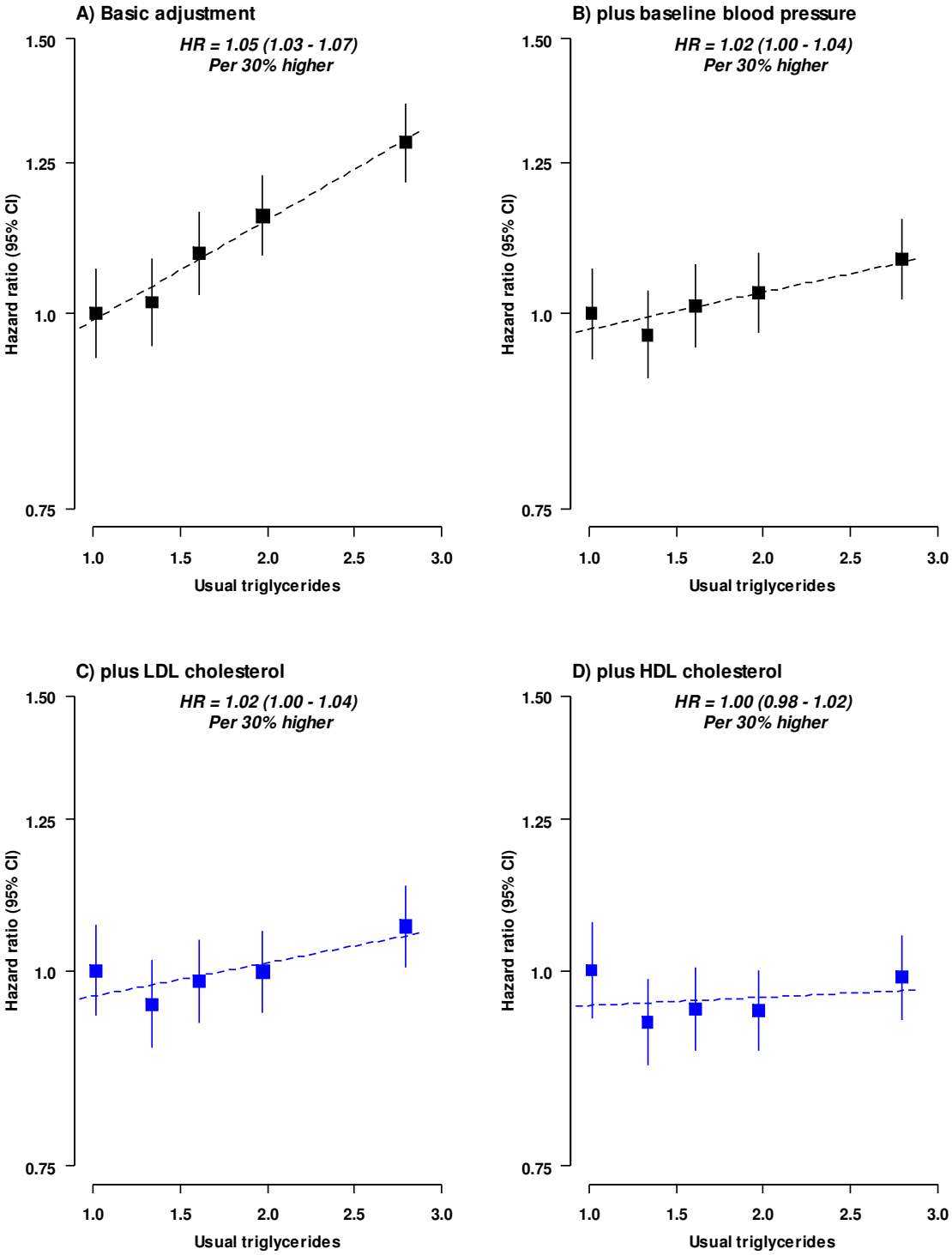
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for Apolipoprotein B100 in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual Apo A1 concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 4.17: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual Apolipoprotein A1 (Apo A1), progressively adjusted for potential confounding factors



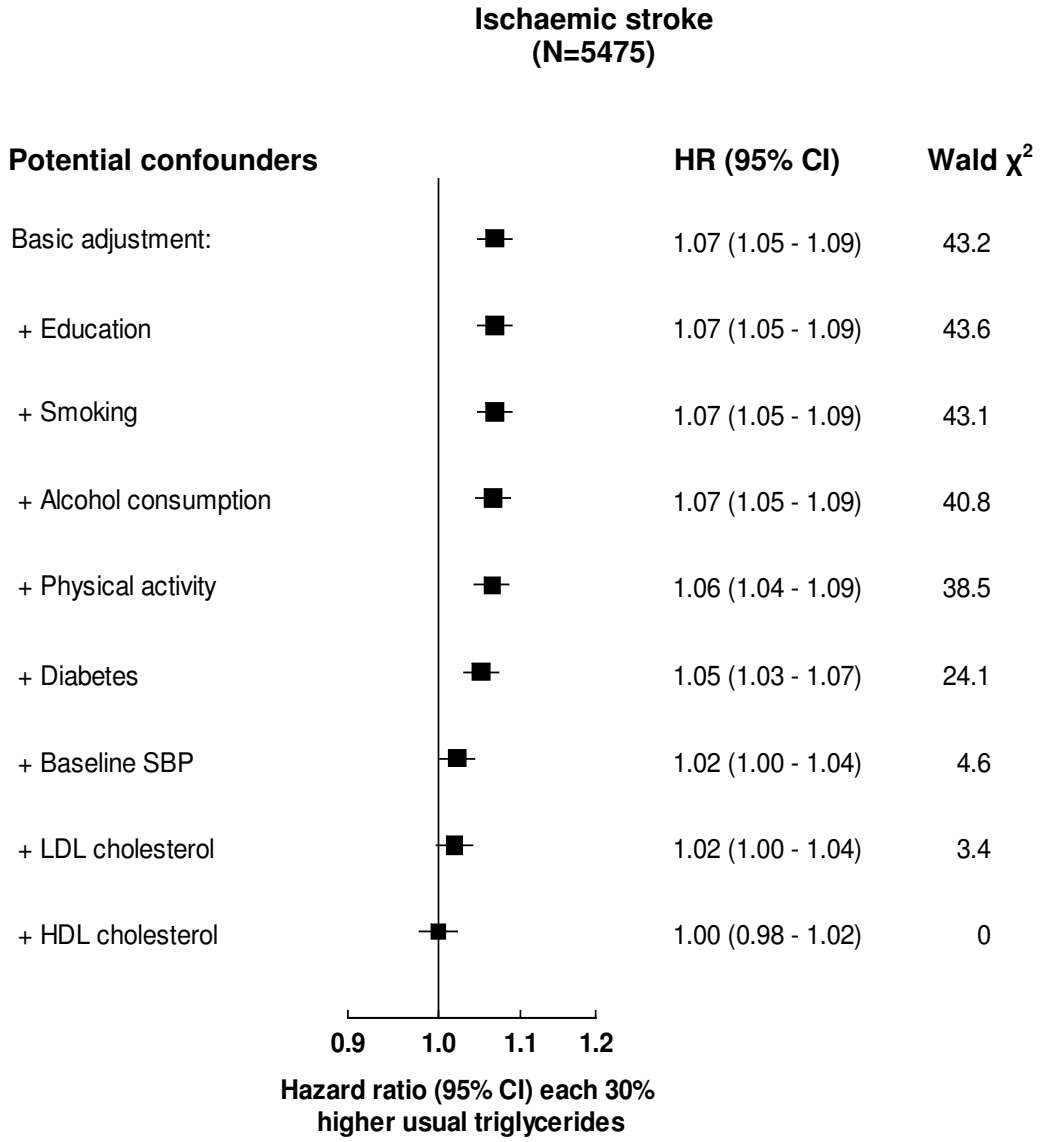
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual Apo A1, per 30 mg/dl higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represented the corresponding 95% CI.

Figure 4.18: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides



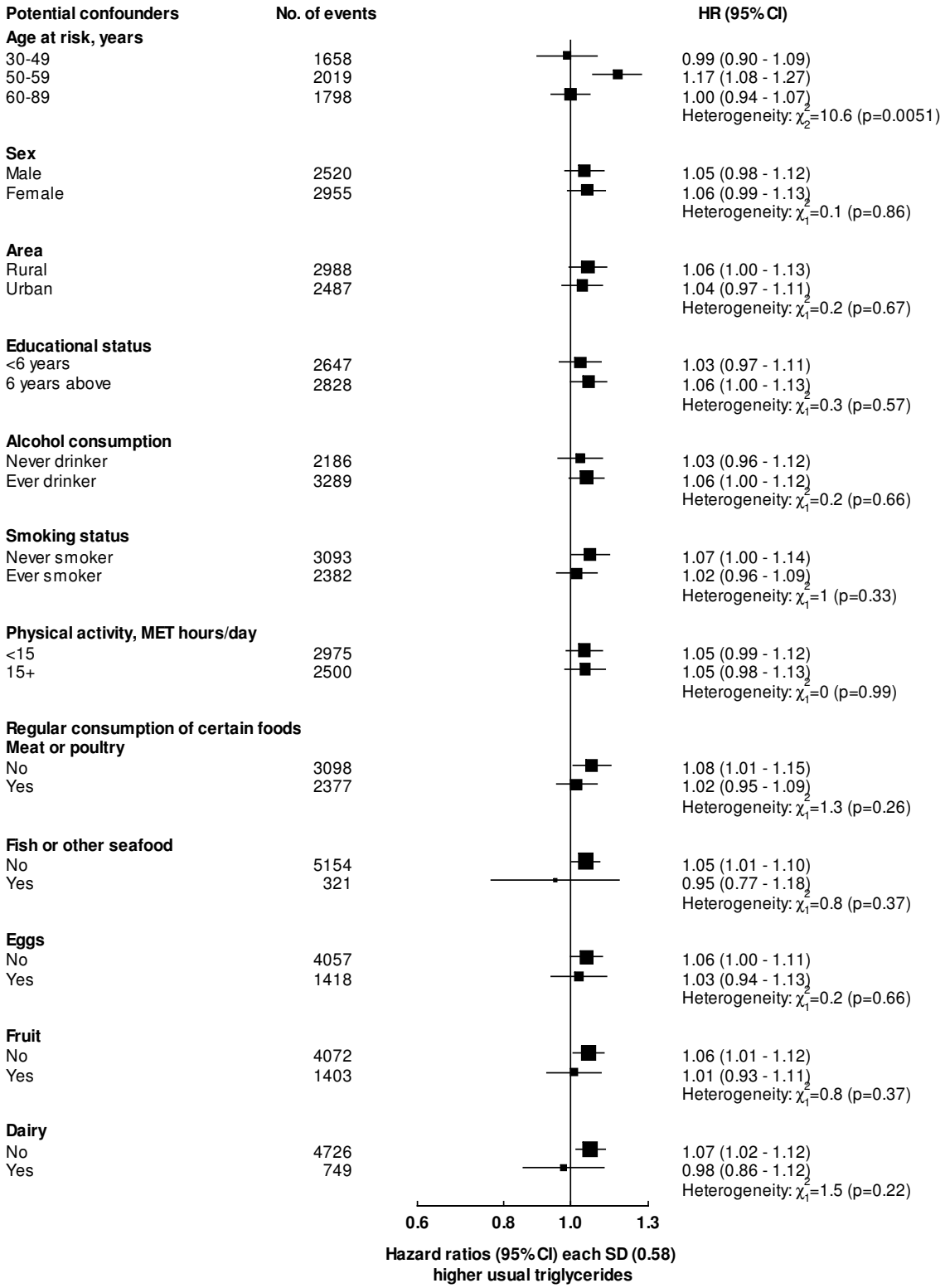
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for LDL cholesterol in addition to those in B); D) further adjusted for HDL cholesterol in addition to those in C). HRs were plotted against mean triglyceride concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 4.19: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides, progressively adjusted for potential confounding factors



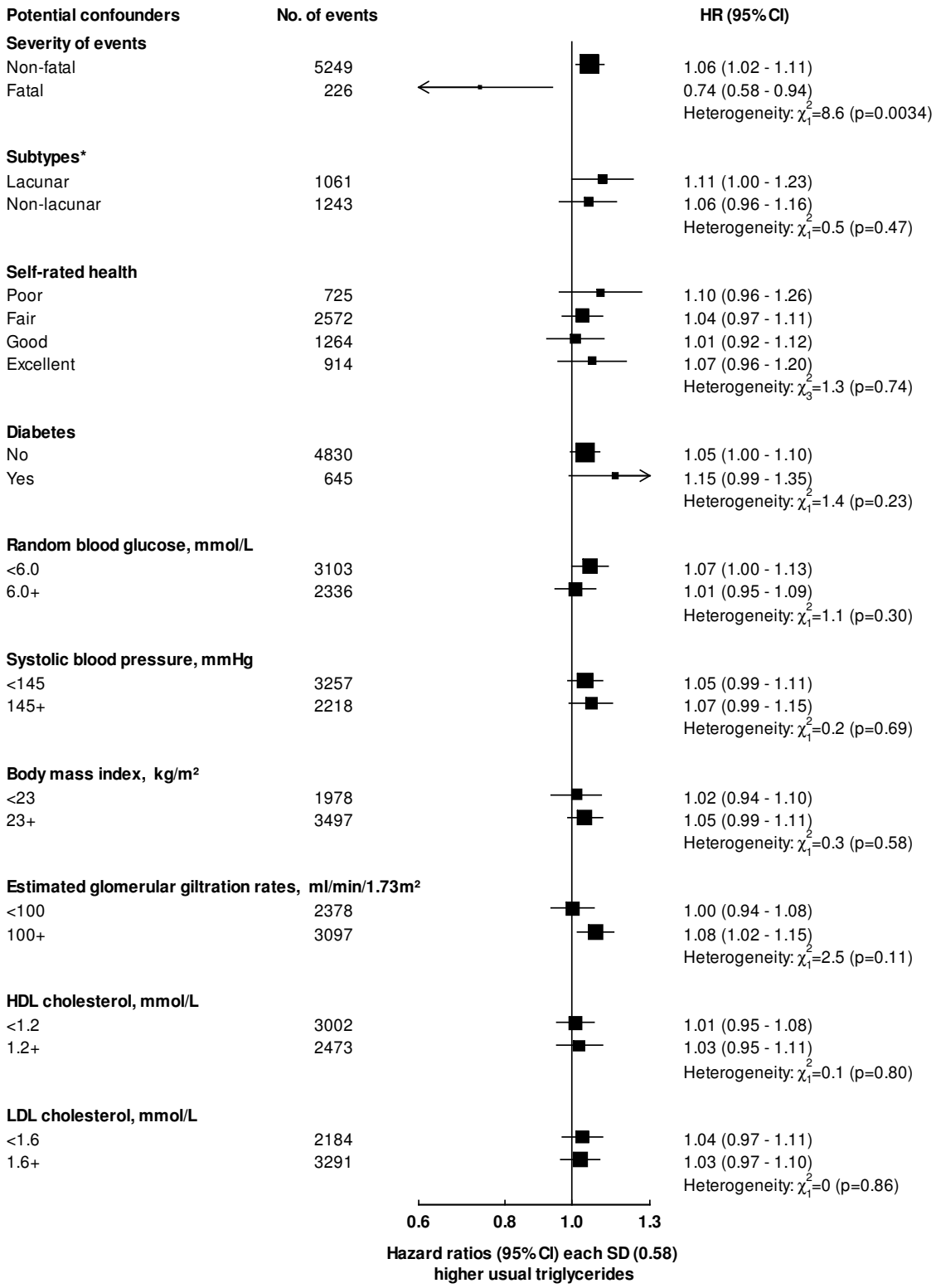
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual triglycerides, per 30% higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represent the hazard ratios with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 4.20: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides in population subgroups of baseline vascular risk factors (I)



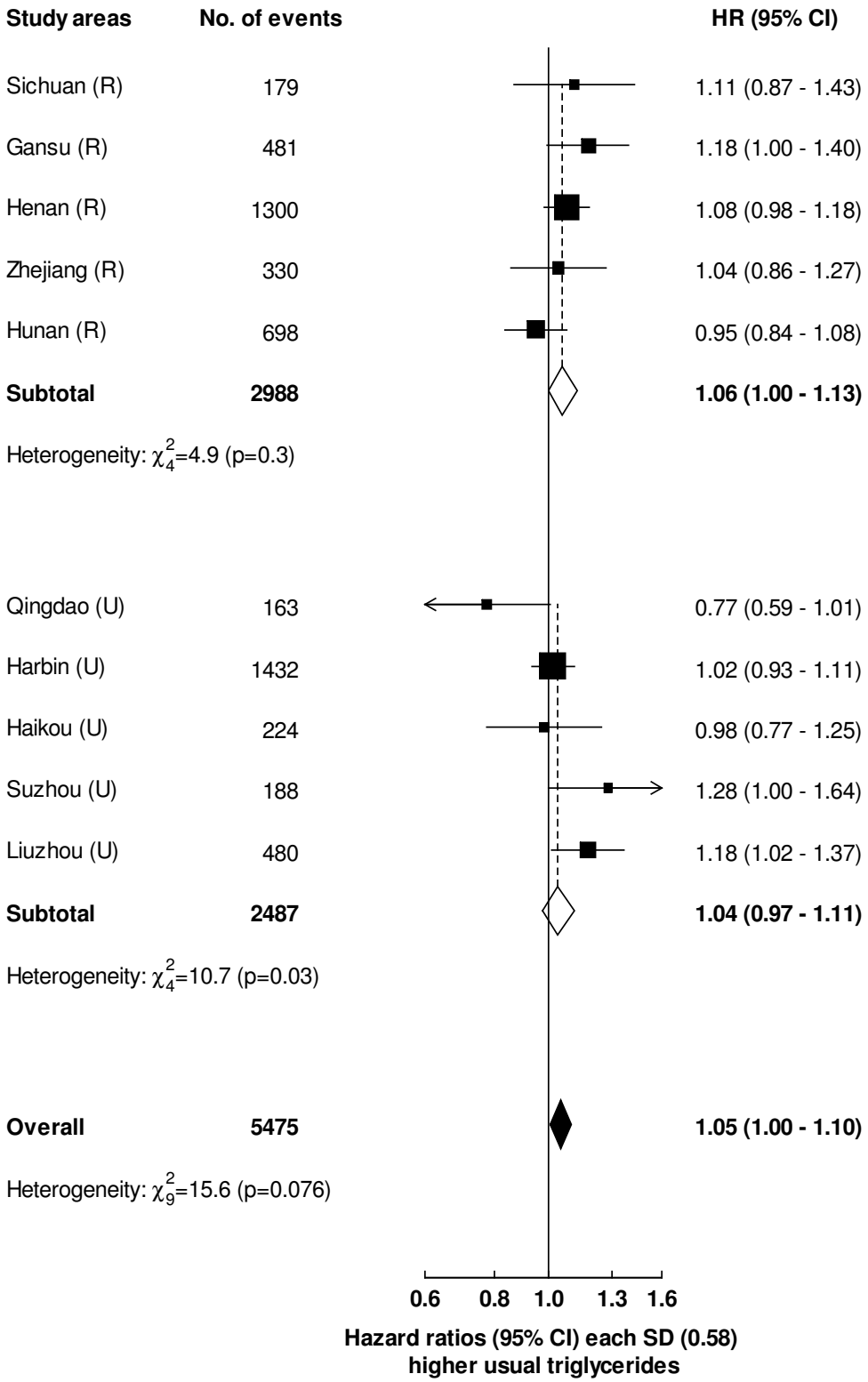
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual triglycerides (Log_e), per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 4.21: Adjusted hazard ratios (HRs) for risk of ischaemic stroke by usual triglycerides in population subgroups of baseline vascular risk factors (II)



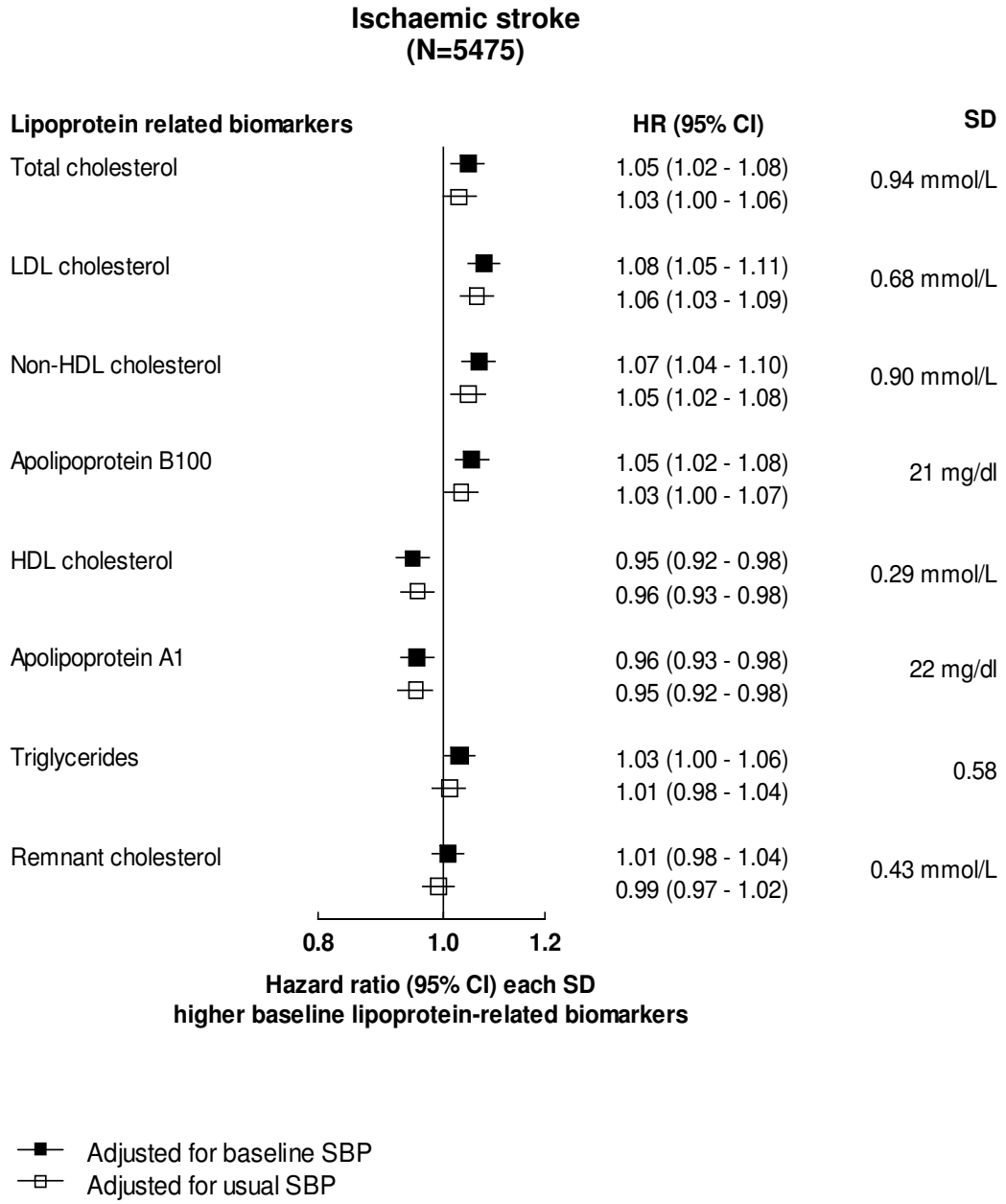
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual triglycerides (Log_e), per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.*Among those adjudicated ischaemic stroke cases (n=2304).

Figure 4.22: Adjusted hazards ratios (HRs) for risk of ischaemic stroke by usual triglycerides, by study region



SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual triglycerides (Log_e), per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 4.23: Comparison of associations of baseline lipid-related biomarkers with risk of ischaemic stroke, with adjustment for baseline or usual systolic blood pressure (SBP)



SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against baseline lipid-related biomarkers, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, diabetes, and baseline SBP, or usual SBP. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs

Chapter 5 Major lipids and risk of intracerebral haemorrhage

This chapter describes the incidence rates of ICH by age, sex, and region, in the overall CKB population. The chief objective of this chapter is to examine the shape and strength of associations of major blood lipids with risk of ICH in a nested case-control study of ICH, after controlling for confounding factors and correction for regression dilution bias in lipid measurements.

5.1 Introduction

HS, including ICH and SAH, accounts for approximately 10-25% of all strokes worldwide,[39] and because of its high case fatality rate, it contributes to a high proportion of stroke deaths.[39] Of HS, SAH accounts for only 5% of all strokes, but occurs at a fairly young age.[232] In 2015, there were around 3.6 million incident HS cases, over 3.3 million deaths, and 73.4 million DALYs lost due to HS, of which, 80% of the incident cases and 63% of the deaths occurred in LMICs.[4] These include China, which accounts for one-fifth of the world's population, and the epidemiology of HS has been characterised by higher incidence rates, and a greater proportion of all strokes, than in Western populations.[10, 52] In China, around 1 million of adult deaths (~9 million in total) were attributable to HS,[7] which accounted for one-third of the total HS deaths worldwide, in 2015.[3]

Compared to the evidence on the relevance of major blood lipids to ischaemic vascular disease, there is relatively limited evidence on about risk of ICH. In particular, in observational studies, concerns have been raised about the inverse associations of LDL-C with risk of ICH.[13, 108] Furthermore, available published evidence on the relevance of HDL-C and triglycerides for risk of ICH is extremely

limited. Hence, appropriately designed studies involving large numbers of well-phenotyped HS cases are needed to address these questions.

Therefore, in this chapter, the entire 0.5 million CKB prospective study participants were used describe the rates of ICH in different settings. The nested case-control study of ICH within the CKB prospective study, involving about 5000 incident ICH cases, and 6000 healthy controls, was then used to assess the associations of major blood lipids and apolipoproteins, with ICH risk, after adjustment for various potential confounders, and correction for regression dilution bias.

5.2 Study design and methods

The description of the incidence rates of ICH were performed in the overall CKB population among those with no prior history of cardiovascular events (i.e., CHD, stroke, or TIA) at recruitment. The overall study population characteristics and design of the CKB prospective study have been reported in detail in previous publications,[156, 157] and have been summarised in **Chapter 3**. The analyses of prospective associations of major blood lipids with risk of ICH were undertaken in the nested case-control study of ICH.

5.2.1 The nested case-control study

The nested case-control study of stroke pathological types, as previously described (**Chapters 3-4**), was designed to answer multiple research questions, with a priority to improve the understanding of the genetic and non-genetic determinants of ICH. The ICH cases and controls were originally selected with a censoring date to 1st January 2014 (when the study was initially designed), among the CKB participants who: a) did not report a doctor-diagnosed CHD,

stroke, TIA, or cancer at baseline; b) did not report use of statins at baseline. Following these criteria, in the overall CKB study population (n=489 762), around 6000 incident ICH events were identified. If one case had both IS and ICH, the stroke cases which occurred earlier was selected. Controls were selected from individuals who had not developed a stroke of any type, or unspecified type, or MI, or CHD by the censoring date, and backwards in time to match ICH cases (ordered by the reverse of the dates of event[166]) by year of birth, sex, and study region. The same controls were used for both IS (see **Chapter 4**) and ICH.

Similar to the analyses on IS (**Chapter 4**), at the time of writing this thesis, the disease definition (i.e., ICH case or control) was re-identified by extending the censoring date to 1st January 2016. With the new disease definitions, those who developed MI, SAH, and other stroke of unknown types, prior to any ICH events were excluded from the subsequent analyses. A few individuals who were initially selected as controls subsequently developed ICH events during the extended follow-up period (n=27), and the disease status of such individuals was changed hence. Due to the initial linkage errors in the health insurance claims database, 111 several previously-defined ICH cases were re-defined as controls (**Table 4.1**).

5.2.2 Statistical analyses

5.2.2.1 Selection of baseline variables

Based on the prior biological knowledge of the associations of major blood lipids with risk of ICH, and the **Literature Review** (**Chapter 2**), the following baseline variables were selected: measurements of TC, LDL-C, HDL-C, triglycerides, Apo A1, Apo B100; age, sex, study region, education, household income, smoking status, alcohol consumption, information on dietary patterns (i.e., regular

consumption of fresh vegetables, meat or poultry, fish or other seafood, egg, soybean, fresh fruit, and dairy products), physical activity (MET hours /day), self-reported health conditions (i.e., self-rated health, diabetes, hypertension, and blood pressure lowering medication); measurements of SBP, BMI, waist circumference, hip circumference, body fat percentage, RBG, and eGFR.

5.2.2.2 Participants exclusions

As described in **Section 4.2.2**, for the description of incidence rates of ICH, participants with a prior history of cardiovascular diseases (CHD, stroke, or TIA) were excluded. For the investigation of the prospective associations of major blood lipids with risk of ICH, participants with missing or unreliable lipid measurements or other covariates were excluded. After this, 489 762 participants, and 11 066 participants (including 4776 ICH cases, and 6290 healthy controls) remained in the subsequent analyses.

5.2.2.3 Definitions and categorisations of baseline variables

The definitions and categorisations of the baseline variables included in the following descriptive or association analyses were as previous described in **Chapter 4** (**Section 4.2.2.3**).

5.2.2.4 Incidence rates of intracerebral haemorrhage in CKB participants

Description of the incidence rates of ICH was performed on the overall CKB study population, from the baseline survey (2004-2008) up to a censoring date of 1st January 2016. Within the 4.4 million person-years' follow-up, 8270 incident ICH cases were identified. The incidence rates were reported, separately, by age-at-risk (5-year group), sex, and region. The crude incidence rates were

calculated as the ratio of the number of ICH cases to the total person-years in each age-at-risk group. The overall incidence rates were estimated by calculating the unweighted average of each component (i.e., age-range or study area in men and women combined) incidence rates.

5.2.2.5 Baseline characteristics of participants in the nested case-control study

The distribution of baseline characteristics of study participants were described using means or proportions by disease status (i.e., ICH cases and controls), directly standardised by sex, age (10-year age group), and region of the study population, where appropriate. Distribution of baseline characteristics of study participants by lipid measurements were estimated on 11 066 individuals, directly standardised by disease status, sex, age (10-year age group), and region of the study population.

5.2.2.6 Prospective associations of major blood lipids with intracerebral haemorrhage

Investigation of the shape and strength of associations of major blood lipids with risk of ICH were similar to that previous described for IS (**Chapter 4: Section 4.2.2.8**). Briefly, Cox regression was used to estimate the multi-covariate-adjusted HRs and the corresponding 95% CIs for risk of ICH associated with usual lipid values. The analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education (no formal education, primary school, secondary school, high school or above), smoking (never, occasional, ex-regular, or current), alcohol consumption (never, ex-regular, occasional, or current), physical activity (<10, 10-<20, 20-<30, 30-<40, or 40+ MET hours /day), and baseline diabetes status

(yes/no). The additional adjustment for baseline SBP was assessed and reported separately, where appropriate. The identical RDRs as estimated in **Chapter 4** (**Section 4.2.2.7**) were applied for the prospective associations of major blood lipids with risk of ICH. As is the case in IS, the primary association analyses were not adjusted for adiposity, since adiposity acts as a determinant of lipids, and other vascular factors (e.g., diabetes), and it is therefore of little independent value in the multivariate regression analyses. Sensitivity analyses were reported by comparing the risk estimates from models with and without adjustment for BMI. The comparison of effects of usual SBP had on the associations was reported in sensitivity analyses. Wald χ^2 statistics and their corresponding *p* values were used to assess evidence of association after the sequential addition of each potential confounder to the association analyses of lipids with risk of ICH. Sensitivity analyses using logistic regression gave similar effect estimates yielded from Cox regression.

All statistical analyses were conducted in SAS version 9.3 (SAS Institute Inc., Cary, North Carolina, USA). Plots were produced using R version 3.2. Of all the analyses and plots, the Cox regression was using a macro developed within CTSU. The log-linear and forest plots were produced using an R package developed within CTSU).

5.3 Results

5.3.1 Incidence rates of intracerebral haemorrhage in CKB participants

Estimations of incidence rates of ICH were conducted on 489 762 participants who had no prior history of cardiovascular disease (i.e., CHD, stroke or TIA) at

baseline, from the recruitment up to the censoring date of 1st January 2016. Among the participants included, 59% were women, and the mean age for men and women were 52 years and 51 years, respectively. The mean SBP at baseline was 130.6 (SD: 19.3) mmHg.

Within a follow-up period of 4.4 million person-years, 8270 incident ICH cases were identified, constituting approximately 20% of the two stroke pathological types combined (8270 ICH and 32 869 IS cases). The overall incidence rates of ICH at age 30-89 years in this population were 187 cases per 100 000 person-years, which approximately doubled with each increasing decade of age: the incidence rates were 53, 85, 181, and 596 cases per 100 000 person-years, respectively, at ages below 50, 50-59, 60-69, and 70 years or above (**Table 5.1**). In each age group, ICH incidence rates were moderately higher in men than in women, with more extreme sex-differences observed at older ages. Excluding the few individuals aged older 75 years, the incidence rates of ICH in this Chinese population aged 30-75 years were 160 cases per 100 000 person-years, with 208 in men and 128 in women, respectively.

Considerable regional variation in ICH incidence rates was also observed in this Chinese population, with a 5-fold difference between study regions: the annual incidence rates of ICH at ages 30-90 years, ranged from 84 in Qingdao to 425 in Hunan per 100 000 person-years in men and 51 in Haikou to 294 in Gansu in women. The ICH incidence rates were slightly higher in Northern regions than in Southern regions. The Northern-Southern gradient was less extreme than that observed for IS. Incidence rates of ICH in rural areas were around double those in

urban areas, as opposite to IS. Consistent with the overall pattern, incidence rates of ICH were higher in men than that in women, in each study region.

5.3.2 Baseline characteristics of participants in the nested case-control study

Descriptive analyses were conducted on 11 066 participants, including 4776 ICH cases and 6290 controls. In this sample population, directly standardised to the age, sex, and study region structure of the study population included, ICH cases were slightly older than the controls (59.1 [SD: 10.0] versus 56.9 [10.8] years) (**Table 5.2**). The proportions of women, and urban residents were similar between ICH cases and controls. Overall, approximately one-third of the participants had completed secondary school or above, 29.6% in ICH cases, and 32.4% in controls, respectively.

Overall, the distribution characteristics of smoking and alcohol consumption, and dietary patterns were similar between ICH cases and controls, whilst, sex-difference was noted in several major lifestyle risk factors. In total, around one-third of the participants (33.7%) were current-smokers, much higher in men than in women (61.5% versus 3.4%). About 15.8% of the participants drank alcohol at least once a week on a regular basis (28.4% in men versus 2.1% in women). Almost all the participants consumed fresh vegetables daily, but daily fresh fruit consumption was less than 20%. There were about one-third, and one-fifth of this sample population who consumed meat, or egg products daily, and the proportions were higher in controls than in ICH cases. Regular consumption of other food groups (e.g., fish or seafood, soybean, and dairy products) was not

common in both ICH cases and controls, and the proportions were consistently lower in ICH cases than those observed in controls.

At baseline, about 15.9% of ICH cases rated their health status as poor, compared to a 9.6% of controls. The prevalence of diabetes was 2-fold greater in ICH cases compared with controls (8.1% versus 4.8%), as was the prevalence of hypertension (25.9% versus 10.6%).

At baseline, the overall distributions of adiposity-related anthropometric measurements were similar between ICH cases and controls (**Table 5.3**). Notably, mean levels of baseline SBP were substantially higher in ICH cases than in controls (153 versus 135 mm Hg). Among those with baseline measurement of RBG in the present study (i.e., 4670 ICH cases and 6218 controls), the mean plasma RBG concentrations were slightly higher in ICH cases than in controls (6.4 versus 6.0 mmol/L). Excluding individuals with potential unreliable measurements of creatinine (above 176 umol/L), the mean eGFR was moderately lower in ICH cases than in controls (96.4 versus 98.3 ml/min/1.73m^2). There were only small differences in levels of physical activity between ICH cases and controls (18.1 versus 19.1 MET hours /day).

Overall, there was little difference in plasma concentrations of lipid-related biomarkers between ICH cases and controls (**Table 5.4**). After adjustment for age, sex, study region, and disease status, the mean (SD) concentrations of TC were 4.5 (0.9) mmol/L, LDL-C were 2.3 (0.7) mmol/L, HDL-C were 1.3 (0.3) mmol/L, and Apo A1 were 135.6 (23.2) mg/dl, Apo B100 were 81.3 (20.3) mg/dl. The median [IQR] of plasma triglyceride concentrations were 1.5 (1.3) mmol/L.

5.3.3 Prospective associations of major lipids with intracerebral haemorrhage

As for IS (**Chapter 4**), given SBP is one of the strongest risk factor of ICH, and the substantial effect that SBP had on the relevance of major lipids for ICH, the main results described and discussed in this thesis were on the SBP adjusted results, unless otherwise stated.

5.3.3.1 Low-density lipoprotein cholesterol, Apolipoprotein B100

There was a log-linear inverse association of usual LDL-C with risk of ICH after stratification by age-at-risk, sex, study region, and adjustment for education, smoking, alcohol consumption, physical activity, baseline diabetes, and SBP (**Figure 5.1**). Across the range of usual LDL-C (approximately 1.7-3.2 mmol/L) examined, each 1 mmol/L higher usual LDL-C concentration was associated with a 14% (95% CI: 8-20%) lower risk of ICH. Compared to individuals in the top quintile of usual LDL-C (mean: 3.0 mmol/L), the adjusted HRs of risk of ICH were 1.22 (95% CI: 1.15-1.31), 1.13 (1.06-1.20), 1.04 (0.98-1.12), and 1.04 (0.98-1.11) for each consecutive higher quintile group, with mean usual LDL-C concentrations of 1.7 mmol/L, 2.0 mmol/L, 2.3 mmol/L, and 2.5 mmol/L, respectively.

Figure 5.2 shows the differences in the strength of association of usual LDL-C with risk of ICH, after sequential adjustment for various vascular risk factors. When stratified by age-at-risk (5-years), sex, and study region only, there was a marginal inverse association of usual LDL-C with risk of ICH (HR 0.94; 95% CI: 0.87-1.00 per 1 mmol/L higher). The HRs were altered marginally after sequential adjustment for other vascular risk factors (i.e., education, smoking, alcohol consumption, physical activity, and baseline diabetes), with the Wald χ^2 increasing

from 4.5 to 6.9. The inverse association was, however, apparent with further adjustment for baseline SBP (HR 0.86; 95% CI: 0.80-0.92 per 1 mmol/L higher). The inverse association of usual LDL-C with risk of ICH was independent of HDL-C (HR 0.86; 95% CI: 0.80-0.92), and of triglycerides (HR 0.89; 95% CI: 0.83-0.95).

Apart from the moderate heterogeneity demonstrated in associations of usual LDL-C with ICH risk due to severity of disease (fatal versus non-fatal cases; $p_{heterogeneity}$ =0.005), BMI (23 kg/m^2 above versus below; $p_{heterogeneity}$ =0.001), and triglycerides (1.7 mmol/L above versus below; $p_{heterogeneity}$ =0.002), no substantial heterogeneity was observed by subgroups of other vascular risk factors, including SBP (145 mm Hg above versus below; $p_{heterogeneity}$ =0.25) (**Figure 5.3-5**). There was a significant inverse association of LDL-C with risk of fatal ICH (n=179; HR 1.18; 95% CI: 1.12-1.23 per 1 SD higher), compared to a marginally significant lower risk of non-fatal ICH cases (n=4597; HR 0.94; 95% CI: 0.88-1.00).

Consistent with those risks observed for LDL-C, there was a log-linear inverse association of Apo B100 with risk of ICH (HR 0.80; 95% CI: 0.75-0.85 per 30 mg/dl higher), which was independent of Apo A1 (**Figure 5.6**). In progressive adjustment for various vascular risk factors (**Figure 5.7**), the strength of the associations of usual Apo B100 with risk of ICH was chiefly affected by adjustment for baseline SBP, with multivariable-adjusted HRs being 0.86 (95% CI: 0.81-0.92) and 0.80 (95% CI: 0.75-0.85), before and after adjustment, respectively. Triglyceride-rich lipoproteins also contain Apo B100, and as expected, additional adjustment for triglycerides moderately attenuated the strength of the association of Apo B100 with risk of ICH (HR 0.83; 95% CI: 0.78-0.89 per 30 mg/dl higher).

5.3.3.2 High-density lipoprotein cholesterol, Apolipoprotein A1

In this population with relatively low LDL-C, there was no evidence of association of usual HDL-C with risk of ICH (**Figure 5.8-9**). The association of HDL-C with risk of ICH remained non-significant after extensive adjustment for various vascular risk factors (**Figure 5.9**), nor were there any apparent effect modification or interaction in various population subgroups (**Figure 5.10-12**). The presence of a significant association between Apo A1 and risk of ICH (**Figure 5.13**) was largely driven by the bottom quintile group (**Figure 5.14**).

5.3.3.3 Triglycerides

There was a log-linear inverse association of triglycerides (Log_e) with risk of ICH, with a multivariable-adjusted HR of 0.94 (95% CI: 0.92-0.96) associated with per 30% higher triglycerides, after stratification by age-at-risk (5-year), sex, study region, and adjustment for education, smoking, alcohol consumption, physical activity, diabetes, and baseline SBP (**Figure 5.15**).

The inverse association of triglycerides with ICH risk was only apparent after the adjustment for baseline SBP (**Figure 5.16**). The log-linear inverse association of triglycerides with risk of ICH was not explained by the correlations between triglycerides and LDL-C or HDL-C, with multi-covariate-adjusted HRs of 0.94 (95% CI: 0.92-0.96) for LDL-C, and 0.93 (95% CI: 0.91-0.96) for HDL-C, respectively.

The HRs appeared stronger in women (HR 0.80; 95% CI: 0.74-0.86) than in men (HR 0.93; 95% CI: 0.87-0.99) ($p_{heterogeneity}$ =0.002) (**Figure 5.17**). The strength of the association of usual triglycerides with risk of ICH appeared to be stronger

among individuals with low HDL-C (i.e., 1.2 mmol/L or below), compared to those with high HDL-C ($p_{heterogeneity}$ =0.003). Given the modest nature of the difference, and the number of testing conducted, this could reflect the play of chance. There was no evidence of heterogeneity of the association of usual triglycerides with risk of ICH across the different subgroups in the study population (**Figure 5.18-19**).

5.3.3.4 Comparison of associations of lipids-related biomarkers with intracerebral haemorrhage

Overall, LDL-C was inversely associated with risk of ICH (HR 0.90; 95% CI: 0.86-0.95 per 1 SD [0.68 mmol/L] higher), with estimates slightly weaker than that for Apo B100 (HR 0.86; 95% CI: 0.82-0.90 per 1 SD [21 mg/dl] higher). Triglycerides also showed an inverse association with risk of ICH (HR 0.87; 95% CI: 0.82-0.91 per 1 SD [0.58] higher triglycerides [Log_e]), independent of LDL-C and HDL-C. In contrast, neither HDL-C nor Apo A1 was associated with risk of ICH (**Table 5.6**). The associations of the lipid-related biomarkers with risk of ICH were unaltered after additional adjustment for BMI (**Table 5.6**).

Figure 5.21 compares the effects of measured (or usual SBP) had on the associations of baseline lipid-related biomarkers with risk of ICH. As is the case in IS, the associations of lipid-related measurements with risk of ICH were chiefly affected after adjustment for baseline SBP, and as expected, adjustment for usual SBP (rather than baseline SBP) had further effects on the associations, but these confidence intervals were overlapped.

Non-HDL-C, in the present study, was strongly correlated with LDL-C (r=0.89), and Apo B100 (r=0.93), but not with HDL-C (r=-0.01). The shape and strength of the association between non-HDL-C and ICH risk were comparable with those

observed for LDL-C and Apo B100 (**Figures 5.20-21**). Remnant cholesterol was inversely associated with ICH risk (**Figures 5.20-21**), similar to that observed for triglycerides (**Figure 5.21**). The strength of the associations of remnant cholesterol with ICH was altered marginally after further adjustment for LDL-C (HR 0.92; 95% CI: 0.89-0.95), and HDL-C (HR 0.92; 95% CI: 0.89-0.95).

5.4 Discussion

5.4.1 Incidence rates of intracerebral haemorrhage in CKB participants

In this large-scale general population-based prospective study, the annual ICH incidence rate at age 30-89 years in this population was 187 per 100 000 person-years, and were higher, in men than in women (247 versus 147), in rural than in urban areas (241 versus 115), and in Northern than in Southern regions (210 versus 173), of China.

Previous studies have reported that stroke incidence in China, and particularly incidence rates of ICH, was greater in China than that in other countries.[45, 57] The GBD 2015 study reported, at the country level, that China had the highest incidence rates of HS in the world (126.3 cases per 100 000 person-years), with rates that were almost two-fold higher than the global average incidence rates (67.8), and much higher than that in HICs (55.7), with age standardised to the WHO standard population.[4, 233] However, the inter-country comparison in GBD study may somewhat be complicated by the varying data availability, and methods of data collection and reporting used in different countries (or studies) that have been included in GBD. In comparison, the WHO MONICA Project conducted during 1984-2004 has been the first, and until now, the most internationally comparable stroke incidence study.[22, 42] The Sino-MONICA-Beijing project,[57] as

previously described, continued until 2004, in which, the clinical criteria rather than CT/MRI findings were used to diagnose the pathological type of stroke in the first few years, and, as such, reliably incidence rates by stroke pathological types were only available for the second decade of the study (1994-2004). In men, the annual incidence of first-ever ICH declined from 70 to 49 per 100 000, and in women, the annual ICH incidence declined from 63 to 22 per 100 000 over the last decade.[57] The ICH incidence rates in the Sino-MONICA-Beijing Project were higher than those reported in high quality studies conducted in Western populations around the same time.[201, 234, 235] The annual incidence rates for ICH at age 25-74 years in the North East Melbourne Stroke Incidence Study for 1996-1997 were 27 per 100 000 in men, and 18 in women.[201] In the Oxford Vascular Study, the estimated incidence rate of ICH in 2004 was 12 (95% CI: 9-15) per 100 000 person-years, albeit the Oxford Vascular Study has included a slightly older population than the MONICA project.[203]

The precise reasons for the high rates of ICH in the Sino-MONICA-Beijing Project in the mid-1990s, and the decline in ICH thereafter, are not fully understood. Other stroke incidence studies in China have also reported high rates of ICH.[10, 49, 51, 205] In another study in China of three urban communities (in Beijing, Shanghai, and Changsha), where the medical service was relatively accessible, between 1991 and 2000, the incidence rates of ICH were 38.1, 27.1, and 77.1 cases per 100 000 person-years, respectively.[51, 236] Notably, of the three cities, Changsha, a city in Hunan province (a regional province included in the CKB study), has been reported to have a higher incidence rate, and a greater proportion of HS compared to other regions in China in previous studies,[54, 208] and in the present thesis. Another study has reported, around the same time period, the annual incidence

rates of ICH in Changsha ranged from 146 in 1991 to 129 in 2000 cases per 100 000 person-years.[49] The difference in ICH incidence rates observed in Changsha between the two studies may be explained by differences in data collecting methods. Compared to the hospital admission data based method in the first study, the latter used a door-to-door active survey method, which might capture milder cases or those who were not admitted to hospitals.[51]

In a recent national survey conducted in 2013 (i.e., the NESS-China study), which has included a similar proportion of individuals from urban areas in China to the present study, the crude incidence rate of ICH was 82.1 (74.0-90.3) events per 100 000 person-years. As previously described, this large NESS-China survey was based on the national disease surveillance system and included 64 urban and 93 rural areas, across 31 provinces in mainland China. A questionnaire-based primary screening for a history of stroke among 4500 residents (approximately 1500 households) within each survey sites was conducted by local CDC physicians, and the stroke events were ascertained by neurologists thereafter. Incidence rates were calculated as the stroke cases newly-occurred a year prior to the survey. However, in the NESS-China study, the mean age of the study population was much younger than that in the CKB study (46 versus 59 years), whereas in the present CKB study, the age-specific ICH incidence rates (56 cases per 100 000 person-years) were slightly lower than that observed in the NESS-China study.

Arguably, a door-to-door approach would potentially capture more stroke events, particularly those milder cases or cases died out of hospitals, and the death registers cannot capture this in some cases. Such approach might not be feasible

in a large prospective epidemiological study (as in the CKB study). By contrast, in the present CKB study, stroke events were collected through multiple previously-established and overlapping data sources, which can reduce the under-counting of any incident strokes. Furthermore, the incidence rates described here in the CKB study were first-ever-in-a-lifetime stroke incidence,[43] where individuals with prior cardiovascular events (i.e., CHD, stroke or TIA) at baseline were excluded, while it is less clear if the recurrent stroke cases during the NESS-China survey period were also counted as incident stroke in the NESS study. Nevertheless, in the present study, the higher ICH incidence rates with increasing age, and the higher incidence rates in men compared with women were consistent with previous studies in China,[49, 57, 58, 205] and elsewhere.[39, 40, 45, 237]

The proportion of total strokes attributable to ICH in China estimated in the present study (20%), which would possibly be slightly lower if SAH and other unspecified strokes were counted as well, was slightly lower than those observed in the hospital-based national stroke registry (23%),[238] the NESS-China survey (25%),[58] and much lower than those observed in some areas of China or in early days (27-55%).[52, 205] The proportion observed in the present study was largely compatible with estimates in other LMICs (14-25%),[39] and was slightly higher than those observed in HICs (9-13%).[39, 235] As elevated blood pressure has been considered to be a major risk factor for ICH,[67, 239, 240] the higher ICH proportion observed in China, if any, may be explained by the poor control and management of hypertension, compared to other HICs.[58, 241-244] The observed higher proportion of ICH in total strokes in China compared with Western populations might also reflect the smaller proportion of strokes due to other cases, such as IS subtypes due to atherosclerosis.

Incidence rates of ICH varied substantially by region, with a five-fold difference between the 10 study regions. The ICH incidence rates were higher in rural than in urban areas, while the converse was true for the IS incidence rates. The observed diverging patterns in incidence rates of IS and ICH between urban and rural areas in China suggest an epidemiological transition is occurring in China within areas,[48, 163, 245] with an already increasing prevalence of atherosclerotic risk factors in urban areas, and poorer management, and poor awareness of major risk factors for ICH in rural China, chiefly of elevated blood pressure, which shows a particularly strong association with risk of ICH in CKB.[246]

5.4.2 Prospective associations of major lipids and intracerebral haemorrhage

5.4.2.1 Low-density lipoprotein cholesterol

The present study, which included around 5000 incident ICH cases, is the largest general population-based prospective study on this topic, and has demonstrated, throughout the range of usual LDL-C examined (i.e., 1.7-3.2 mmol/L), there was an inverse association of LDL-C with risk of ICH. Due to the relative rarity of HS events, most of the evidence concerning the relevance of LDL-C to HS comes from only a few studies,[13, 108, 111, 125, 247] which, in aggregate, have suggested an inverse association of either TC,[13, 108, 111] or LDL-C,[126] with risk of ICH. The large KNHS study of 3947 HS cases identified through linkage to routinely collected data has found that each 1 mmol/L higher measured plasma TC concentrations was associated with an adjusted HR of 0.91 (95% CI: 0.87-0.95) for risk of HS, after adjustment for conventional vascular risk factors (i.e., age, sex, BMI, height, blood glucose, hypertension, alcohol consumption, smoking, physical activity,

income, and study region).[111] The relative risks of HS associated with 1 mmol/L higher TC in the KNHS study were less extreme than, but still consistent with, those observed in the present study (HR 0.86; 95% CI: 0.80-0.92), perhaps due to the fact that the KNHS study was unable to correct for regression dilution bias, and the potential "surrogate dilution bias" of using TC,[248] and were also unable to adjust for blood pressure adequately. More extreme estimates of the proportional risks were observed in other studies with relatively small numbers of HS cases,[109, 125] but the confidence intervals were still comparable with the present study.

The KNHS study suggested that the inverse association of TC with risk of HS was restricted to hypertensive individuals (HR 0.89; 95% CI: 0.84-0.94),[111] as was also suggested in the MRFIT study, which recorded limited number of ICH deaths (n=83),[108] although, the confidence intervals were largely compatible across different levels of blood pressure (no specific HRs were reported among individuals without hypertension). In the PSC study of approximately 2700 HS cases, there was a weak inverse association of usual TC with risk of mortality from HS, which appeared to be more extreme in individuals with high blood pressure (e.g., above 145 mmHg).[13] The combined results from the MRFIT study and the PSC study were similar to those in the PSC study alone.[13, 108] It has thus been hypothesized that the LDL-C is likely to be more relevant for deep ICH than for lobar ICH,[249] the latter of which is attributable to beta-amyloid accumulation.[249, 250] In the present study, which included more ICH cases than any of the previous studies, there was no evidence of the association of LDL-C with risk of ICH varied by baseline SBP ($p_{heterogeneity}$ =0.25). In the Anglo-Scandinavian Cardiac Outcomes Trial – Lipid Lowering Arm (ASCOT-LLA), among the hypertensive patients, there is no apparent excess risk of total stroke overall, or fatal stroke

alone, associated with lowering LDL-C.[15] Even though the present study has limited statistical power to investigate the potential interaction between blood pressure and LDL-C, if anything, this may well suggest that any interaction is likely to be negligible, compared to the individual role of either blood pressure or cholesterol.

In the KNHS study, it has been claimed that the inverse association of TC with ICH risk was confined to those with high levels of GGT (a proxy of alcohol consumption),[111] but the confidence intervals of the risk estimates were largely overlapping across levels of GGT (Results were reported visually in figures, and group-specific HRs were not available).[111] The inverse association of LDL-C with risk of ICH observed in the present study is less likely to be affected by alcohol consumption, for there was little difference of the multivariable-adjusted risk estimates before and after adjustment for alcohol consumption. In addition, among a subset where the measurements of GGT were available (n=10 266), there was no evidence of heterogeneity in associations of LDL-C with the risk of ICH across different levels of GGT ($p_{heterogeneity}$ =0.59). Consistent with the KNHS study, the present study may still lack statistical power to detect difference in associations across levels of GGT, if any.

The inverse association of LDL-C with ICH observed in the present study is likely to be causal. With more than 500 confirmed haemorrhagic stroke cases in the CTT study, each 1 mmol/L lower LDL-C with statin therapy was associated with a non-significant higher risk of HS (286 versus 242; RR 1.15 [95% CI: 0.96-1.36]).[137] In the SPARCL trial, allocation of participants with stroke to either atorvastatin or placebo, there was a significant higher risk of HS (55 versus 33; RR 1.44 [95% CI:

1.06-1.95]).[141] Combining the CTT and the SPARCL trials yielded a statistically significant higher risk of HS associated with each 1 mmol/L lower LDL-C concentration (RR 1.21; 1.04-1.42; p=0.01). The probable higher risk in HS due to lowering LDL-C by statin treatment is being studied in an ongoing project.[251] The proportional excess risk of ICH associated with each 1 mmol/L lower LDL-C in the present study (HR 1.10; 95% CI: 1.05-1.15) was comparable, albeit was less extreme than, that estimated in the randomised trials.[15]

Taken together, all the available evidence from both randomised trials and observational studies tends to support the hypothesis that lowering LDL-C is causally associated with a higher risk of ICH.[13, 108, 247] The mechanisms for the excess risk of ICH caused by lowering LDL-C are not fully understood. The likely pathogenesis of ICH is likely to involve the rupture of micro-aneurysms resulting from arterionecrosis in small intracerebral arteries.[250, 252, 253] It has been suggested that very low cholesterol concentrations may weaken the endothelium of intracerebral arteries, which is accelerated by the presence of hypertension,[254] and possibly initiates the arterionecrosis process.

The absolute excess risks in ICH associated with lowering LDL-C are chiefly dependent on the underlying rates of ICH, and would be greater in populations with high rates of ICH, for example, cerebrovascular patients, and individuals with hypertension, or other Asian populations where the underlying HS rates are high. In the present study, the overall rates of ICH were 0.18% per annum, which translates to a 9-18 excess ICH cases for each 10 000 participants treated with statin therapy for 5 years, compared to 5-10 extra cases in European and North American populations.[25518] However, statin therapy has been found to reduce the

overall risk of stroke (i.e., IS and HS combined) in many different settings irrespective of the underlying risk of vascular disease,[16, 141, 255] and more importantly, the increase in HS risk is outweighed by the reduction in the risk of IS, as well as in other occlusive vascular events and deaths.[15] Therefore, further reductions in LDL-C concentrations either by medication,[15, 16] or dietary modification,[256-258] are still warranted in Western populations.

5.4.2.2 High-density lipoprotein cholesterol and triglycerides

In this Chinese population with comparatively low LDL-C concentrations, there was no evidence of association of HDL-C, or of Apo A1 with risk of ICH. In contrast, there was a log-linear inverse association of usual triglycerides with ICH risk in this Chinese population, irrespective of LDL-C and HDL-C concentrations.

To date, the largest single prospective study to investigate the association of HDL-C with risk of ICH has been the AMORIS study, which included 865 HS cases. Combining the group-specific estimates (by sex or fatal status) in the AMORIS study using inverse-variance-weighted method, there was no significant overall association of HDL-C with risk of ICH (HR 0.95; 95% CI: 0.89-1.01 per 1 SD [0.39 mmol/L] higher).[116] Several other studies with relatively limited number of ICH cases (~200) did not find any association of HDL-C with risk of ICH.[115, 117, 126] In contrast, in another summary-level meta-analysis of 8 studies (5 were prospective and 3 were nested case-control studies), each 1 mmol/L higher HDL-C was associated with a pooled OR of 1.05 (95% CI: 0.88-1.24), involving an OR of 1.17 (95% CI: 1.02-1.35) for ICH and an OR of 0.46 (95% CI: 0.26-0.91) for SAH.[151] The presence of a significant association of HDL-C with risk of ICH appeared to be chiefly accounted by three studies with limited adjustment for

confounders. Importantly, the individual studies included in this meta-analysis have all been included in an earlier IPD meta-analysis (the PSC study[13]), in which, there was no evidence of any associations of HDL-C with ICH, with even greater number of prospective studies and ICH cases included.

As is the case with HDL-C, the AMORIS study has been the largest single prospective study to investigate the relevance of triglycerides for risk of ICH.[116] In the AMORIS study, there was no evidence of an overall association of triglycerides with risk of HS (HR 1.02; 95% CI: 0.95-1.09 per SD higher triglycerides [estimated independently for this thesis by taking the inverse-variance-weighted average across sex, and severity of events in the published report]).[116] As previously described, the AMORIS study was unable to sufficiently adjust for smoking, diabetes, and blood pressure. In the present CKB study of over 4000 incident ICH cases, without adjustment for diabetes and blood pressure, there was weak evidence of an inverse association of plasma triglyceride concentrations with risk of ICH (HR 0.96; 95% CI: 0.92-1.00 per SD [0.58] higher triglycerides [Log_e]). Neither the ERFC study or APCSC study reported any significant associations of triglycerides with HS risk, but the numbers of cases in each study were limited, with about only 500 and 200 HS cases, respectively.[14, 122] In contrast, in another two studies with even more limited numbers of HS cases, triglycerides have been reported to be inversely associated with risk of ICH.[259, 260] A pooled study of CHS and ARIC, involving 135 incident ICH cases showed that higher triglycerides were associated with a lower risk of ICH (RR 0.56; 95% CI: 0.37-0.84).[260] In the Three-City study, there was a marginal inverse association of triglycerides with risk of HS risk (HR 0.46; 95% CI: 0.21-1.01), with only 34 HS cases in this study.

The CKB included over 4700 incident ICH cases, more than the total number included in all these previous studies combined, and has demonstrated triglycerides were inversely associated with risk of ICH (HR 0.94; 95% CI: 0.92-0.96 per 30% higher triglycerides), independently of other lipid fractions. Hence, the relevance of triglycerides for risk of ICH is unlikely to be explained by its correlations with LDL-C (r=0.15). In CKB, triglycerides were moderately correlated with BMI (r=0.34), and consistently, adiposity, given SBP, has also been observed to be inversely associated with ICH risk in different populations.[261, 262] It is unclear whether triglycerides per se are protective against ICH or represent a biomarker of other underlying pathological mechanisms for ICH.

5.4.3 Strengths and limitations

Consistent with the methodology to estimate incidence rates of IS, estimation of the ICH incidence rates involving use of multiple overlapping surveillance systems – the mortality, disease surveillance system, and the health insurance systems – to identify stroke cases, should have substantially reduced the potential undercounting of ICH cases.[43] Collection of data on disease outcomes through a number of pre-existing data sources reduced the risk of selection and recall bias, as well as limiting the effects of biases inherent in any individual data source.[263, 264] Although, compared to studies using death certificates alone, the complementary use of multiple-sources in event identification is one of the major strengths in the present study, access to health facilities is still limited in some rural areas in China, and thus, some fatal ICH cases might not be captured in the present study. Another main strength of the CKB study is the high rates of CT/MRI

scanning for stroke diagnosis, which allows more accurate classification of stroke pathological types.

The present study included the largest number of incident ICH cases from a single population for analysis of lipid-associated risks for ICH, which reduced potential biases from the correlation between lipid distribution and ICH incidence rates in different populations, or other sources of heterogeneity between populations. In the present study, directly measured LDL-C concentrations were used rather than estimated values calculated using the Friedewald equation, which reduced measurement error in other lipid fractions and the underlying assumptions of the formula. Furthermore, the diversity of the data collected in the CKB enabled a comprehensive adjustment for potential confounding factors, and thus, the residual confounding is likely to have been greatly reduced. Moreover, the correction for regression dilution bias has enabled more accurate estimates of the long-term effects on risk of ICH associated with differences in major blood lipids.

However, although larger numbers of ICH cases were recorded in the CKB study, compared with most previous studies, the present study still has limited statistical power in some subgroup analyses, particularly blood pressure. Whilst the adjustment of major blood lipids for regression dilution bias is one of the strengths in the present analyses, the RDR estimates in the present CKB study may still be subject to residual confounding, due to the relatively limited number of participants with re-measurements. RDR estimates in the CKB study were, however, largely compatible with those seen in previous studies of similar follow-up time.[13, 14]

5.5 Summary

In summary, around 20% of incident strokes were attributable to ICH in this population, which was less extreme than estimates in previous studies in Chinese populations and was slightly higher than those in Western populations. The overall incidence rates of ICH in the present study were around 187 cases per 100 000 person-years, and were higher than estimates in Western populations. In the present study with large number of ICH cases, LDL-C was inversely associated with ICH incident, with no evidence that the strengths of the association varied by levels of SBP. There was no evidence of association of HDL-C with ICH risk in this population with relatively low LDL-C. Irrespective of measurements of LDL-C, and HDL-C, plasma triglyceride concentrations were inversely associated with ICH risk.

Even assuming a causal association of lower LDL-C with higher risk of ICH, the public health implications are probably limited for most Western populations, but the absolute excess risks of ICH have important implications for Chinese populations, and some other Asian populations. In the present study, the strength of associations of major blood lipids with risk of ICH were chiefly affected after adjustment for blood pressure, and thereby provides further evidence for the need for a better control and management of blood pressure for the prevention of ICH at population levels.

In the present study, the precise reasons for the inverse association of plasma triglyceride concentrations with risk of ICH are unclear, and thus, it is unclear whether triglycerides are protective against ICH, or instead, a biomarker of other underlying mechanisms that cause ICH. Hence, well-powered genetic studies are needed to assess the causal relevance of these associations.

5.6 Tables

Table 5.1: Number and crude incidence rate of intracerebral haemorrhage, by sex and area (among 489 762 participants)*

	Men			Women			Total		
	Person-years at risk ($x10^3$)	No. of intracerebral haemorrhage	Crude rate per 100 000 person-years	Person-years at risk ($x10^3$)	No. of intracerebral haemorrhage	Crude rate per 100 000 person-years	Person-years at risk ($x10^3$)	No. of intracerebral haemorrhage	Crude rate per 100 000 person-years
Age at risk (years)									
30-39	81.32	29	36	133.81	26	19	215.13	55	26
40-44	218.35	120	55	356.44	118	33	574.79	238	41
45-49	267.64	228	85	433.82	262	60	701.46	490	70
50-54	268.20	340	127	418.25	349	83	686.45	689	100
55-59	286.25	516	180	434.53	571	131	720.78	1087	151
60-64	238.98	666	279	344.79	604	175	583.78	1270	218
65-69	177.32	742	418	233.08	587	252	410.41	1329	324
70-74	139.58	842	603	167.78	700	417	307.36	1542	502
75-79	81.13	701	864	93.24	467	501	174.37	1168	670
80-89	18.29	208	1137	21.96	194	883	40.25	402	999
All (30-89)	*1777.06*	*4392*	*247*	*2637.7*	*3878*	*147*	*4414.78*	*8270*	*187*
Study area									
Qingdao (U)	138.49	123	89	167.82	93	55	306.32	216	71
Harbin (U)	174.95	395	226	269.41	331	123	444.35	726	163
Haikou (U)	90.77	134	148	162.74	120	74	253.51	254	100
Suzhou (U)	193.16	222	115	278.43	190	68	471.59	412	87
Liuzhou (U)	155.39	323	208	256.24	240	94	411.63	563	137
Sichuan (R)	186.41	499	268	309.37	479	155	495.78	978	197
Gansu (R)	166.05	673	405	272.61	824	302	438.66	1497	341
Henan (R)	235.19	637	271	311.43	570	183	546.62	1207	221
Zhejiang (R)	214.22	373	174	315.50	293	93	529.71	666	126
Hunan (R)	222.44	1013	455	294.16	738	251	516.61	1751	339
All	*1777.07*	*4392*	*247*	*2637.71*	*3878*	*147*	*4414.78*	*8270*	*187*

* Excluding participants who reported doctor-diagnosed coronary heart disease or stroke or transient ischaemic attack at baseline; U=Urban; R=Rural.

Table 5.2: Selected baseline characteristics of the nested case-control study participants*

Characteristics	Intracerebral haemorrhage cases (n = 4776)	Controls (n = 6290)
Age (years), %		
30-39	3.8	8.0
40-49	15.5	18.7
50-59	29.0	28.9
60-69	33.2	28.5
70-79	18.5	15.9
Mean (SD)	59.1 (10.0)	56.9 (10.8)
Sex (Women), %	48.0	48.0
Area (Urban), %	22.6	21.3
Highest education completed, %		
No formal school	30.9	26.9
Primary school	39.5	40.7
Secondary school	19.2	20.1
High school or above	10.4	12.3
Smoking category, %		
Never-smoker	51.3	52.0
Occasional smoker	6.4	6.5
Ever regular	8.7	7.8
Current smoker	33.6	33.7
Alcohol consumption, %		
Never-drinker	48.8	46.3
Occasional drinker	6.7	5.6
Ever regular	28.3	32.5
Current drinker	16.2	15.6
Regular consumption of certain foods, %		
Fresh vegetables	96.8	97.3
Meat or poultry	33.5	35.7
Egg	21.2	26.6
Fresh fruit	15.1	18.6
Soybean	8.1	9.1
Dairy products	5.9	9.0
Fish or seafood	3.7	4.7
Self-rated health, %		
Poor	15.9	9.6
Fair	45.8	45.5
Good	25.8	28.9
Excellent	12.5	16.0
Diabetes, %	8.1	4.8
Hypertension, %	25.9	10.6

*Directly standardised to the baseline age (10-year), sex, and study region structure of the CKB study population, where appropriate. **Included doctor-diagnosed and screen-detected diabetes at baseline, see above Section (4.2.2.3) for definition of screen-detected diabetes.

Table 5.3: Selected baseline physical measurements of the nested case-control study participants

Characteristics	Intracerebral haemorrhage cases, % (n = 4776)	Controls, % (n = 6290)
Body mass index (kg/m^2)		
<20	18.0	18.7
20-<22	21.5	21.9
22-<24	21.5	23.9
24-<26	17.0	17.9
26+	22.1	17.7
Mean (SD)	23.3 (3.5)	23.0 (3.1)
Waist circumference (cm)		
<72	21.5	23.6
72-<78	21.4	23.8
78-<83	18.6	19.2
83-<89	17.9	17.0
89+	20.6	16.3
Mean (SD)	80.5 (9.9)	79.2 (9.1)
Hip circumference (cm)		
<84	21.8	22.4
84-<88	22.0	23.5
88-<92	21.0	21.9
92-<96	17.3	17.1
96+	17.9	15.2
Mean (SD)	89.5 (6.5)	89.1 (6.0)
Body fat percentage (%)		
<18	21.9	22.1
18-<24	23.6	24.7
24-<30	24.2	24.0
30-<35	14.1	14.9
35+	16.2	14.3
Mean (SD)	25.7 (7.0)	25.3 (6.4)
Systolic blood pressure (mmHg)		
<120	10.1	24.6
120-<135	18.2	30.8
135-<145	14.1	16.0
145-<160	21.0	16.1
160+	36.6	12.4
Mean (SD)	153 (28)	135 (20)
Random blood glucose (mmol/L)*		
<4.8	16.6	20.1
4.8-<5.8	34.6	37.3
5.8-<7.8	35.5	33.1
7.8-<11.1	8.2	6.7
11.1+	5.1	2.8
Mean (SD)	6.4 (3.0)	6.0 (2.0)

Table continues on next page.

Table 5.3: Selected baseline physical measurements of the nested case-control study participants

Characteristics	Intracerebral haemorrhage cases, % (n = 4776)	Controls, % (n = 6290)
Estimated glomerular filtration rate (ml/min/1.73m^2)* **		
<90	28.0	22.7
90-<100	27.5	29.6
100-<110	25.0	28.2
110-<120	15.0	15.3
120+	4.6	4.3
Mean (SD)	96.4 (12.6)	98.3 (9.9)
Total physical activity (MET hours /day), %		
<10	34.3	28.7
10-<20	30.4	32.8
20-<30	16.1	18.1
30-<40	11.1	12.2
40+	8.1	8.1
Mean (SD)	18.1 (12.1)	19.1 (11.1)

MET=Metabolic equivalent of task; *Directly standardised to the baseline age (10-year), sex, and study region structure of the CKB study population, where appropriate; **Among individuals with measurements of random blood glucose at baseline (5437 IS cases versus 6218 controls); ***Among individuals with creatinine measurements lower than 175 umol/L (5469 IS cases versus 6277 controls);

Table 5.4: Distribution of lipoprotein-related biomarkers of baseline samples among the nested case-control study participants*

Characteristics	Ischaemic stroke cases, % (n = 4776)	Controls, % (n = 6290)
Total cholesterol (mmol/L)		
<4.0	29.9	27.4
4.0-<4.5	21.6	23.4
4.5-<5.0	20.7	19.9
5.0-<5.5	13.1	14.5
5.5+	14.7	14.8
Mean (SD)	4.5 (0.9)	4.6 (0.8)
LDL cholesterol (mmol/L)		
<1.8	24.8	23.4
1.8-<2.3	30.5	30.6
2.3-<2.9	27.5	28.9
2.9-<3.4	11.0	11.7
3.4+	6.2	5.5
Mean (SD)	2.3 (0.7)	2.3 (0.7)
HDL cholesterol (mmol/L)		
<0.8	3.1	2.5
0.8-<1.0	15.6	14.8
1.0-<1.2	29.2	29.0
1.2-<1.4	24.7	25.8
1.4+	27.4	28.0
Mean (SD)	1.3 (0.3)	1.3 (0.3)
Triglycerides (mmol/L)		
<1.0	21.6	21.7
1.0-<1.5	27.7	27.5
1.5-<2.0	18.5	18.1
2.0-<3.0	19.0	18.9
3.0+	13.2	13.7
Median (IQR)	1.5 (1.2)	1.5 (1.2)
Apolipoprotein A1 (mg/dl)		
<110	11.4	9.8
110-<130	33.0	33.0
130-<140	18.0	18.7
140-<160	23.9	25.6
160+	13.7	12.9
Mean (SD)	135.3 (22.4)	135.7 (21.4)
Apolipoprotein B100 (mg/dl)		
<60	15.7	13.7
60-<80	35.4	34.4
80-<100	31.9	34.4
100-<120	13.5	14.5
120+	3.4	3.0
Mean (SD)	80.5 (20.0)	81.7 (18.7)

IQR=Interquartile range; *Directly standardised to the baseline age (10-year), sex, and study region structure of the CKB study population, where appropriate.

Table 5.5.1: Mean values of the lipoprotein-related biomarkers among the nested case-control study participants (n=11 066)

	TC, mmol/L	LDL-C, mmol/L	HDL-C, mmol/L	Triglyceride, log_e	Apo A1, mg/dl	Apo B100, mg/dl
Sex						
Men	4.39 (0.86)	2.21 (0.63)	1.23 (0.30)	0.40 (0.58)	133.01 (23.34)	78.71 (19.21)
Women	4.74 (0.91)	2.37 (0.66)	1.28 (0.28)	0.52 (0.55)	138.37 (20.70)	84.14 (19.50)
Study area						
Qingdao (U)	5.31 (1.09)	2.81 (0.79)	1.27 (0.33)	0.53 (0.64)	135.24 (23.44)	98.28 (22.81)
Harbin (U)	4.84 (0.92)	2.51 (0.70)	1.20 (0.30)	0.53 (0.68)	129.41 (22.17)	89.15 (20.71)
Haikou (U)	5.38 (1.06)	3.04 (0.81)	1.25 (0.28)	0.45 (0.58)	134.03 (20.65)	100.16 (22.25)
Suzhou (U)	4.56 (0.84)	2.31 (0.65)	1.29 (0.30)	0.33 (0.60)	137.72 (23.14)	81.15 (19.99)
Liuzhou (U)	4.88 (0.95)	2.62 (0.73)	1.22 (0.27)	0.48 (0.61)	131.86 (20.50)	90.04 (20.40)
Urban areas	*4.88 (0.96)*	*2.57 (0.73)*	*1.24 (0.29)*	*0.46 (0.62)*	*133.09 (21.53)*	*89.40 (21.19)*
Sichuan (R)	4.59 (1.00)	2.25 (0.72)	1.40 (0.39)	0.37 (0.61)	148.18 (28.45)	77.66 (21.15)
Gansu (R)	4.19 (0.82)	2.04 (0.58)	1.21 (0.26)	0.47 (0.52)	130.96 (21.37)	74.42 (18.04)
Henan (R)	4.45 (0.86)	2.30 (0.62)	1.16 (0.26)	0.41 (0.57)	127.27 (20.52)	81.38 (18.79)
Zhejiang (R)	4.54 (0.90)	2.25 (0.65)	1.32 (0.32)	0.35 (0.52)	142.33 (23.27)	80.74 (20.60)
Hunan (R)	4.66 (0.90)	2.31 (0.67)	1.28 (0.30)	0.53 (0.57)	138.21 (21.71)	82.23 (19.69)
Rural areas	*4.46 (0.88)*	*2.21 (0.63)*	*1.26 (0.3)*	*0.45 (0.55)*	*136.25 (23.31)*	*78.86 (19.12)*
Overall	*4.55 (0.86)*	*2.28 (0.63)*	*1.26 (0.29)*	*0.45 (0.55)*	*135.57 (21.82)*	*81.17 (19.06)*

SD=Standard deviation; U=Urban areas; R=Rural areas; TC=Total cholesterol; LDL-C=Low-density lipoprotein cholesterol; HDL-C=High-density lipoprotein cholesterol; Apo=Apolipoprotein; Directly standardised to the baseline age (10-year), sex, region, and disease status structure of the study population included, where appropriate. Values are mean (SD) unless otherwise stated.

13, 248

Table 5.5.2: Selected baseline characteristics of the nested case-control study participants by levels of total cholesterol (n=11 066)

Characteristics	Total cholesterol (mmol/L)					P for trend
	≤3.76	3.77 - 4.25	4.26 - 4.70	4.71 - 5.28	≥5.29	
Number of participants	2194	2238	2218	2216	2200	
Demographic factors						
Age, Mean (SD), years	56.3 (11.8)	57.3 (11.4)	57.7 (11.2)	57.8 (10.7)	59.1 (11.8)	<.0001
Women, %	34.6	41.4	47.9	54.8	60.3	<.0001
Urban, %	12.2	17.3	19.8	26.1	33.8	<.0001
High school or above, %	9.4	10.2	11.6	12.6	11.5	0.0005
Household income >20 000 yuan/year, %	26.2	26.5	27.0	27.1	27.7	0.18
Lifestyle factors						
Male ever smoker, %	87.2	86.4	86.1	86.7	84.2	0.15
Female ever smoker, %	7.5	8.0	7.2	7.6	7.5	0.84
Male ever drinker, %	70.9	72.4	72.8	71.5	72.9	0.43
Female ever drinker, %	29.0	29.7	33.1	33.1	29.3	0.55
Physical activity, Mean (SD), MET hours/day	19.4 (11.6)	19.4 (11.8)	18.9 (11.6)	18.5 (12.5)	17.4 (14.1)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	30.4	31.9	35.4	35.1	39.9	<.0001
Fish or other seafood	2.5	3.1	4.6	4.2	4.9	<.0001
Egg	22.7	24.8	23.2	26.0	24.8	0.061
Fresh fruit	16.2	17.1	16.1	17.8	16.9	0.35
Dairy products	6.1	6.3	6.2	9.1	8.8	<.0001
Health conditions, %						
Diabetes	5.3	4.5	5.3	6.8	9.3	<.0001
Hypertension	15.0	15.8	16.9	18.8	19.0	<.0001
Poor health	14.5	11.5	11.2	11.9	13.0	0.21

Values are mean (SD) unless otherwise stated. All values are directly standardised to the age (10-year group), sex, and study area distribution of the entire CKB population, where appropriate.

Table 5.5.3: Selected baseline characteristics of the nested case-control study participants by levels of LDL cholesterol (n=11 066)

Characteristics	LDL cholesterol (mmol/L)					P for trend
	≤1.71	1.72-2.06	2.07-2.39	2.40-2.81	≥2.82	
Number of participants	2196	2242	2200	2206	2222	
Demographic factors						
Age, Mean (SD), years	56.1 (11.4)	57.5 (10.9)	57.6 (10.8)	58.2 (10.3)	59.1 (11.9)	<.0001
Women, %	39.1	42.8	48.5	51.4	56.5	<.0001
Urban, %	12.0	16.7	18.9	24.4	36.7	<.0001
High school or above, %	9.8	10.3	11.4	12.2	12.9	<.0001
Household income >20 000 yuan/year, %	26.7	26.3	26.6	27.2	27.6	0.24
Lifestyle factors						
Male ever smoker, %	86.9	87.1	86.1	86.9	83.6	0.087
Female ever smoker, %	8.4	6.9	7.4	7.3	7.1	0.32
Male ever drinker, %	72.8	72.9	71.8	72.9	69.6	0.22
Female ever drinker, %	31.7	29.5	32.1	33.7	29.0	0.68
Physical activity, Mean (SD), MET hours/day	19.6 (11.9)	19.3 (11.9)	18.9 (11.6)	18.1 (12.0)	17.4 (13.6)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	31.8	33.0	34.4	35.4	39.2	<.0001
Fish or other seafood	2.9	4.5	3.5	4.8	4.9	0.0052
Egg	25.5	24.1	23.3	24.8	23.7	0.34
Fresh fruit	15.5	16.5	15.1	17.9	18.3	0.0031
Dairy products	7.0	6.2	6.9	8.4	8.8	0.0005
Health conditions, %						
Diabetes	5.7	5.6	5.1	6.2	8.6	<.0001
Hypertension	15.4	15.8	17.6	17.5	20.2	<.0001
Poor health	13.3	13.9	11.9	11.6	12.0	0.026

Conventions as in Table 5.5.2

Table 5.5.4: Selected baseline characteristics of the nested case-control study participants by levels of HDL cholesterol (n=11 066)

Characteristics	HDL cholesterol (mmol/L)					P for trend
	≤1.01	1.02-1.14	1.15-1.29	1.30-1.47	≥1.48	
Number of participants	969	2757	3094	2951	1295	
Demographic factors						
Age, Mean (SD), years	57.6 (12.4)	57.6 (10.8)	57.8 (10.5)	58.1 (10.8)	58.5 (11.6)	0.0103
Women, %	34.4	44.2	49.2	53.7	53.8	<.0001
Urban, %	23.7	22.2	21.7	21.5	19.6	0.026
High school or above, %	13.7	12.6	11.6	9.4	11.0	<.0001
Household income >20 000 yuan/year, %	28.5	26.7	27.2	27.0	27.0	0.73
Lifestyle factors						
Male ever smoker, %	88.3	86.5	84.2	85.9	86.3	0.19
Female ever smoker, %	9.3	7.0	8.3	6.5	8.3	0.55
Male ever drinker, %	64.5	68.2	71.4	75.7	79.6	<.0001
Female ever drinker, %	27.3	29.9	30.3	32.1	36.0	0.0001
Physical activity, Mean (SD), MET hours/day	17.6 (11.4)	18 (11.3)	18.8 (12)	19.4 (12.4)	19.5 (13.6)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	34.6	34.6	33.1	35.5	37.9	0.021
Fish or other seafood	4.6	4.6	4.3	4.1	4.5	0.50
Egg	24.1	23.0	24.4	24.2	27.9	0.015
Fresh fruit	16.1	18.7	18.3	15.7	15.0	0.0063
Dairy products	7.6	7.1	7.3	8.2	9.4	0.012
Health conditions, %						
Diabetes	11.0	8.2	5.5	4.9	4.3	<.0001
Hypertension	23.2	19.6	17.8	13.8	13.0	<.0001
Poor health	13.4	13.7	12.1	11.7	11.8	0.020

Conventions as in Table 5.5.2

Table 5.5.5: Selected baseline characteristics of the nested case-control study participants by levels of triglycerides (n=11 066)

Characteristics	Triglycerides (mmol/L)					P for trend
	≤0.96	0.97-1.30	1.31-1.75	1.76-2.51	≥2.52	
Number of participants	2207	2192	2246	2215	2206	
Demographic factors						
Age, Mean (SD), years	57.6 (11.5)	58.3 (10.8)	57.8 (10.5)	57.7 (10.6)	57 (10.5)	0.01
Women, %	37.7	46.0	50.7	52.7	53.4	<.0001
Urban, %	23.8	19.4	20.1	22.5	25.0	0.048
High school or above, %	11.0	10.0	11.2	12.0	13.3	0.0005
Household income >20 000 yuan/year, %	27.1	25.9	26.8	26.6	27.6	0.48
Lifestyle factors						
Male ever smoker, %	85.9	86.2	84.9	85.5	86.8	0.73
Female ever smoker, %	7.2	6.8	6.9	7.4	8.2	0.21
Male ever drinker, %	70.7	73.1	70.0	72.7	74.6	0.064
Female ever drinker, %	32.8	32.7	32.9	29.2	31.0	0.082
Physical activity, Mean (SD), MET hours/day	19.4 (11.7)	19.4 (11.8)	18.6 (11.5)	18.6 (11.9)	17.6 (12.1)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	33.9	34.9	34.1	35.2	36.3	0.060
Fish or other seafood	3.7	4.4	3.5	4.5	5.1	0.027
Egg	24.7	24.2	23.3	25.6	24.5	0.67
Fresh fruit	15.5	16.5	16.2	17.3	18.9	0.0006
Dairy products	7.6	6.2	7.5	7.6	9.6	0.002
Health conditions, %						
Diabetes	3.5	4.6	4.9	7.2	11.2	<.0001
Hypertension	11.8	15.8	16.3	19.4	21.7	<.0001
Poor health	13.4	12.6	12.2	11.0	12.7	0.19

Conventions as in Table 5.5.2

Table 5.5.6: Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein A1 (n=11 066)

Characteristics	Apolipoprotein A1 (mg/dl)					P for trend
	≤117.2	117.3-128.2	128.3-138.9	139.0-152.7	≥152.8	
Number of participants	2207	2222	2222	2198	2217	
Demographic factors						
Age, Mean (SD), years	56.3 (12.5)	57 (11.2)	57.7 (10.8)	58.8 (10.3)	59.6 (10.7)	<.0001
Women, %	33.8	44.0	52.4	55.9	54.3	<.0001
Urban, %	25.0	22.7	21.2	22.2	18.2	<.0001
High school or above, %	11.8	12.0	11.3	10.5	10.2	0.019
Household income >20 000 yuan/year, %	26.8	26.0	27.8	28.2	25.4	0.97
Lifestyle factors						
Male ever smoker, %	87.5	84.7	84.7	84.2	89.2	0.38
Female ever smoker, %	7.9	8.3	6.5	7.3	7.5	0.62
Male ever drinker, %	64.9	69.3	71.8	74.4	81.2	<.0001
Female ever drinker, %	27.3	28.5	30.5	31.9	35.4	<.0001
Physical activity, Mean (SD), MET hours/day	18.1 (11.9)	18.2 (11.4)	19 (12.2)	18.8 (13.5)	19.1 (13.7)	0.0024
Regular consumption of certain foods, %						
Meat or poultry	31.9	33.4	33.9	35.7	37.8	<.0001
Fish or other seafood	4.2	3.7	4.3	4.6	4.8	0.071
Egg	22.1	23.0	25.0	24.0	28.3	<.0001
Fresh fruit	17.5	18.0	17.9	16.1	14.4	0.0014
Dairy products	7.1	7.1	7.6	8.5	8.0	0.062
Health conditions, %						
Diabetes	6.6	7.1	6.1	6.0	5.8	0.071
Hypertension	19.8	18.8	17.0	14.2	14.6	<.0001
Poor health	14.1	13.7	11.1	12.4	11.9	0.013

Conventions as in Table 5.5.2

Table 5.5.7: Selected baseline characteristics of the nested case-control study participants by levels of Apolipoprotein B100 (n=11 066)

Characteristics	Apolipoprotein B100 (mg/dl)					P for trend
	≤63.6	63.7-75.0	75.1-85.6	85.7-97.6	≥97.7	
Number of participants	2208	2209	2226	2212	2211	
Demographic factors						
Age, Mean (SD), years	56.7 (11.1)	57 (10.9)	57.6 (10.6)	58.2 (10.5)	58.6 (11.6)	<.0001
Women, %	39.2	42.7	47.9	52.7	57.6	<.0001
Urban, %	12.6	15.5	19.9	24.3	36.7	<.0001
High school or above, %	9.3	11.0	10.9	12.5	13.3	<.0001
Household income >20 000 yuan/year, %	25.5	26.3	26.1	27.5	29.1	0.0007
Lifestyle factors						
Male ever smoker, %	85.1	88.9	87.1	83.5	85.4	0.099
Female ever smoker, %	7.7	7.2	7.9	7.2	7.6	0.87
Male ever drinker, %	73.3	73.4	70.0	72.3	71.8	0.26
Female ever drinker, %	32.2	30.0	32.5	31.9	29.1	0.24
Physical activity, Mean (SD), MET hours/day	19.6 (11.6)	19.9 (11.5)	18.5 (11.7)	18.2 (11.7)	16.9 (14.2)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	31.4	33.0	34.6	35.0	38.9	<.0001
Fish or other seafood	2.3	3.5	4.3	4.5	5.4	<.0001
Egg	24.4	22.7	24.2	25.0	24.1	0.55
Fresh fruit	15.4	16.9	16.1	16.4	18.7	0.011
Dairy products	6.5	5.8	7.6	8.3	8.2	0.0006
Health conditions, %						
Diabetes	4.1	4.3	5.6	6.8	10.5	<.0001
Hypertension	14.8	13.8	18.0	17.8	21.3	<.0001
Poor health	13.4	11.8	12.3	11.7	12.7	0.51

Conventions as in Table 5.5.2

Table 5.5.8: Selected baseline characteristics of the nested case-control study participants by levels of non-HDL cholesterol (n=11 066)

Characteristics	non-HDL cholesterol (mmol/L)					P for trend
	≤2.54	2.55-3.00	3.01-3.42	3.43-3.98	≥3.99	
Number of participants	2209	2214	2223	2206	2214	
Demographic factors						
Age, Mean (SD), years	56.4 (11.7)	57.2 (10.8)	57.8 (10.6)	57.9 (10.6)	58.7 (11.8)	<.0001
Women, %	35.8	42.0	48.8	53.2	59.3	<.0001
Urban, %	11.8	15.8	19.9	25.4	36.3	<.0001
High school or above, %	10.0	10.8	11.4	12.4	12.9	0.0001
Household income >20 000 yuan/year, %	26.1	26.5	26.6	27.6	28.2	0.037
Lifestyle factors						
Male ever smoker, %	86.7	87.4	87.7	84.7	83.9	0.028
Female ever smoker, %	7.6	6.4	7.5	7.5	7.6	0.64
Male ever drinker, %	73.9	72.2	71.7	72.0	70.3	0.075
Female ever drinker, %	31.4	30.4	32.4	32.8	28.8	0.42
Physical activity, Mean (SD), MET hours/day	20 (12.1)	19.6 (11.9)	18.9 (11.6)	18.1 (12.2)	17.1 (14.2)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	30.2	33.2	34.2	35.1	40.8	<.0001
Fish or other seafood	2.3	4.2	3.7	4.6	5.4	<.0001
Egg	24.2	22.9	25.5	24.8	24.1	0.49
Fresh fruit	16.0	16.8	16.5	16.7	19.1	0.020
Dairy products	6.7	6.0	7.0	8.7	9.0	<.0001
Health conditions, %						
Diabetes	4.3	3.7	5.5	7.1	10.1	<.0001
Hypertension	13.2	14.7	16.5	19.3	21.0	<.0001
Poor health	14.0	12.6	11.0	11.4	13.7	0.38

Conventions as in Table 5.5.2

Table 5.5.9: Selected baseline characteristics of the nested case-control study participants by levels of remnant cholesterol (n=11 066)

Characteristics	Remnant cholesterol (mmol/L)					P for trend
	≤0.70	0.71-0.85	0.86-1.02	1.03-26	≥1.27	
Number of participants	2215	2225	2206	2205	2215	
Demographic factors						
Age, Mean (SD), years	57.8 (11.2)	57.4 (11.1)	57.6 (10.7)	57.9 (10.8)	57.4 (11.5)	0.59
Women, %	35.0	42.0	48.5	56.5	59.6	<.0001
Urban, %	16.6	17.7	22.6	24.8	28.1	<.0001
High school or above, %	10.1	10.7	11.0	12.2	12.9	0.0002
Household income >20 000 yuan/year, %	27.5	27.3	25.6	26.9	28.7	0.27
Lifestyle factors						
Male ever smoker, %	86.0	86.3	87.4	84.6	84.4	0.28
Female ever smoker, %	6.2	6.4	6.0	8.7	7.9	0.0058
Male ever drinker, %	73.0	71.5	72.1	70.4	72.4	0.46
Female ever drinker, %	31.9	32.7	30.7	31.1	30.1	0.19
Physical activity, Mean (SD), MET hours/day	19.8 (12.9)	19.1 (11.9)	19.2 (11.9)	18.4 (12.1)	17 (13.5)	<.0001
Regular consumption of certain foods, %						
Meat or poultry	31.4	35.1	33.9	35.4	37.7	<.0001
Fish or other seafood	2.9	3.9	4.1	4.4	5.2	<.0001
Egg	22.9	24.4	24.7	25.2	25.7	0.035
Fresh fruit	17.2	16.2	17.8	16.3	18.4	0.34
Dairy products	6.7	5.9	8.2	8.4	8.8	<.0001
Health conditions, %						
Diabetes	4.0	4.3	5.4	5.9	11.5	<.0001
Hypertension	13.4	14.9	15.3	19.5	21.8	<.0001
Poor health	12.5	12.6	11.9	12.2	13.7	0.44

Conventions as in Table 5.5.2

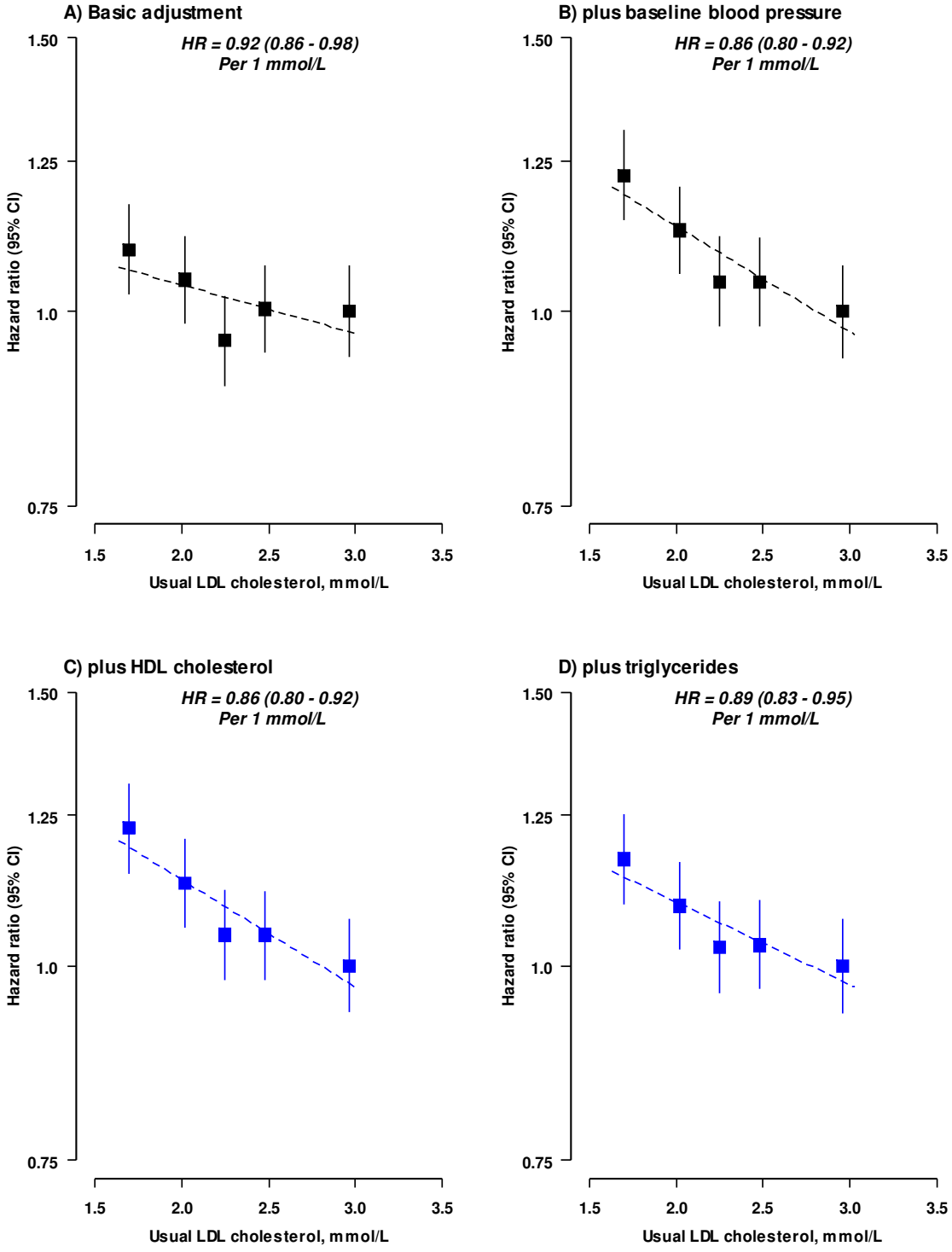
Table 5.6: Comparison of hazard ratios (HR) of major lipid-related biomarkers with intracerebral haemorrhage, before and after adjustment for body mass index (BMI)*

Lipoprotein biomarkers (SD)	HR (95% CI)	
	Without BMI	With BMI
LDL cholesterol (0.68 mmol/L)	0.90 (0.86-0.95)	0.91 (0.87-0.96)
Triglycerides (Log_e) (0.58)	0.87 (0.82-0.91)	0.88 (0.84-0.93)
HDL cholesterol (0.29 mmol/L)	1.00 (0.96-1.05)	0.98 (0.94-1.02)
Apolipoprotein B100 (21 mg/dl)	0.86 (0.82-0.90)	0.87 (0.84-0.91)
Apolipoprotein A1 (22 mg/dl)	0.93 (0.89-0.98)	0.92 (0.87-0.96)

SD=Standard deviation; BMI=Body mass index; *Analyses were stratified by age-at-risk (5-year), sex, and study region; adjusted for education, smoking, alcohol consumption, physical activity, diabetes, and baseline systolic blood pressure (or with BMI). HRs were presented for differences in SD unit of the measured value of each lipoprotein related biomarker.

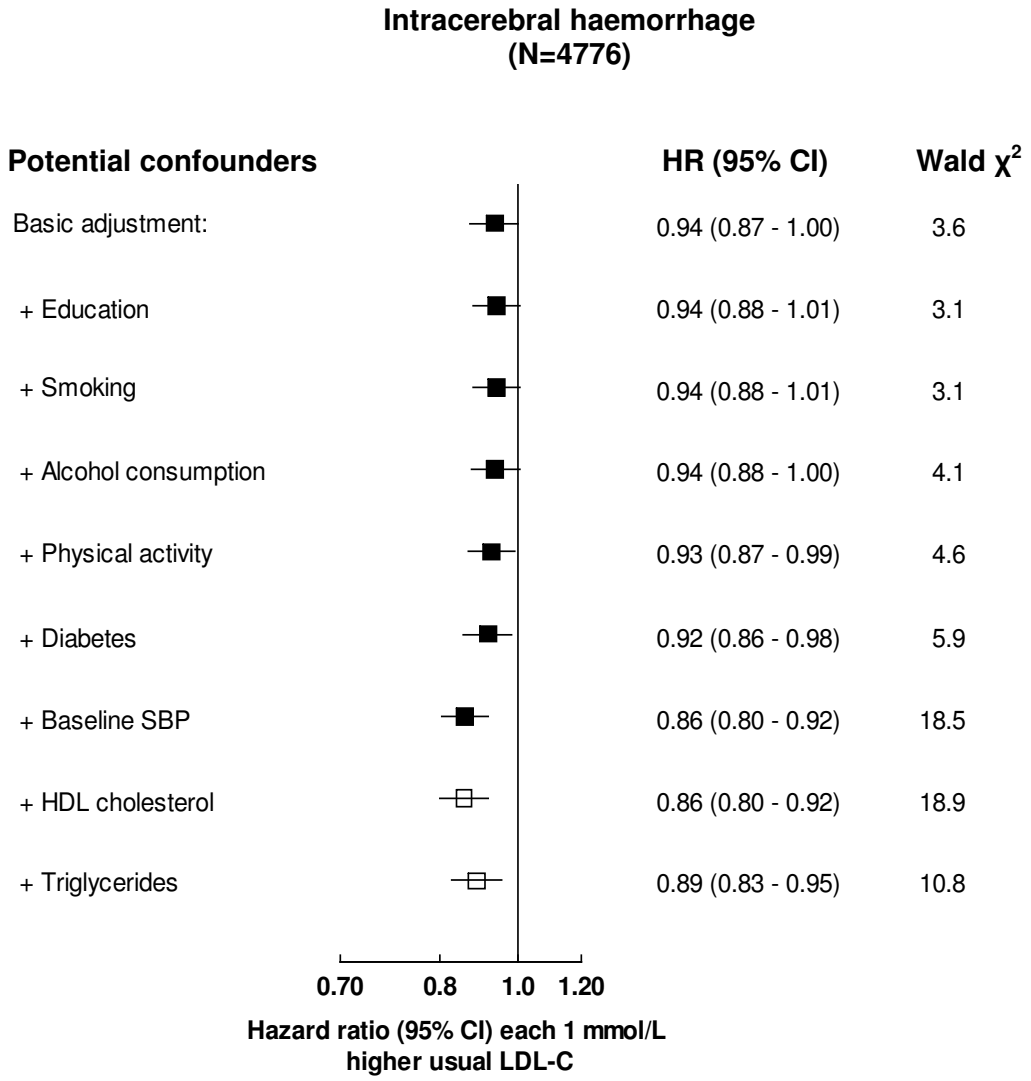
5.7 Figures

Figure 5.1: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol



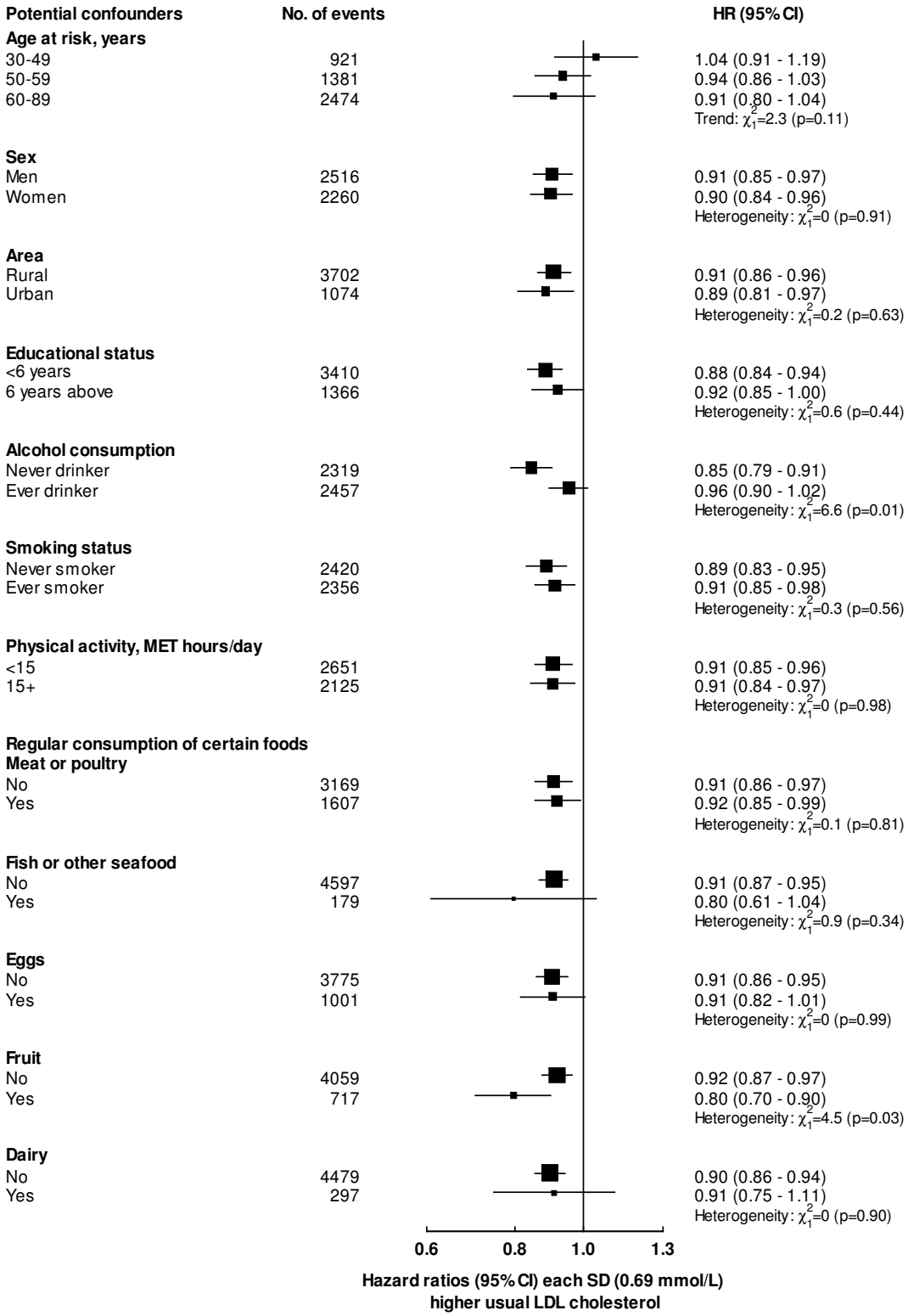
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for HDL cholesterol in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual LDL cholesterol concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 5.2: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol, progressively adjusted for potential confounding factors



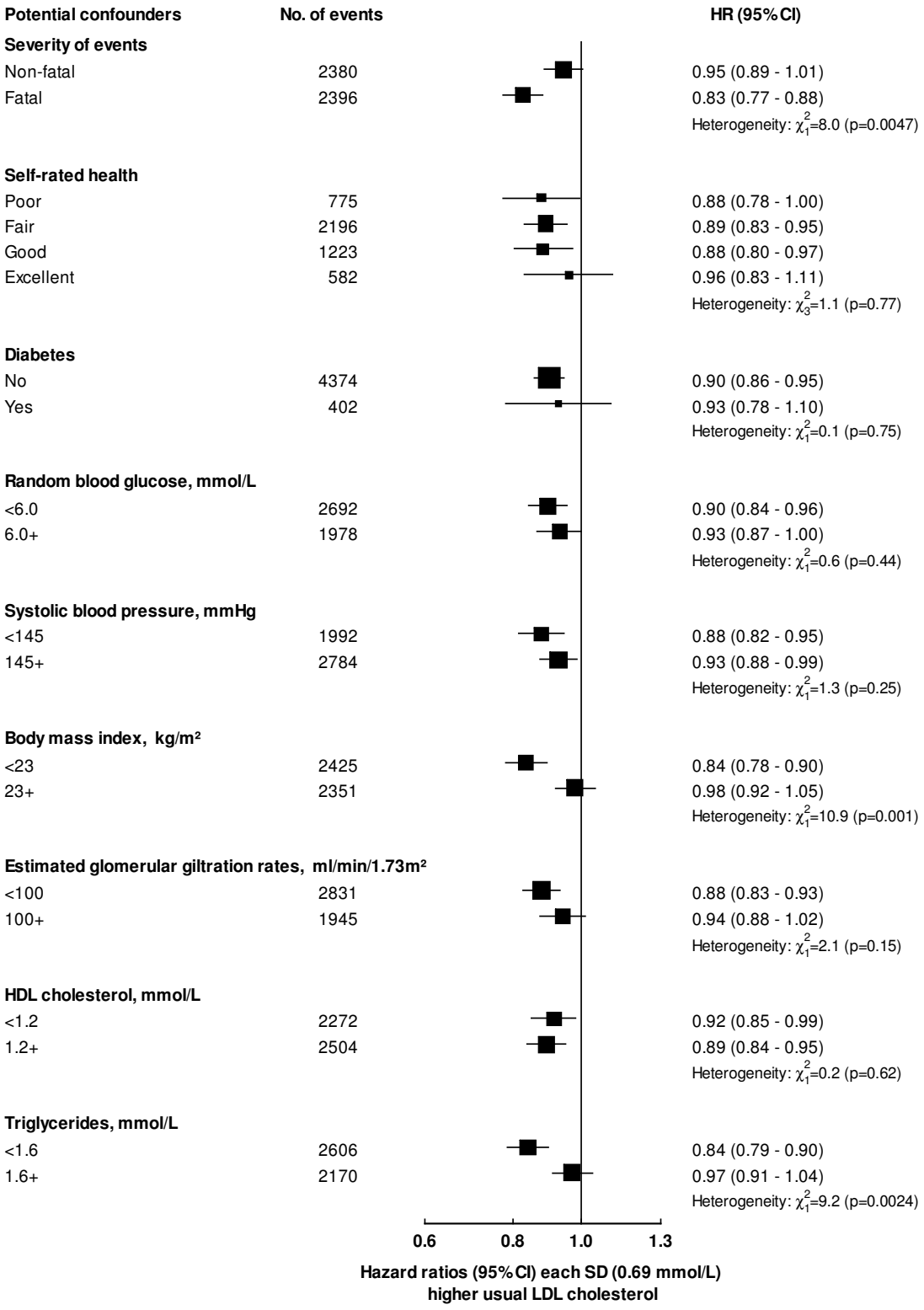
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual LDL cholesterol, per 1 mmol/L higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 5.3: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol in population subgroups of baseline vascular risk factors (I)



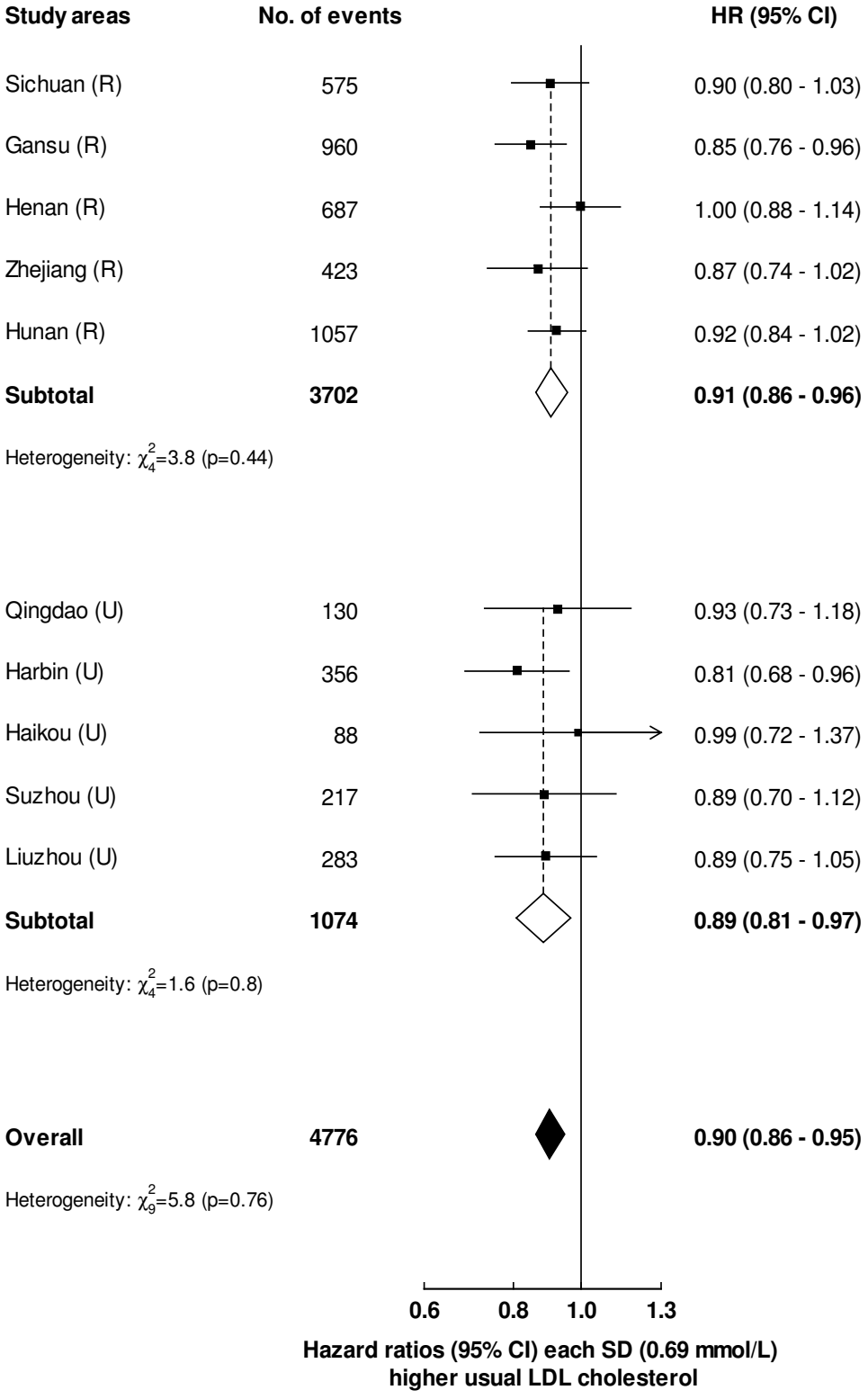
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.4: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol in population subgroups of baseline vascular risk factors (II)



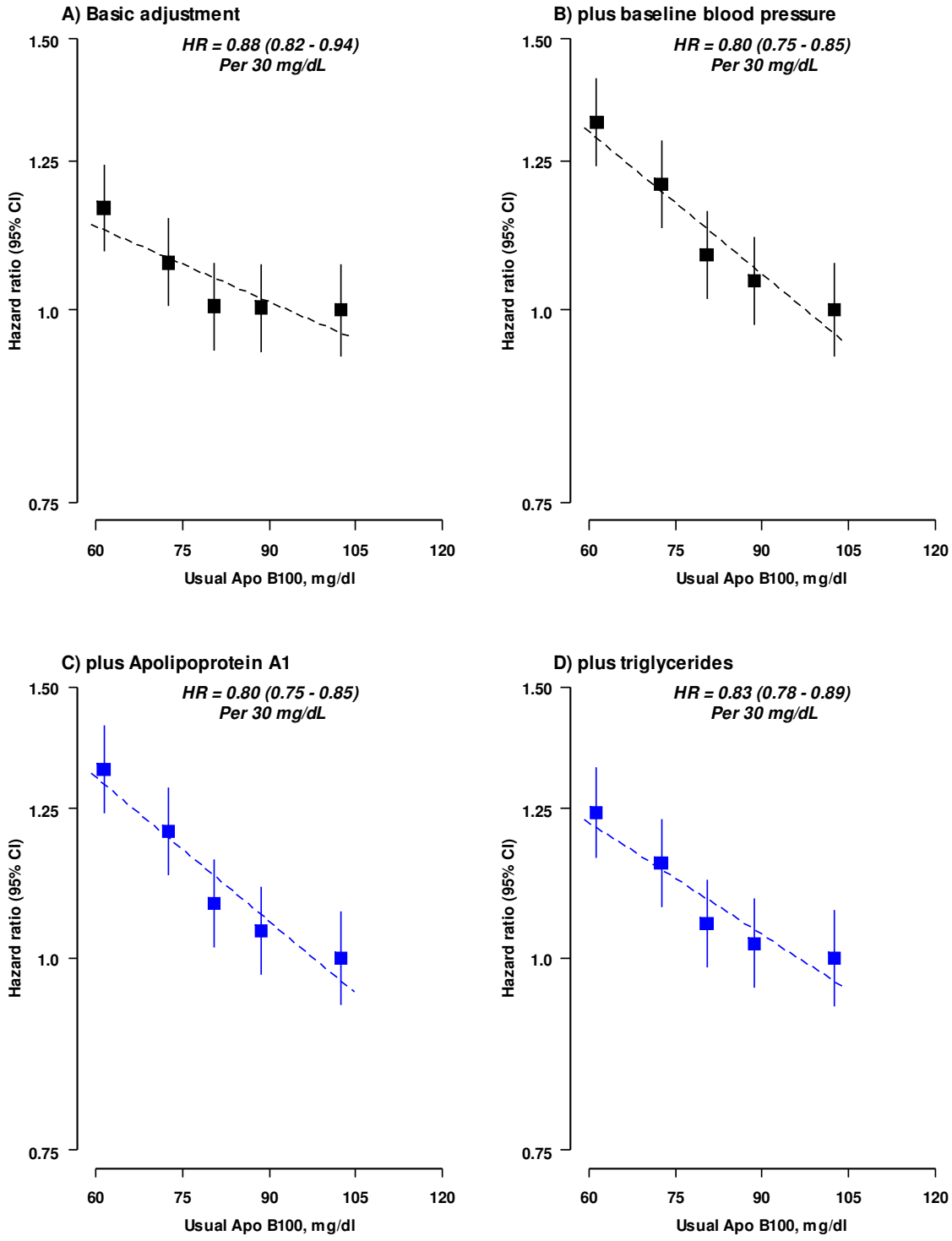
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.5: Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by usual LDL cholesterol, by region



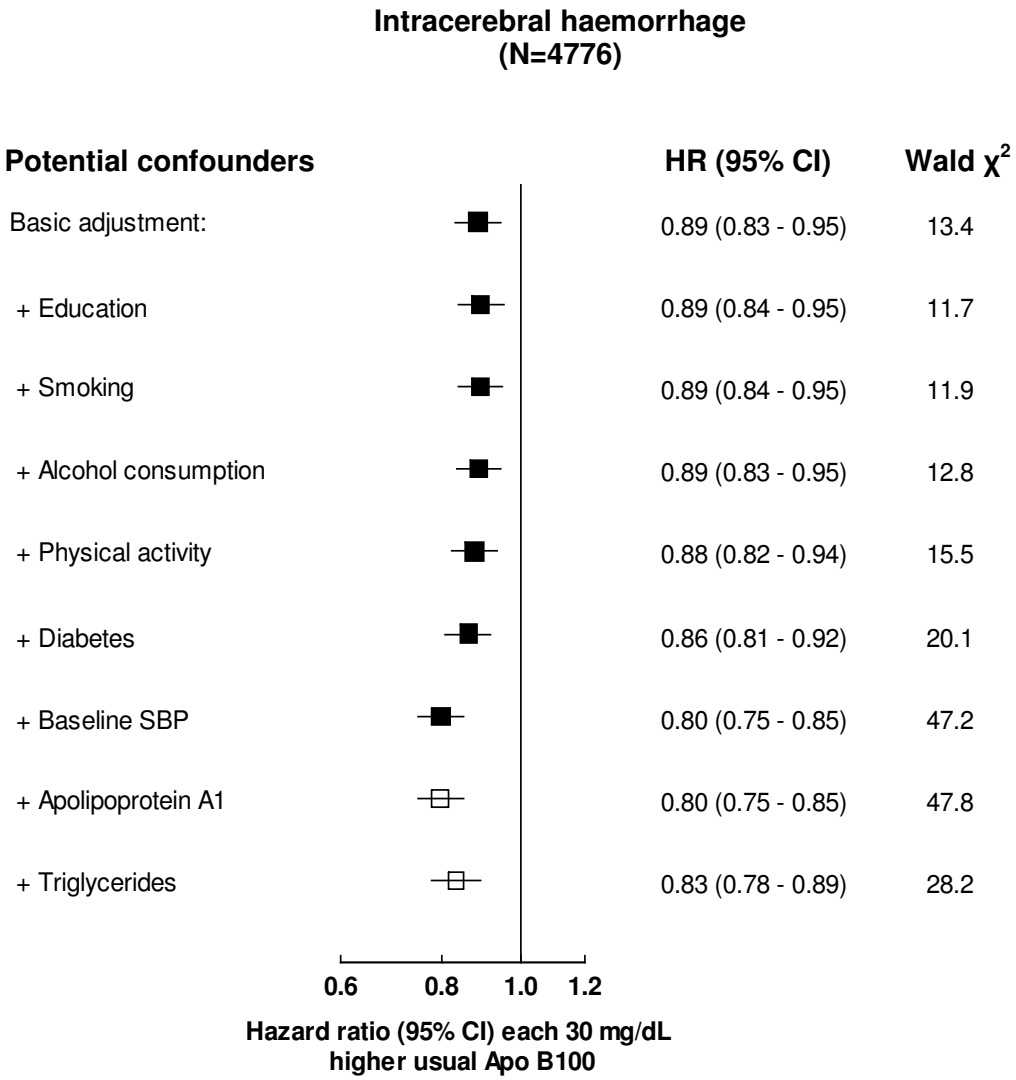
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.6: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein B100 (Apo B100)



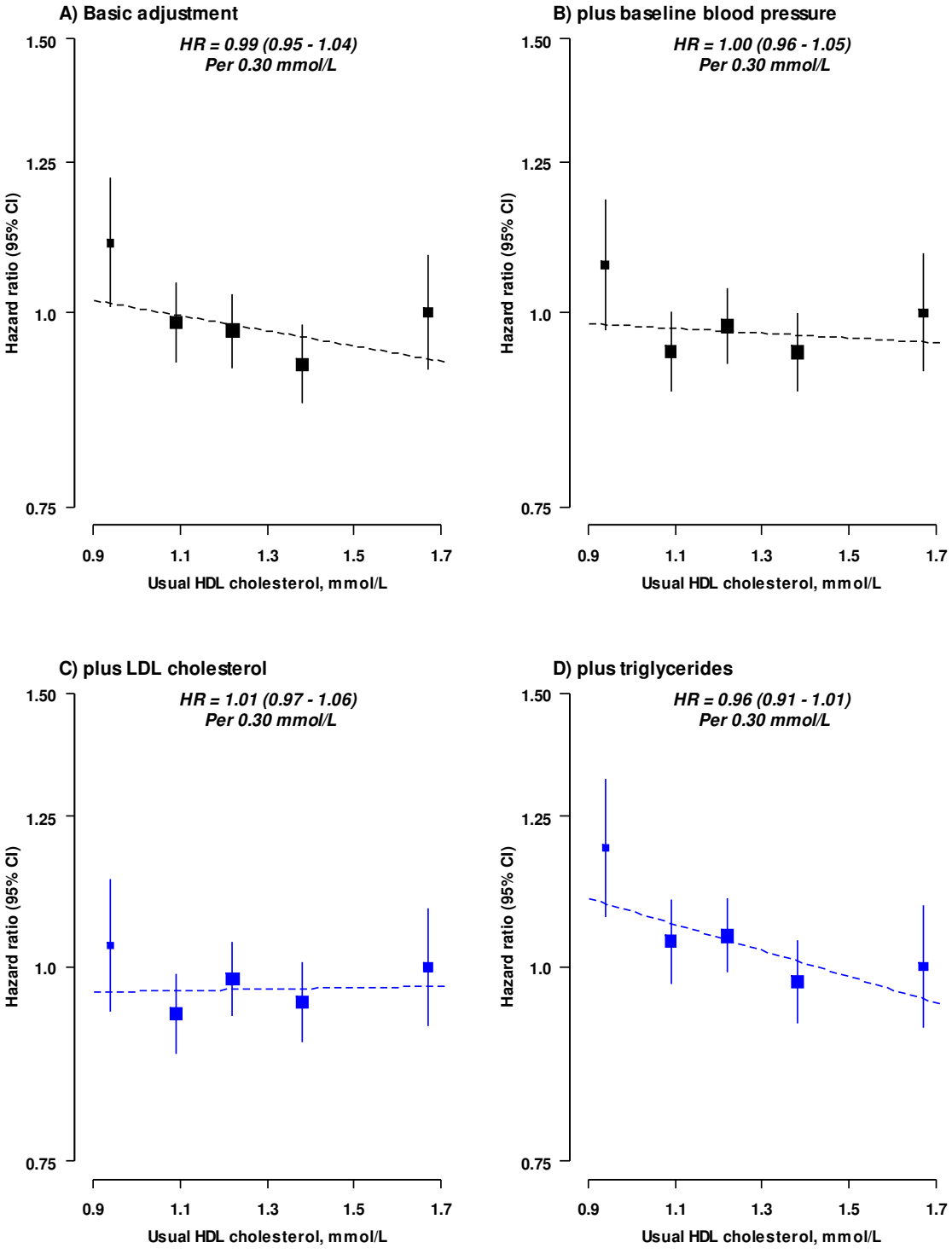
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for Apolipoprotein A1 in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual Apo B100 concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 5.7: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein B100 (Apo B100), progressively adjusted for potential confounding factors



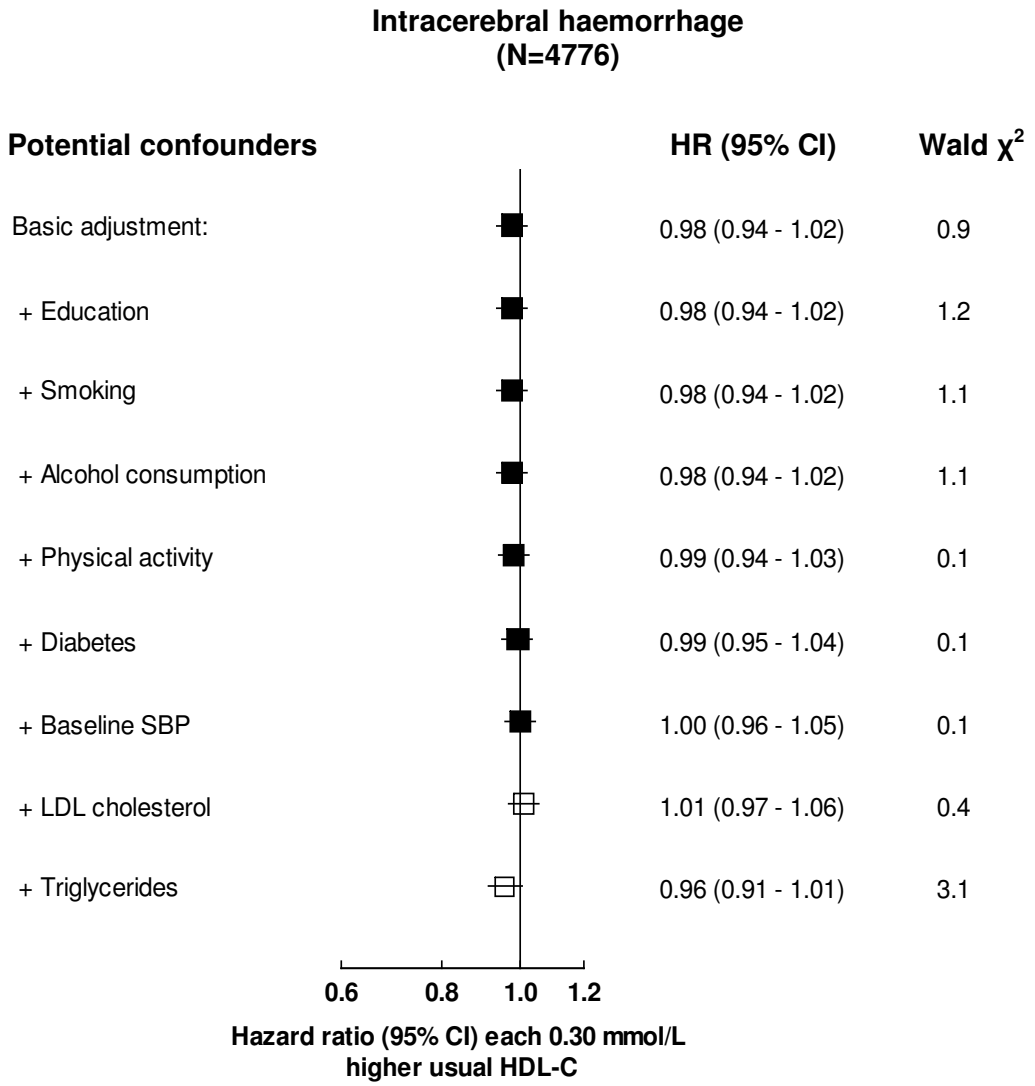
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual Apo B100, per 30 mg/dl higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CIs.

Figure 5.8: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol



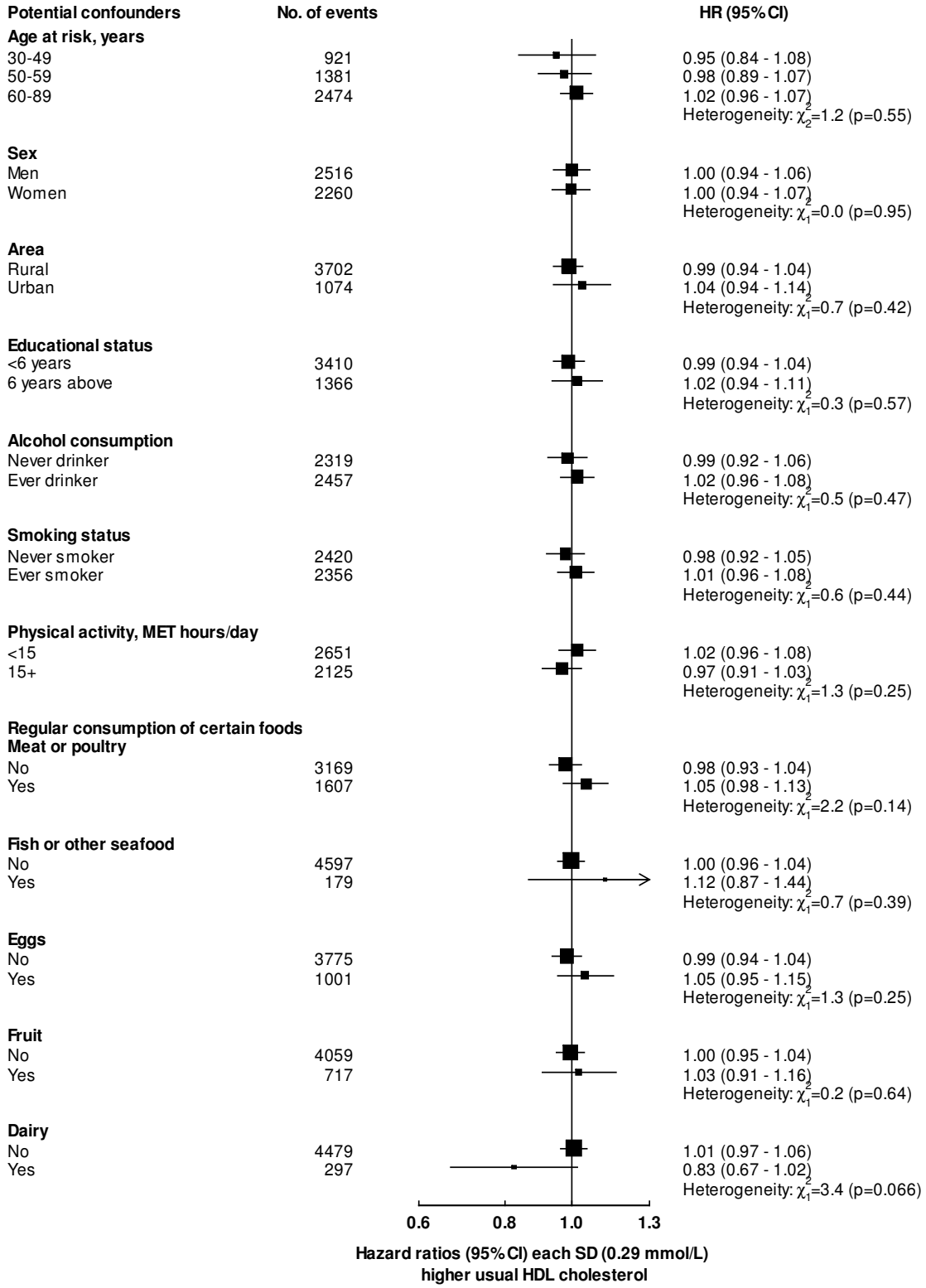
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for LDL cholesterol in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual HDL cholesterol concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 5.9: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol, progressively adjusted for potential confounding factors



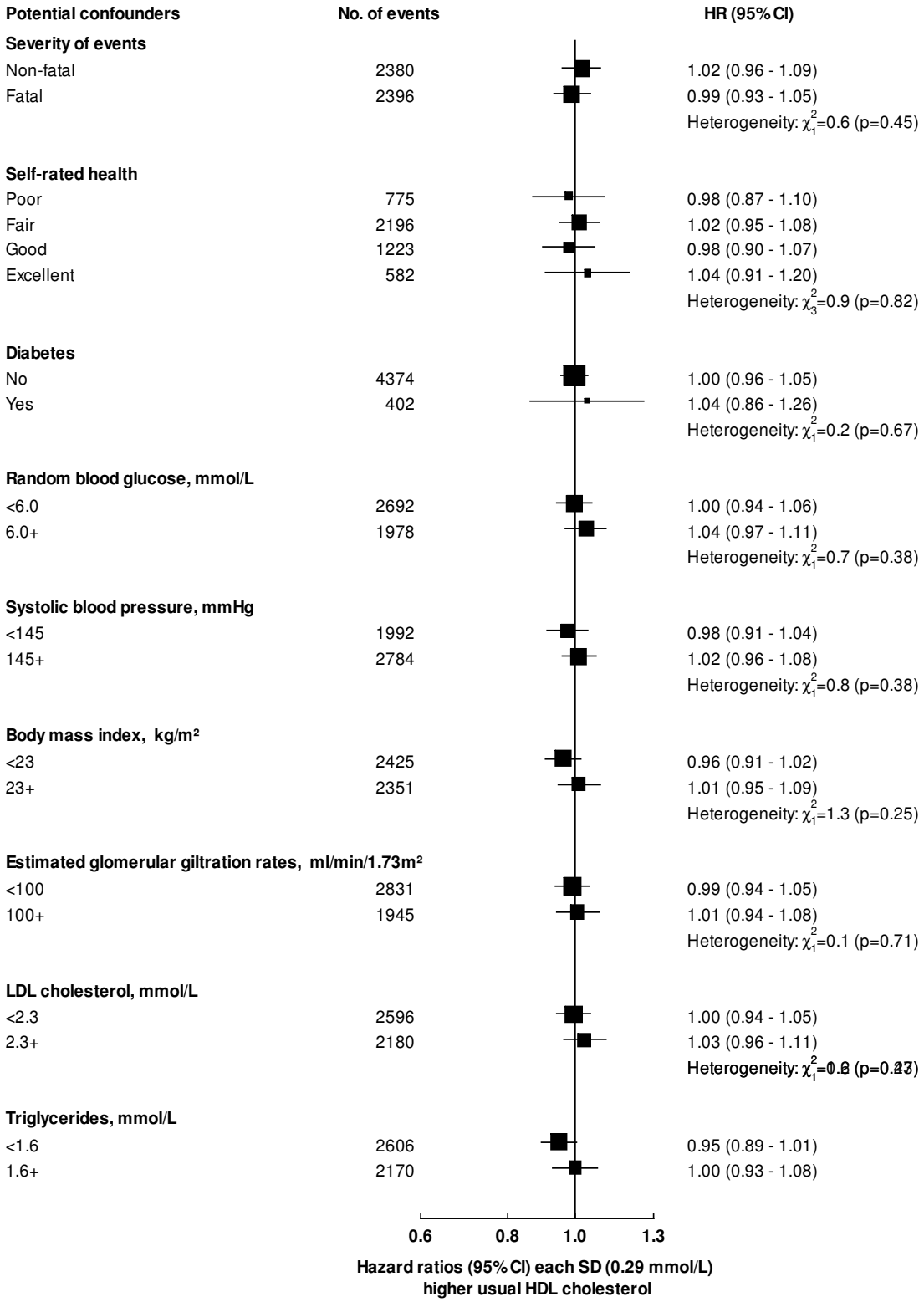
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual HDL cholesterol, per 0.30 mmol/L higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CIs.

Figure 5.10: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol in population subgroups of baseline vascular risk factors (I)



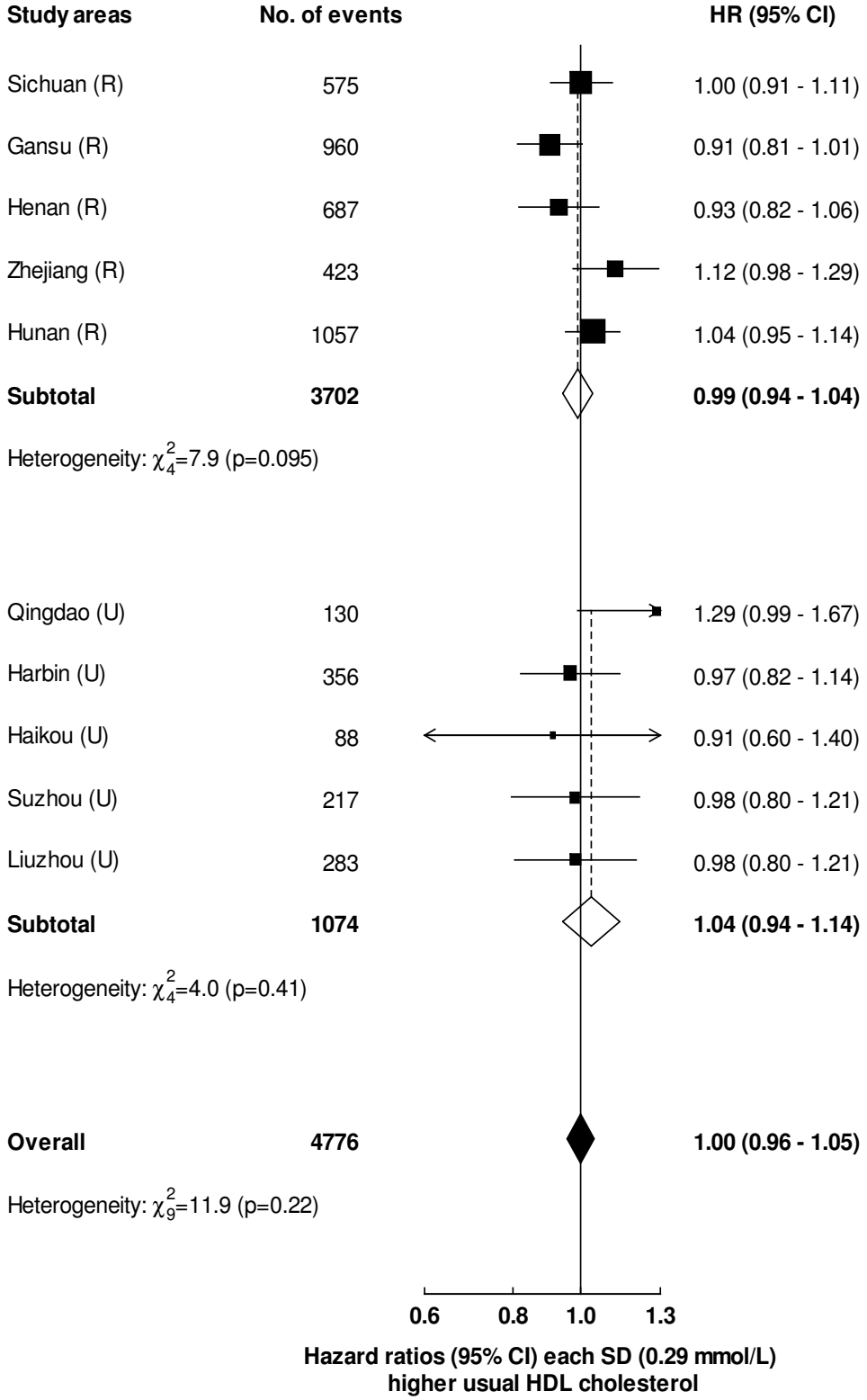
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual HDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.11: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol in population subgroups of baseline vascular risk factors (II)



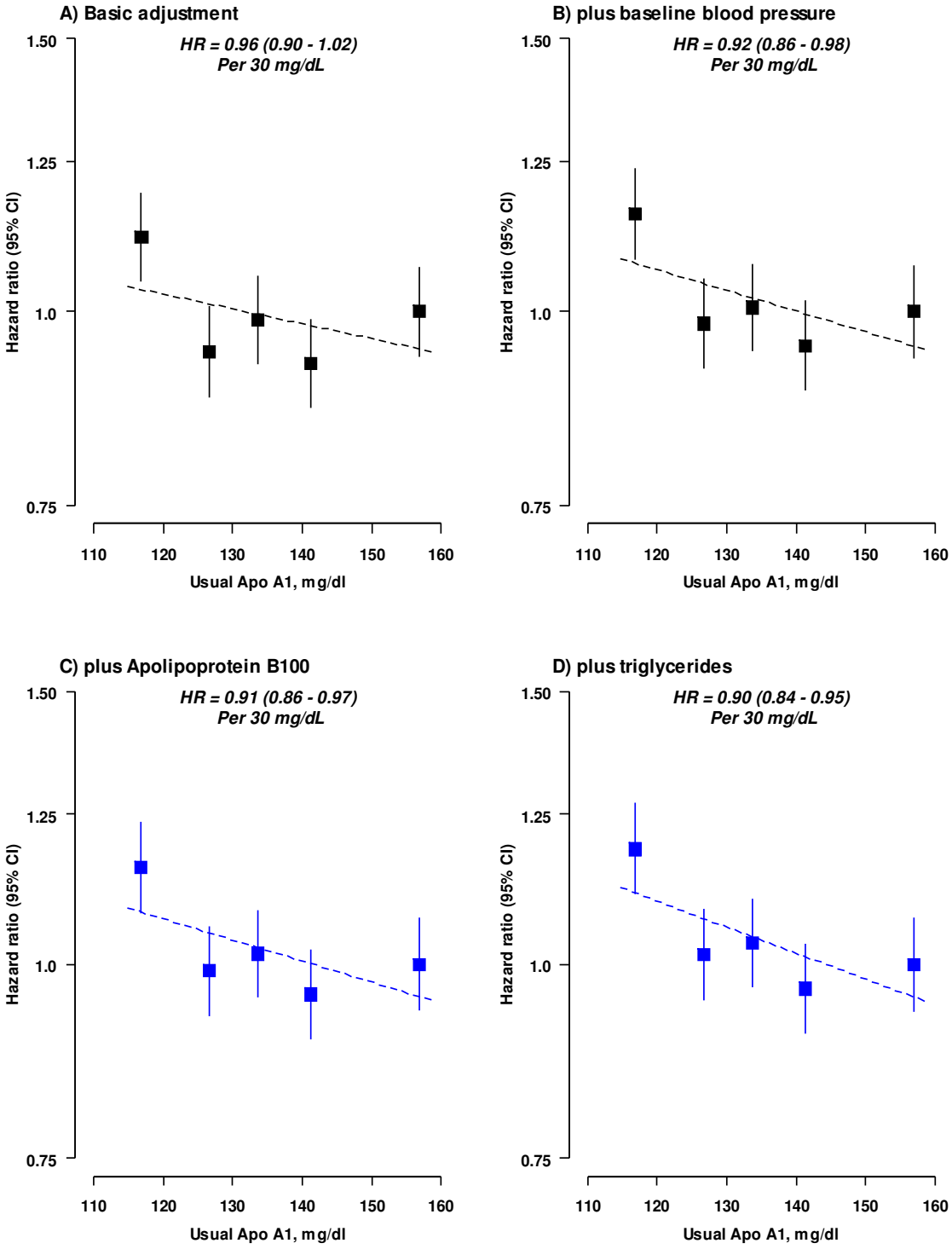
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual LDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.12: Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by usual HDL cholesterol, by region



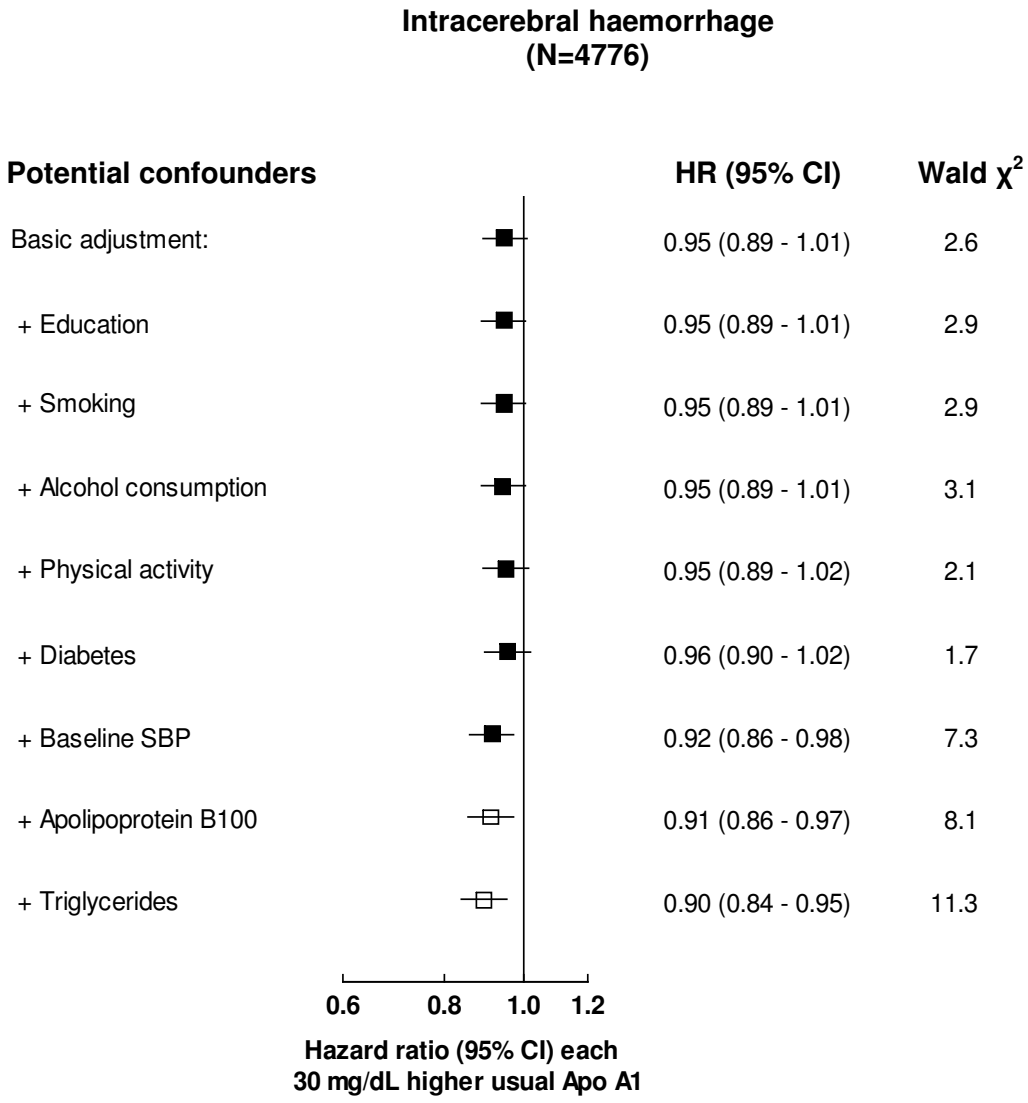
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual HDL cholesterol, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.13: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein A1 (Apo A1)



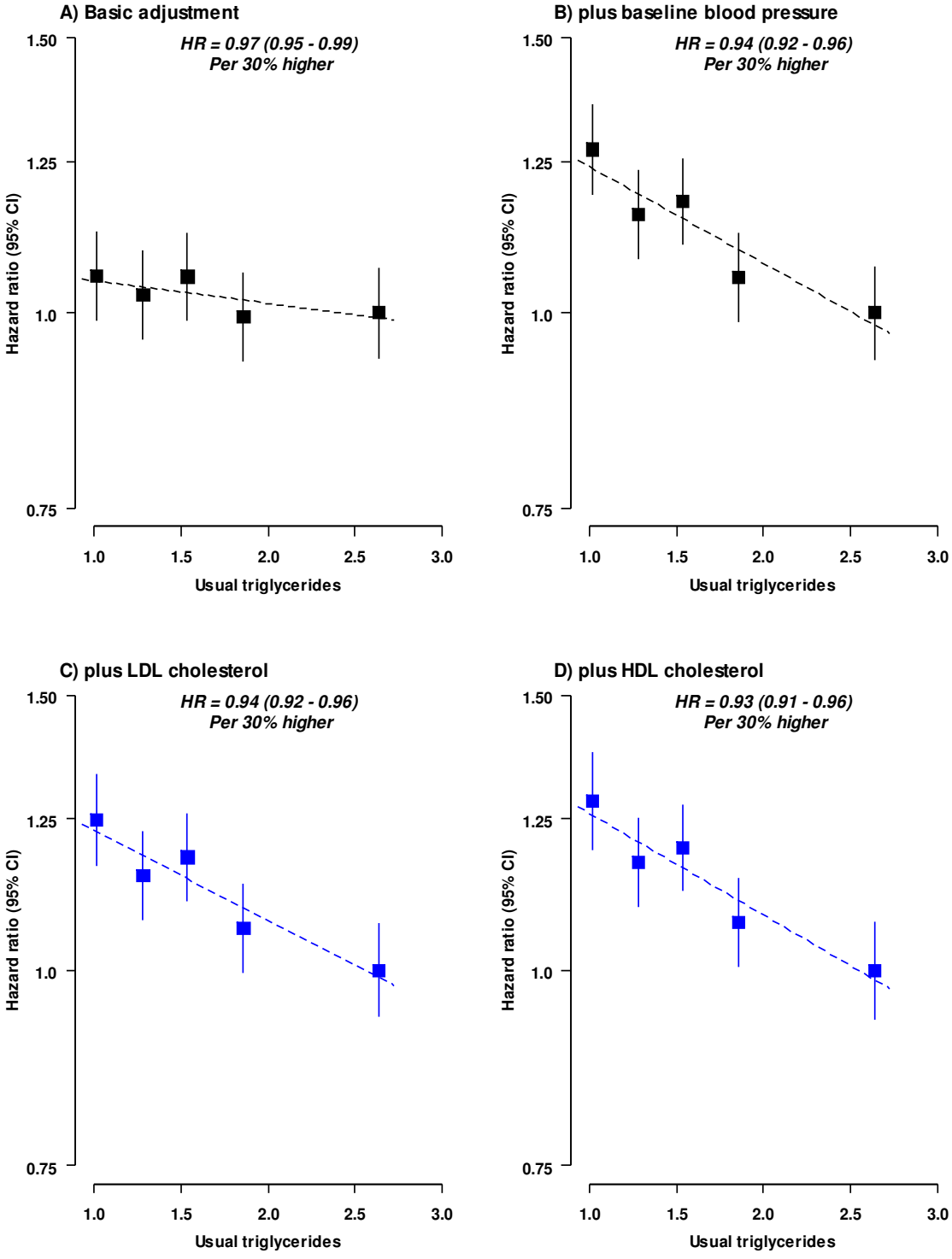
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for Apolipoprotein B100 in addition to those in B); D) further adjusted for triglycerides in addition to those in C). HRs were plotted against mean usual Apo A1 concentrations in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 5.14: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual Apolipoprotein A1 (Apo A1), progressively adjusted for potential confounding factors



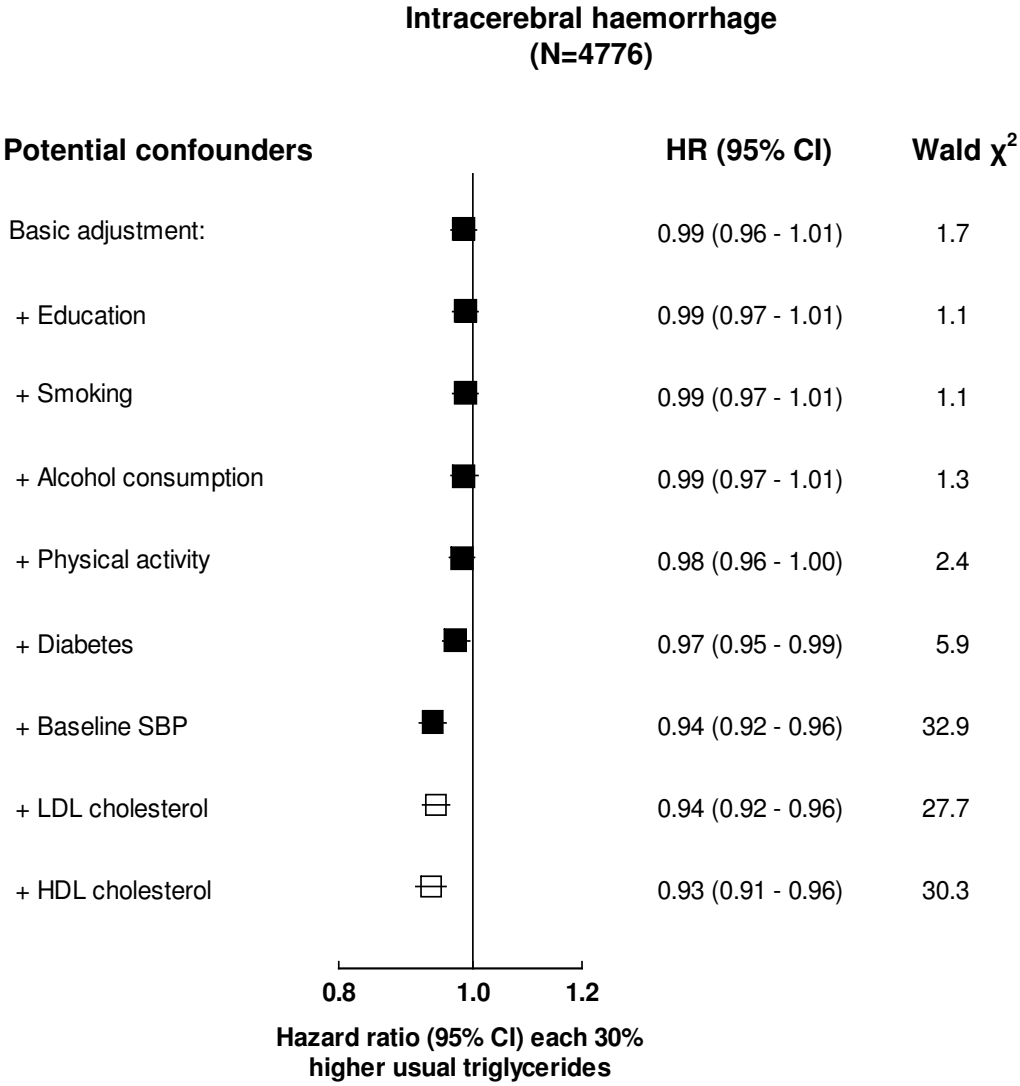
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual Apo A1, per 30 mg/dl higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 5.15: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides



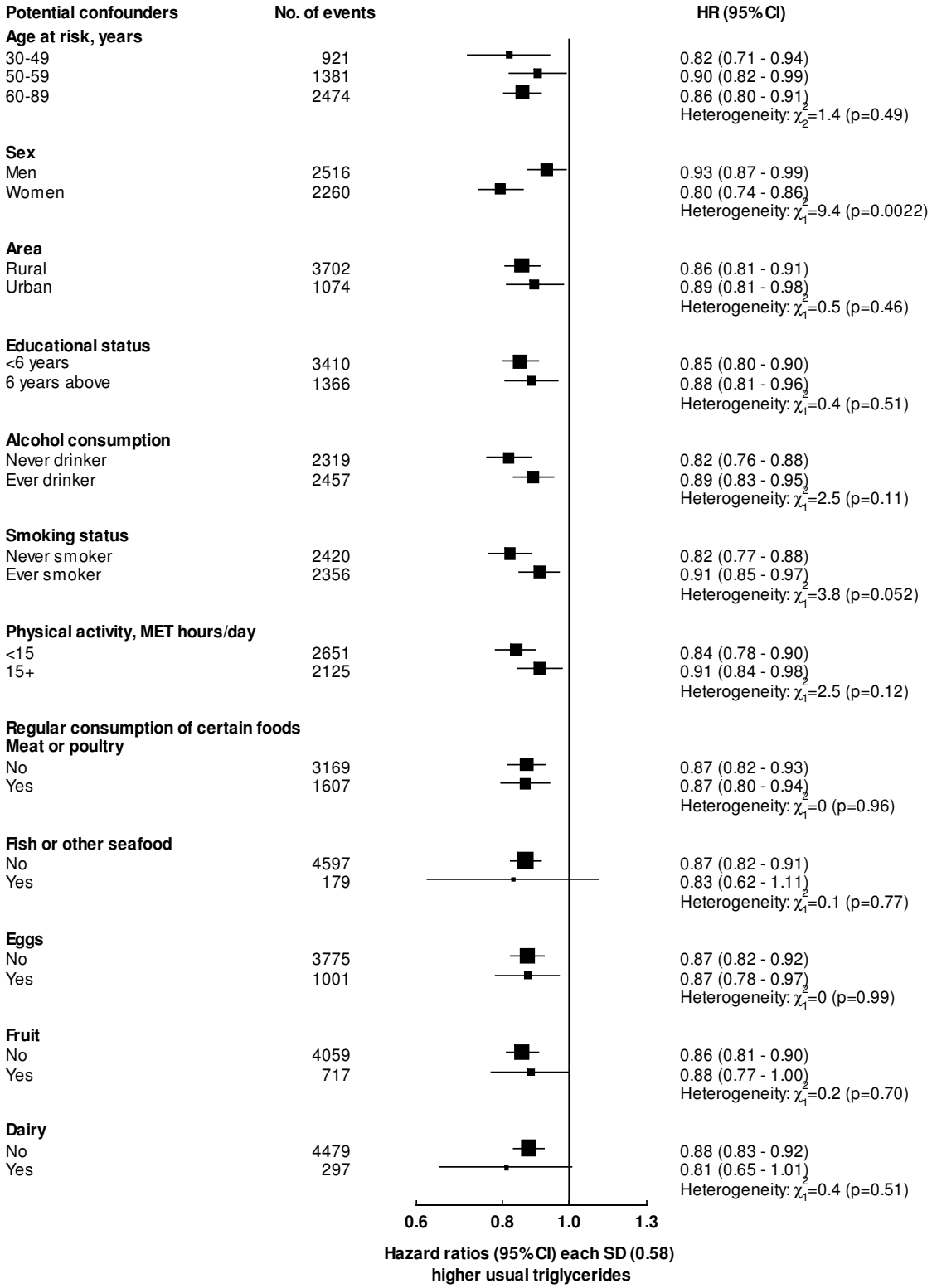
Cox regression was stratified by study region, 5-year age-at-risk, and sex, based on which: A) Basic adjustment was further adjusted for education, smoking, alcohol consumption, physical activity, and baseline diabetes; B) further adjusted for baseline systolic blood pressure in addition to those stated in A); C) further adjusted for LDL cholesterol in addition to those in B); D) further adjusted for HDL cholesterol in addition to those in C). HRs were plotted against mean usual triglycerides in each category. Each square had an area inversely proportional to the effective variance of the log risk in the specific group, with vertical lines representing the 95% confidence intervals.

Figure 5.16: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides, progressively adjusted for potential confounding factors



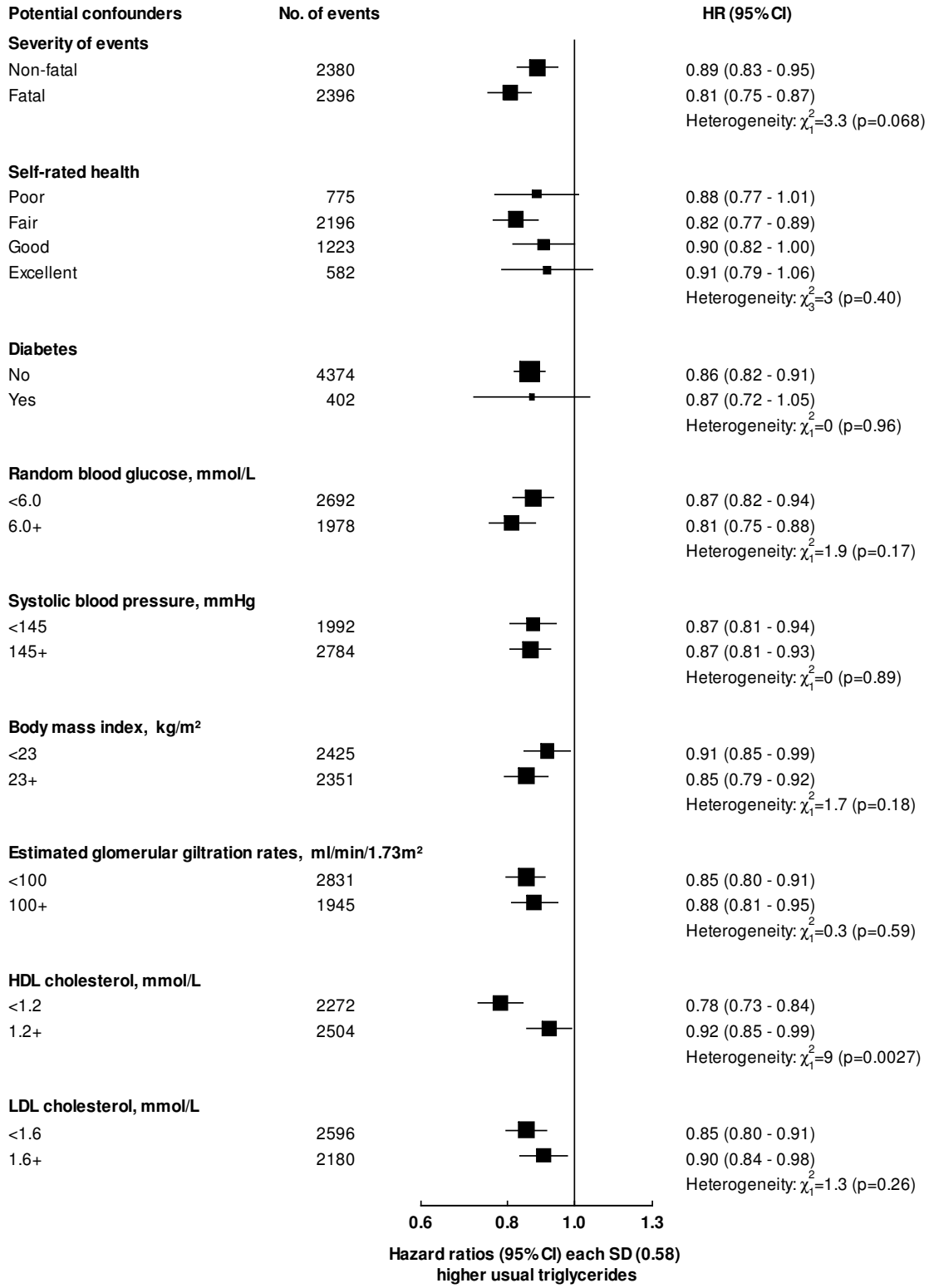
Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual triglycerides, per 30% higher. Basic adjustment was stratified by study region, 5-year age-at-risk, and sex, based on which, analyses were progressively adjusted for potential confounding factors. Squares represented the HRs with area inversely proportional to the variance. Horizontal lines represent the corresponding 95% CI.

Figure 5.17: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides in population subgroups of baseline vascular risk factors (I)



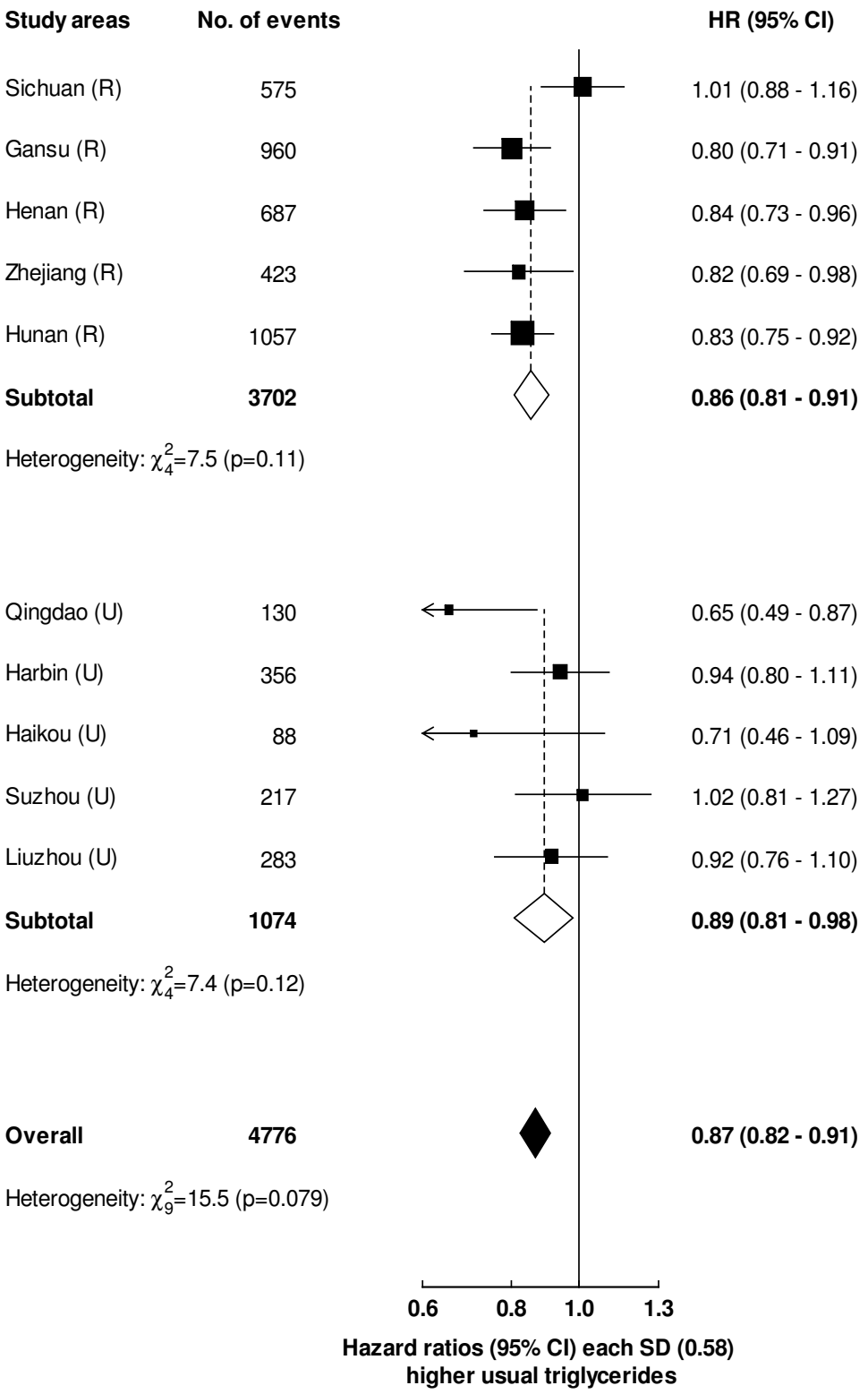
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against usual triglycerides (Log_e), per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.18: Adjusted hazard ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides in population subgroups of baseline vascular risk factors (II)



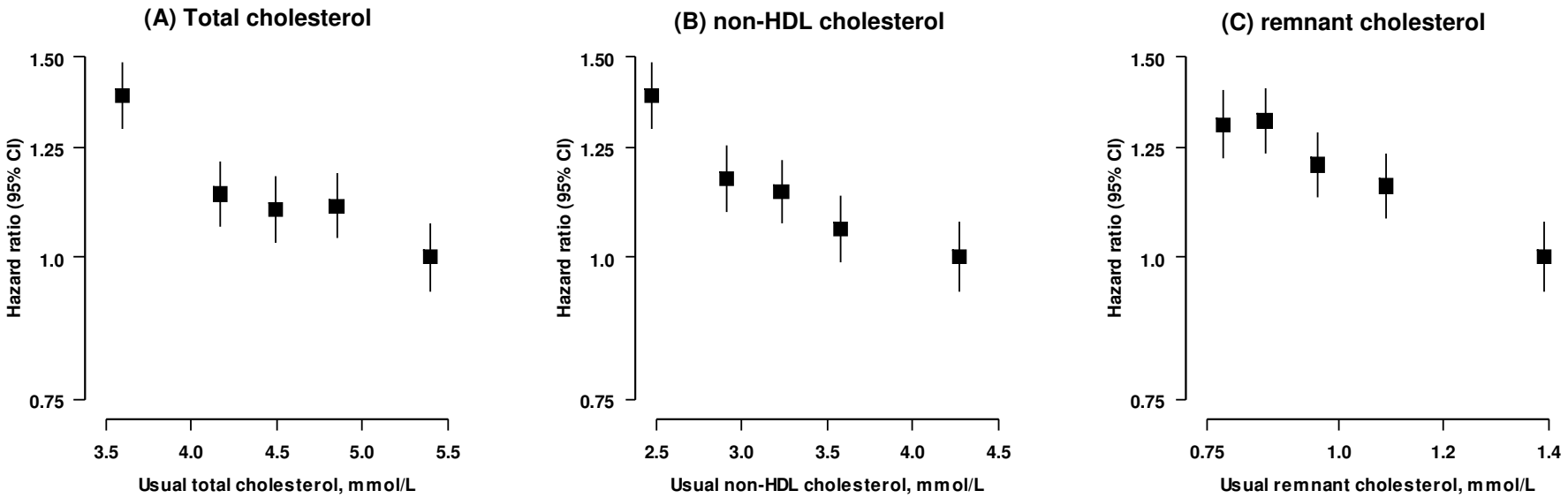
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual triglycerides (Log_e), per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, region, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.19: Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by usual triglycerides, by region



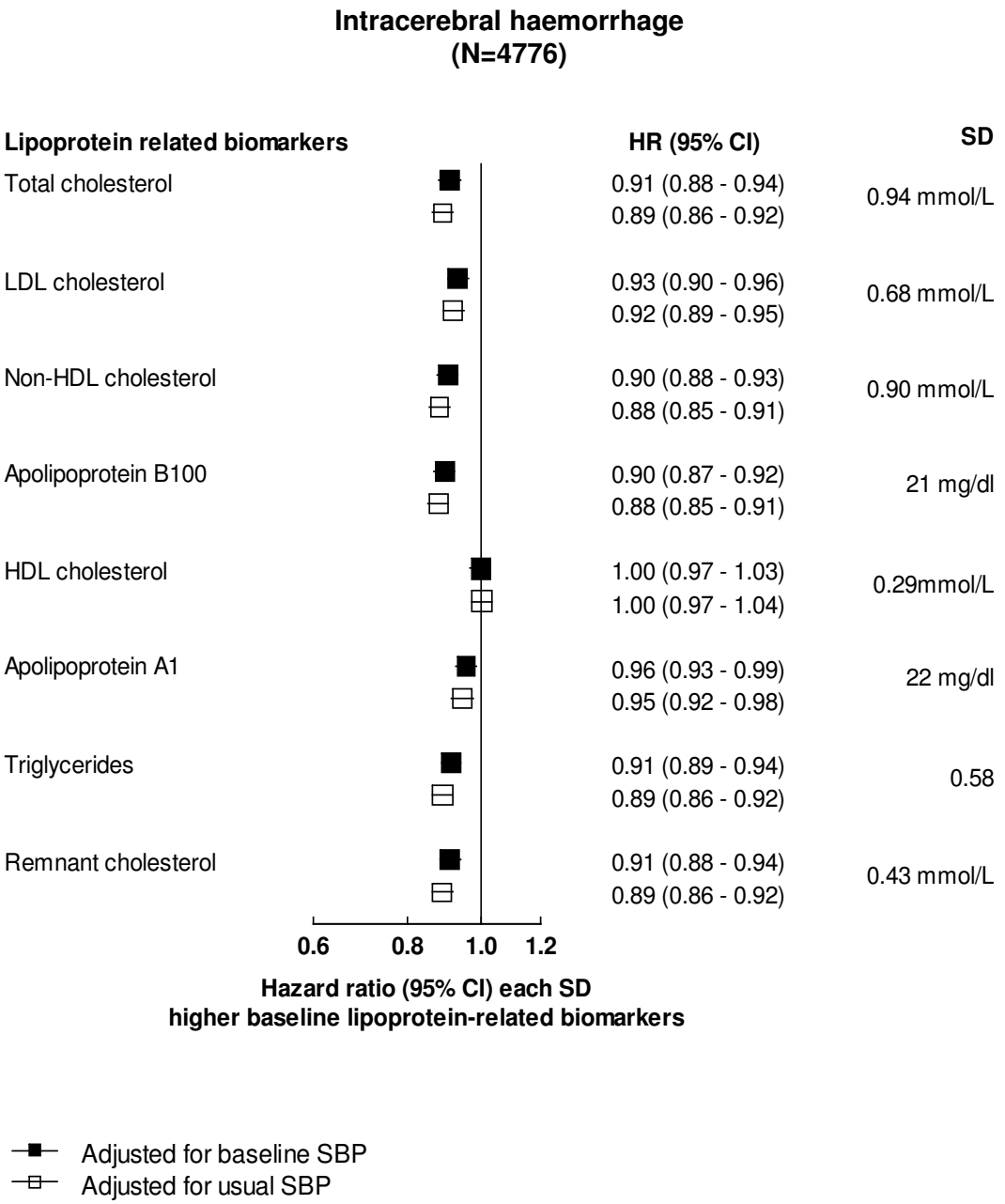
SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of intracerebral haemorrhage against usual triglycerides (Log_e), per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Figure 5.20: Adjusted hazards ratios (HRs) for risk of intracerebral haemorrhage by (A) total cholesterol, (B) non-HDL cholesterol, and (C) remnant cholesterol



Cox regression was used to yield adjusted HRs and the corresponding 95% confidence intervals (CI) for risk of intracerebral haemorrhage against usual levels of lipid fractions in each quintile. Analyses were stratified by study region, 5-year age-at-risk, and sex; and adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure. Each square had area inversely proportional to the effective variance of the log risk in the specific group, with vertical line indicating the variance derived only from that category.

Figure 5.21: Comparison of associations of baseline lipid-related biomarkers with risk of intracerebral haemorrhage, with adjustment for baseline or usual systolic blood pressure (SBP)



SD=Standard deviation; Cox regression was used to yield adjusted HRs and the corresponding confidence intervals (CIs) for risk of ischaemic stroke against baseline lipid-related biomarkers, per 1 SD higher. Analyses were stratified by age-at-risk (5-year), sex, and adjusted for education, smoking, alcohol consumption, physical activity, diabetes, and baseline SBP, or usual SBP. Squares represented the HRs with area inversely proportional to the variance of the log HR, with horizontal lines representing the 95% CIs.

Chapter 6 Major lipids and stroke subtypes: Mendelian randomisation studies

This chapter investigates the causal relevance of the associations of three major blood lipids with stroke pathological types reported in the **Chapters 4-5**, using a Mendelian randomisation approach in a nested case-control study of stroke pathological types in the CKB study. The Global Lipids Genetics Consortium (GLGC), which is the largest GWAS study of lipids, was used as the source of genetic information to select the lipid-related genetic variants.[92] The 1000 Genomes Project (Phase 3) was used as the major reference panel to provide the basic information on genetic markers in populations of Chinese ancestry.[265]

6.1 Introduction

The observational associations of major blood lipids, including LDL-C, HDL-C, and triglycerides with risks of IS and ICH in this Chinese population, have been characterised in previous chapters (i.e., **Chapters 4** and **5**). However, despite of large number of the confirmed stroke cases from a single population, direct measurements of plasma lipid assays, and analyses adjusted for a wide range of potential confounders, assessment of the causal relevance of major blood lipids for risks of stroke pathological types in observational studies is constrained by general limitations of observational studies (e.g., residual confounding and reverse causality), and by limited ability to distinguish IS from ICH. Hence, it is not possible to infer conclusions about the causality of associations of such lipid fractions with risk of stroke pathological types using evidence from the conventional observational analyses alone. In principle, large-scale reliable RCT evidence can overcome many of the limitations and provide the strongest evidence to establish

causality. In practice, however, interventions targeting one lipid fraction often have effects on other risk factors, and RCTs are also constrained by expense, relatively short duration of treatment, and are also restricted to specific populations (e.g., high-risk populations).

MR studies afford an opportunity to overcome several of the limitations outlined above, and to improve the understanding of the causal relevance of modifiable risk factors to disease outcomes. MR studies exploit Mendel's second law that the inheritance of one trait is independent of other traits, and thus reduce the susceptibility to residual confounding.[102] Furthermore, an individual's genotype is determined at conception, and thus associations between the genotype and the disease outcomes cannot be affected by reverse causality.[100] Additionally, genetic variants that are related to modifiable exposures (e.g., LDL-C) will generally be related to such exposures throughout life and, hence, their use for causal inference can also avoid attenuation by random measurement errors (regression dilution bias).[100, 103] Overall, MR approaches have been widely applied in studies on CHD risks and have confirmed the causal role for both LDL-C,[101, 129] and also, for triglyceride-mediated pathways for CHD.[83, 124] However, MR studies appear to have refuted the causal role of HDL-C,[78, 129] albeit substantial uncertainties remain regarding the relevance of HDL functionality for CHD risks.[78] However, the causal relevance of major blood lipids for stroke pathological types has not yet been reliably assessed, particularly in East Asians.

To investigate the causality of the associations observed in **Chapters 4-5**, this chapter uses genetic variants for major blood lipids as instrumental variables to examine the causal relevance of LDL-C, HDL-C, and triglycerides for risk of IS,

and ICH on a nested case-control study of stroke pathological types in the large CKB study.

6.2 Study design and study methods

6.2.1 Study population and study design

The study design and the selection criteria of stroke cases and controls have been described in previous chapters (i.e., **Chapters 3-5**). In the following genetic analyses, for each stroke case, approximately two healthy controls were identified by age, sex, and study region, with prioritisation to ICH cases. As is the case in Chapter 4-5, the new-definition of disease status was used by extending the censoring date to 1st January 2016. To minimise the potential index event bias,[266, 267] participants who developed MI, SAH, and stroke of unknown types, prior to any IS, or ICH events, were excluded. Consistent with the analysis of the observational associations, only the first incident stroke events for each type were counted, and ICH was counted if both types of stroke (i.e., ICH and IS) were recorded on the same day. Overall, there were 4911 ICH cases, 5567 IS cases, 9742 controls included in the following analyses.

6.2.2 Instrumental variables

The selection of genetic variants was chiefly based on the underlying biological mechanisms of the genetic markers. Based on available data, three slightly different approaches were applied when conducting the MR studies on LDL-C, HDL-C, and triglycerides, respectively. A GRS of eight SNPs was used as an instrument variable for LDL-C. Common *CETP* haplotypes were constructed by incorporating five SNPs within the *CETP* gene region, and used as an instrument

variable for HDL-C. A GRS of four SNPs involved in triglycerides lipolysis process was used as an instrumental variable for triglycerides.

6.2.2.1 Low-density lipoprotein cholesterol

Selection of genetic variants

The SNPs used to construct the genetic instrument for LDL-C were selected among those which had been associated at $p<5x10^{-8}$ with LDL-C in the GLGC study.[92] There were 28 LDL-C SNPs captured in the CKB genotyping panel. As, in the GLGC study, some of the LDL-C SNPs were also correlated with HDL-C, and triglycerides, those with strongest evidence of associations with LDL-C (i.e., LDL-C as the primary trait), or if LDL-C was not the primary trait but only secondary to TC, were selected in the thesis. Of the LDL-C SNPs, those with primary traits of HDL-C (i.e., rs1800775, rs2126259, rs3764261), triglycerides (i.e., rs16996148, rs2954029, rs58542926, rs662799, rs964184), or lipoprotein (a) (i.e., rs10455872, rs3798220, rs3798221, rs6415084) were excluded.

After excluding LDL-C related SNPs that were correlated to other lipid traits, the remaining SNPs (n=16), with large effect estimates in the GLGC study were prioritised and one SNP was selected for each individual gene. Due to the relatively small difference of the p-values between LDL-C ($p=1.57X10^{-9}$) and HDL-C ($p=6.58X10^{-8}$), *C12orf30* (rs11066320) was removed. After these exclusions, eight LDL-C associated SNPs located in different genes remained, including rs2000999 (*HPR*), rs11206510 (*PCSK9*), rs12916 (*HMGCR*), rs579459 (*ABO*), rs1713223 (*APOB*), rs6511720 (*LDLR*), rs602633 (*SORT1*), and rs7412 (*APOE*). **Table 6.1** shows the basic information on the eight selected SNPs, based on data from the GLGC, and the 1000 Genome Project.[265]

Among the eight SNPs, two SNPs were on chromosome 1 (i.e., rs11206510 and rs602633), and two SNPs were on chromosome 19 (i.e., rs6511720 and rs7412). Linkage disequilibrium (LD) statistics were calculated among Chinese population included in the 1000 Genome Project (n=208).[265] The R^2 were 0.005 and 0.0007, respectively, for pairs of rs11206510 and rs602633, and rs6511720 and rs7412 (**Table 6.2**). Hence, both pairs of SNPs were included when constructing the GRS for LDL-C.

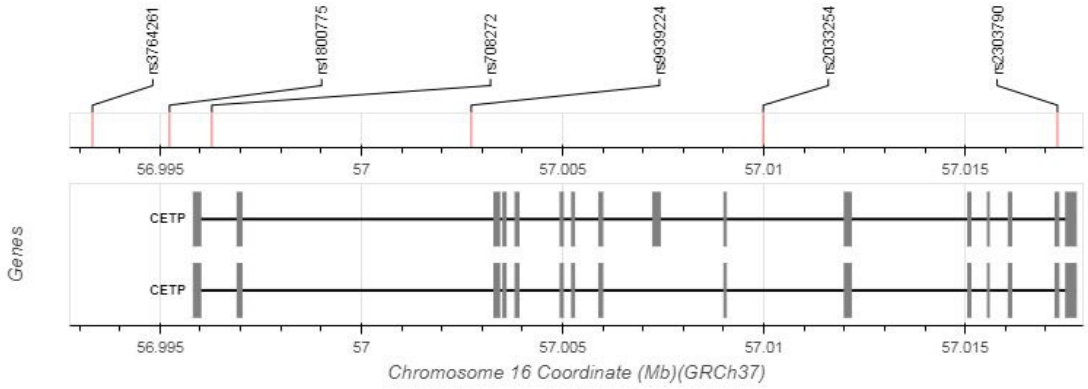
Construction of genetic risk scores

To maximise the variation on plasma LDL-C concentrations explained by the genetic instrumental variable, a GRS was constructed by adding multiple genetic variants either as an unweighted or weighted sum of LDL-C raising alleles.[268] The unweighted GRS was constructed simply by addition of the LDL-C raising alleles each individual has. The weighted GRS was subsequently constructed by multiplying the number of the LDL-C raising alleles by the external effects on LDL-C observed in the GLGC,[92] and then adding these together.

6.2.2.2 High-density lipoprotein cholesterol

Similar inclusion criteria as described for LDL-C were used for HDL-C, seven SNPs within two gene regions were identified within the available CKB genotyping data, including *LIPC* (rs1532085) and *CETP* (6 SNPs: rs3764261, rs1800775, rs708272, rs9939224, rs2033254, and rs2303790) genes. *LIPC* encodes hepatic triglyceride lipase, whereas CETP is a plasma protein that facilitates the exchange of cholesteryl esters and triglycerides between lipoproteins (i.e., triglyceride-rich lipoproteins and HDLs).

Figure 6.1 shows the relative positions of the six genetic variants within the *CETP* gene region, and **Table 6.3** provides the relevant LD information. Common *CETP* haplotypes were constructed with the genetic data analysis package SimHap,[269] which was also used to examine the associations between common *CETP* haplotypes and plasma concentrations of HDL-C, LDL-C, triglycerides, and levels of BMI, and SBP. SimHap uses estimation-maximisation based methods for estimation of haplotypes from un-phased genotype data,[270] and provides the posterior probabilities of each possible haplotype. All possible haplotype configurations were resolved for each individual within the program itself. In the thesis, the most likely haplotype was assigned for each individual, with a haplotype frequency above 5%.



Centromere ***Telomere***

Figure 6.1: Diagrammatic representation of the *CETP* region, showing the relative position of informative variants, based on Chinese population (n=208) in 1000 Genome Project (Phase 3)

6.2.2.3 Triglycerides

Selection of genetic variants

Previous GWAS studies have identified relatively limited number of genetic variants associated solely with triglycerides, but not with other lipid fractions

(n=16). Of the 16 triglyceride associated loci (out of 39) in the GLGC study, the effect sizes of each on triglycerides were relatively moderate, and none have been genotyped in the available CKB genetic data. The triglyceride associated genetic variants included in the present study were chiefly involved in the lipolysis process, and consequently, all of them were also highly correlated with HDL-C. **Table 6.4** shows information on selected triglyceride associated loci in the present study, based on the 1000 Genome Project and the GLGC study. Genetic variants for *LPL* (rs12678919) and *TRIB* (rs2954029) are on the same chromosome (Chromosome 8), and the D' and R^2 between these two SNPs were 0.107 and 0.001, respectively. Both genetic variants were included in the construction of the GRS for triglycerides.

Construction of genetic risk scores

The unweighted GRS for triglycerides was constructed by adding the effect alleles an individual has, and the weighted GRS for triglycerides was constructed through multiplying the number of effect alleles by the triglyceride effect sizes as reported in the GLGC study for each genetic variant, and then adding up.

6.2.3 Lipid measurements

The baseline plasma lipid assays were available for 16 216 individuals (i.e., 4728 ICH cases, 5259 IS cases, and 6229 ICH controls), using direct assays measured at the CTSU laboratory (**Chapter 3**). Briefly, LDL-C and HDL-C were measured using direct homogenous method, and triglycerides were assayed using enzymatic approach.

6.2.4 Stroke ascertainment

A stroke adjudication exercise is ongoing to assess the accuracy of diagnosis of stroke pathological types and its subtypes, in addition to several other main diseases of interest in the CKB study, details of which have been described in **Chapter 3** (**Section 3.3.3**). Preliminary results of this adjudication exercise indicated a very high level of use of CT/MRI imaging in 93% of all stroke cases, and a high confirmation rate of stroke pathological types, albeit there is some misclassification within subtypes of IS, and ICH, respectively.

6.2.6 Statistical analysis

The baseline characteristics of the study population were described by disease status, after direct standardisation to the age at baseline (10-year), sex, and study region structure of the overall study population included in the present study.

An additive (per allele) genetic model was used for individual genetic variants, assuming the effect of each additional allele of interest was additive.[271] General linear regression, or logistic regression were used, respectively, for continuous variable, or binary variable, after adjusting for sex, age at baseline, and age at baseline squared. All analyses were performed separately by region, and then combined across region through an inverse-variance-weighted fixed-effect meta-analysis to yield an overall estimate of the effect size.

The external weighted GRS was used as an instrumental variable for LDL-C. Causal effects of each 1 mmol/L higher genetically-determined LDL-C was estimated using the ratio (or Wald) method,[272, 273] as this method relies only on a linear relationship between the conditional expectation of the outcome (or in the

binary case, the logistic function of the probability of the outcome and the risk factor).[274] Linear regression or logistic regression were used, respectively, to estimate the associations of the GRS with the risk factors, and with stroke risks, adjusted for sex, age at baseline, and age at baseline squared. The regression analyses were performed by region, separately, and then combined across regions through inverse-variance-weighted meta-analyses to construct the overall β_1 (se_1) and β_2 (se_2). The causal estimates were calculated as $\beta_3=\beta_2/\beta_1$, with standard errors derived from $se_3=se_2/\beta_1$.[95, 273]

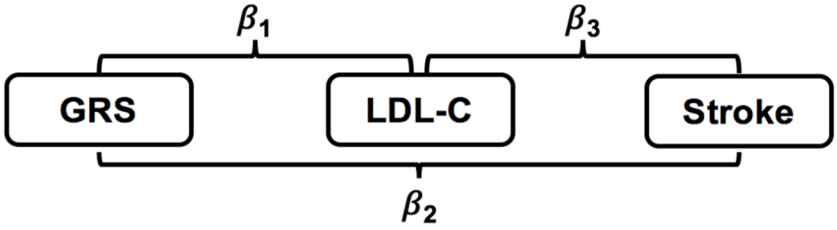


Figure 6.2: Graphical display of the beta estimates in ratio method

GRS=Genetic risk score; LDL-C=Low-density lipoprotein cholesterol; β_1 is the effect size of the instrumental variable (i.e., GRS) on the risk factor (i.e., LDL-C [in mmol/L]), derived from general linear regression, after accounting for differences in age at baseline, age at baseline squared, sex, and region. β_2 is the log odds ratio estimates yielded from logistic regression of stroke (binary levels) against the GRS, adjusted for age at baseline, age at baseline squared, sex, and region. β_3 was calculated through dividing β_2 by β_1, and se_3 was estimated as the ratio of se_2 to β_1.[273, 275, 276]

Figure 6.2 displays the relationship between the beta estimates used to analyse the genetic associations of LDL-C with risk of stroke pathological types. P values were presented without adjusting for multiple testing. All analyses were conducted with SAS® version 9.3 (SAS Institute Inc.), and R version 3.3.

6.3 Results

6.3.1 Baseline characteristics of participants in the nested case-control study

Descriptive analyses were conducted on 20 220 participants, including 5567 IS cases, 4911 ICH cases, and 9742 common controls. After direct-standardisation to the age, sex, and study region structure of the study population examined, IS cases were slightly younger than controls (54.6 [11.0] versus 57.9 [11.0] years), while the ICH cases were slightly older (59.0 [10.4] years) (**Table 6.5**). Proportions of women, and urban residents were largely similar between ICH cases and controls, while IS cases had a slightly lower proportion of women, and a much higher proportion (two-fold) of urban residents, compared to controls.

Overall, among the selected participants in the nested case-control study of stroke, the distribution characteristics of lifestyle factors were similar between both types of stroke cases, and controls. Sex-differences were observed for several major lifestyle risk factors in this nested case-control study. Overall, around one-third of the participants were current-smokers (around 61% in men, and 3% in women). About one-sixth of the participants reported drank alcohol at least once on a weekly regular basis (around 31% in men, and 2% in women).

Overall, in the present study, less than 20% of the individuals consumed fresh fruit daily, with around one-third, and one-fifth reported, respectively, consuming meat, or egg products daily. The regular consumption of other foods (i.e., fish or other seafood, and dairy products) was not common in this population, and was consistently lower in ICH cases than in either IS cases or controls.

Importantly, mean levels of baseline SBP were much higher in ICH cases than in controls (153 versus 135 mmHg), which was twice the difference in SBP between IS cases (145 mmHg) and controls. Similar patterns were observed for DBP, but the differences were less extreme. The mean levels of baseline BMI were similar between cases and controls. Overall, there was little difference in mean plasma concentrations of lipid-related biomarkers between stroke cases and controls.

At baseline, proportions of prevalent diabetes (including both doctor-diagnosed and screen-detected through RBG) in stroke cases were almost double that in controls (11.1% in IS, 9.0% in ICH, versus 5.5% in controls), as were the proportions with hypertension (25.0% in IS, 26.3% in ICH, versus 11.4% in controls). About 15% of stroke patients, irrespective of the pathological types, rated their health status as poor, compared to around 10% in controls.

6.3.2 Causal estimates of major lipids with risk of strokes

6.3.2.1 Low-density lipoprotein cholesterol

The externally-weighted GRS for LDL-C, estimated by combining the available LDL-C associated genetic variants, was strongly associated with plasma LDL-C concentrations, but was not associated with the other major blood lipid fractions, or other major vascular risk factors. In the present genetic analysis, the ORs, for each 1 mmol/L higher genetically-determined LDL-C, were 1.28 (95% CI: 1.05-1.56) for IS, and 1.04 (95% CI: 0.87-1.25) for ICH, respectively.

Figure 6.3 compares the results of individual SNP associations with plasma LDL-C concentrations in the CKB study with those in the GLGC study. In the GLGC study, approximately 6% of the participants were Asians, while the remainder were

populations of European ancestry. Overall, the effect sizes of the selected LDL-C genetic variants estimated in the present study were similar to those reported in the GLGC study, albeit as expected, had slightly wider confidence intervals. In the present CKB study, the relevance of the genetic marker of *APOE* (rs7412) gene was stronger than that reported in the GLGC study ($p < 0.0001$).

In CKB, none of the eight LDL-C genetic variants were associated with stroke risks at the GWAS threshold (i.e., $p<5X10^{-8}$), after accounting for differences in study region, age, and sex (**Table 6.6**). Regarding risk of incident IS, the proportional relevance was positive for six genetic variants, although the confidence intervals were all compatible with null. There was no evidence of any association of LDL-C genetic variants with risk of ICH.

The GRS was constructed by multiplying the number of LDL-C raising alleles by their effects on plasma LDL-C concentrations as estimated in the GLGC study. An important assumption of MR studies is that the investigated instrumental variable is associated with the outcome of interest only through the exposure, and not mediated through other risk factors for stroke. The validity of this assumption was tested by relating the GRS for LDL-C with known measured risk factors (**Figure 6.4**). There was strong evidence of a positive association of the externally-weighted GRS for LDL-C with plasma LDL-C concentrations, and around 7.2% of the total variance in LDL-C concentrations was explained by the GRS (F-statistic around 1400). In contrast, the GRS for LDL-C was not correlated with HDL-C, triglycerides, or other major vascular risk factors (e.g., baseline BMI and SBP).

The confidence intervals of the causal estimates were calculated as the first term of the “delta method” expansion for the ratio of two random variables, and the

results differed marginally, compared to those estimated using Fieller's Theorem (Data not shown).[95, 277] The externally-weighted GRS for LDL-C was positively associated with IS risk, with an OR of 1.28 (95% CI: 1.05-1.56), per 1 mmol/L higher genetically-determined LDL-C concentration (**Figure 6.5**). In contrast, in the present study, there was no evidence of any association of the GRS for LDL-C with ICH (OR 1.04; 95% CI: 0.87-1.25).

Figures 6.6 and 6.7 report the proportional relevance of genetically-determined LDL-C for stroke risks by baseline SBP levels. The associations of genetically-determined LDL-C with risk of IS were directionally consistent by levels of baseline SBP, with slightly higher proportional risks observed among individuals with baseline SBP levels below 145 mmHg (OR 1.42; 95% CI: 1.11-1.82), compared to those with baseline SBP of 145 mmHg or above (OR 1.04; 95% CI: 0.74-1.48). The associations of the genetically-determined LDL-C with ICH appeared to differ by levels of SBP ($p_{heterogeneity}$ =0.003), with ORs of 1.16 (95% CI: 0.89-1.51), and 0.86 (95% CI: 0.64-1.15), among individuals with baseline SBP of 145 mmHg below versus above. In addition, the genetic relevance of LDL-C appeared to be greater for non-lacunar stroke than lacunar stroke, but the confidence intervals were compatible with each other (**Figure 6.8**). However, the numbers of cases included in the present study were too small for such subgroup analyses to be informative, and the weak evidence of differences suggested in certain subgroups may reflect the play of chance.

6.3.2.2 High-density lipoprotein cholesterol

In the present study, the single *LIPC* genetic variant was not associated with stroke risks of any type. Of the possible 25 five-SNP based *CETP* haplotypes,

there were five common haplotypes (i.e., with a frequency of 5% or above), with frequencies ranging from 7% for h.CCGTC to 35% for haplotypes h.CCGGT, respectively. Overall, the five haplotypes constituted approximately 91% of all the potential information included in the five *CETP* variants. **Table 6.7** shows differences in mean plasma HDL-C concentrations, and other major vascular risk factors, associated with each addition of the common *CETP* haplotype. In the present study, there was strong evidence of associations of h.AAAGT (positive), h.CCGTC (negative), and h.CAGGT (negative) with plasma HDL-C concentrations, with p values of $2.2X10^{-64}$, $9.0X10^{-13}$, and $1.5X10^{-8}$ respectively. In contrast, there was no evidence of any association of h.CAAGT, and only weak evidence of association of h.CCGGT with plasma HDL-C concentrations. Of the common haplotypes, h.AAAGT tended to be positively, and negatively associated with plasma concentrations of LDL-C, and triglycerides, and not for other common haplotypes. None of the haplotypes were associated with levels of baseline SBP, or BMI. Overall, by combining all the common haplotypes together through inverse-variance weighted method, there was weak evidence of a positive association of HDL-C with each additional common haplotype included here (p=0.02), and no evidence of association with other potential vascular risk factors (**Table 6.7**).

Table 6.8 shows the associations of the five common *CETP* haplotypes with stroke risks. Allowing for the uncertainty in haplotype assignments, none of common haplotypes were associated with risk of either IS or ICH. The F-statistics for each common *CETP* haplotype ranged from 0.1 observed in h.CAAGT to 350 in h.AAAGT for associations with stroke risk. Combining the above risk estimates throughout the five common haplotypes, the ORs, associated with each 0.3

mmol/L higher HDL-C were 1.05 (95% CI: 0.84-1.30) for IS, and 0.92 (95% CI: 0.74-1.13) for ICH, respectively. The genetic relevance of HDL-C determined by the *CETP* haplotypes appeared to differ by stroke types (p=0.02). However, given the modest difference, this may largely reflect the play of chance.

6.3.2.3 Triglycerides

Table 6.4 shows the selected characteristics for the triglyceride-associated genetic variants studied, based on the GLGC study and the 1000 Genome Project. Notably, of the four triglyceride-associated genetic variants, the -1131T>C (rs662799) promoter polymorphism of the apolipoprotein A5 (*APOA5*) gene was three times more common in Chinese populations (26%), compared with European populations (8%). As in the hydrolysis process of the triglyceride-rich lipoproteins by lipoprotein lipase, the reduction in plasma triglyceride concentrations is accompanied by the transfer of cholesterol into HDL participles, resulting in an inverse relationship between plasma concentrations of triglycerides and HDL-C. Consequently, the previously identified triglyceride-associated genetic variants are frequently inversely associated with HDL-C, which was also observed in CKB (**Figure 6.9**). The effect estimates of the four genetic variants on triglycerides were similar to those estimated in the GLGC study, and their effects on HDL-C were also similar.

Figure 6.10 displays the associations of the externally-weighted GRS for triglycerides with baseline plasma concentrations of triglycerides, HDL-C, and LDL-C; levels of BMI, and SBP. There was strong evidence of a positive association of the GRS with triglycerides ($p=8.9X10^{-152}$), and the GRS explained around 4.4% of the total triglyceride variance (F-statistic: 763). There was an

inverse association of the GRS for triglycerides with HDL-C ($p=4.6X10^{-74}$). Each 1 SD higher of the externally-weighted GRS was associated with 0.20 SD higher plasma concentrations of triglycerides, and 0.14 SD lower plasma concentrations of HDL-C. However, there was no evidence of any association of the genetically-determined triglycerides with either stroke pathological types. Each 30% higher genetically-determined triglycerides was associated with an OR of 0.94 (95% CI: 0.81-1.09) for IS, and 0.91 (95% CI: 0.79-1.05) for ICH, respectively. The absence of any significant associations of the genetically-determined triglycerides with either stroke types could reflect the lack of statistical power. Given the observed HRs associated with each SD higher triglycerides for IS (1.05 [95% CI: 1.00-1.10]) and for ICH (0.86 [95% CI: 0.82-0.90]), the statistical power of the present genetic analyses to detect a significant association was only 9% for IS, and 44% for ICH, respectively.

6.4 Discussion

6.4.1 Low-density lipoprotein cholesterol

In the present study, with 5567 incident IS cases and 4911 incident ICH cases, in addition to 9742 common controls, an externally-weighted GRS obtained by combining eight LDL-C genetic variants, was positively associated with IS risk, but not with ICH risk. This LDL-C GRS of eight SNPs explained around 7.2% of the total LDL-C variance in this Chinese population. Importantly, each 1 mmol/L higher genetically-determined LDL-C was associated with an OR of 1.28 (95% CI: 1.05-1.56) for IS, which was consistent with the estimates from observational studies ($p_{heterogeneity}=0.8$). In contrast, there was no significant association of the genetically-determined LDL-C with risk of ICH (OR 1.04; 95% CI: 0.87-1.25),

however, these estimates were still compatible with the findings in the observational studies ($p_{heterogeneity}$ =0.06).

There have been few large studies assessing the genetic relevance of LDL-C for stroke risks, in contrast to the widely existing evidence for the causal relevance of LDL-C for CHD. The METASTROKE study, which was initiated to investigate the genetic associations for IS and its subtypes, is a large collaboration study of 15 IS cohort studies, involving a total of about 12 000 IS cases, and 62 000 controls, all of European ancestry.[278, 279] In the METASTROKE study a GRS of 12 SNPs (reported in an abstract only), weighted by their effects on LDL-C as reported in the GLGC study, was associated with a non-significant 8% higher risk of IS (12 580 cases versus 62 546 controls; OR 1.08; 95% CI: 0.93-1.25 per 1 mmol/L higher), with no evidence of heterogeneity in the effects of the GRS for LDL-C with IS subtypes ($p_{heterogeneity}$=0.9).[130] In contrast, the same LDL-C GRS was associated with a two-fold greater risk of CHD (OR 2.14; 95% CI: 1.67-2.73), among individuals included in the HPS and PROCARDIS studies (8491 cases versus 9275 controls).[130] Comparison of the genetic relevance of LDL-C to CHD risk is currently unavailable in the CKB study.

A most recent METASTROKE publication reported that higher LDL-C conferred by a *PCSK9* genetic variant (R46L) appeared to have significantly weaker ($p_{heterogeneity}$=0.02), if any, associations with risk of IS (OR 0.92; 95% CI: 0.61-1.41 per 1 mmol/L genetically-raised LDL-C) than with risk of CHD (OR 1.69; 95% CI: 1.34-2.13).[280] In another recent report of 3657 presumed IS cases, the association of genetically-determined LDL-C with risk of IS (OR 1.52; 95% CI: 0.90-2.58) was also less extreme than that for risk of CHD (OR 2.24; 95% CI: 1.57-3.19), but

these estimates were statistically compatible ($p_{heterogeneity}=0.2$).[281] The differences between the two studies may be explained by the selection of LDL-C genetic variants, the heterogeneity between study populations, or relatively smaller number of IS cases in the latter report. The loss-of-function genetic variant (rs11591147) used in the recent METASTROKE study publication was rare (effect allele frequency: 1.5%) and might well be subject to differences introduced in populations. The genetic variant used in the METASTROKE was not considered in the latter report, of which, the seven *PCSK9* genetic variants included were relatively common in Western populations (effect allele frequencies ranged from 15% to 84%).[280, 281]

Estimates in the two reports from the METASTROKE study,[130, 280] and Ference's report,[281] were less extreme than what the present CKB study has suggested, but were generally statistically compatible. Indeed, stroke research has frequently been complicated by diagnostic limitations and phenotypic heterogeneity, as well as by an inability to distinguish IS subtypes. However, each of these studies has relatively high rates of brain-imaging use for stroke diagnoses. Notably, the number of IS cases involved in both the CKB study, and Ference's report,[281] were only half those included in the METASTROKE study, and hence, the estimates were probably less robust than those reported in the METASTROKE study.[130, 280] However, the nature of more homogeneous populations could have contributed to the slightly greater risk estimates for IS in the present CKB study.

PCSK9 variants in the METASTROKE study had significantly weaker effects, if any, on risk of IS, than for CHD.[280] The estimates were generally consistent with other two genetic studies, suggesting weaker effects of LDL-C on IS risk than

CHD.[130, 281] By contrast, the FOURIER trial of PCSK9-inhibitor therapy (treated for a mean of around 2.2 years),[282] together with the large CTT meta-analysis of large trials of statin therapy (treatment duration of around 5 years),[15] have demonstrated that lowering LDL-C had similar reductions in risk of both IS and MI (which was also noted by Hopewell and colleagues in their report[280]). The METASTROKE study illustrates the complexities using MR findings of genetic variants on different outcomes to anticipate results in randomised trials. This could possibly be resolved, where available, to compare the effects of the potential genetic variants with a reference standard variants on different outcome, for example, comparing *PCSK9* variants with *HMGCR* variants, the latter of which had relatively clear biological pathways.[281] Such discrepancies also raises concerns about comparisons of the relevance of LDL-C for risk of both IS and IHD with studies conducted in different populations. Compared to the more prominent role of atherosclerosis in CHD, IS has different subtypes, some of which are amenable to LDL-C lowering treatment (such as the large arterial atherosclerotic IS), whereas some are more likely to be affected by non-lipid factors, like hypertension and atrial fibrillation.[33] The comparison of relevance of LDL-C to IS and IHD are frequently made between studies in different populations, or using different datasets, which may well be at different levels of atherosclerotic risk. By contrast, in the large randomised trials, random-allocation has ensured that participants allocated to either treatment, or placebo, are strictly comparable in terms of their atherosclerotic risk for different outcomes.[18]

In contrast to the positive association of the genetically-determined LDL-C with IS, in the present general population-based genetic study, there was no evidence of association of the genetically-determined LDL-C with risk of ICH. However, the

confidence intervals estimated in the present genetic study were statistically compatible with estimates of the observational findings in **Chapter 5** ($p_{heterogeneity}$ =0.06). To date, there were no other studies investigating the association of LDL-C with ICH using MR approach, which perhaps reflected the rarity of ICH in other populations. The absence of a significant association in the present study seems to suggest the inverse relationship in observational studies may be inflated, due to residual confounding or reverse causality from other chronic conditions, even though the present nested case-control study of stroke excluded participants who reported of major vascular events at baseline.

In previous observational studies with relatively limited number of HS cases, it has also been suggested the inverse relationship of LDL-C with risk of HS was confined among individuals with hypertension, although the confidence intervals were substantially overlapping within each other.[108, 111] Such observation has not been supported by the observational associations of LDL-C with risk of ICH included in the thesis (**Chapter 5**). In the present genetic study, the genetically-determined LDL-C appeared to have directionally different associations with risk of ICH, by levels of baseline SBP (of 145 mmHg above versus below), but interpretation of these subgroup findings requires caution, as the number of cases was too few to be informative.

6.4.2 High-density lipoprotein cholesterol

The present study did not find any association of HDL-C determined by either the genetic marker in *LIPC* or the common *CETP* haplotypes (i.e., the haplotype frequency was above 5%). However, because of the small proportion of the variance in HDL-C concentrations explained by the genetic variants in the present

study, there was insufficient statistical power to address the causal relevance of HDL-C determined by either *LIPC* or the common *CETP* haplotypes to stroke risks.

Despite the highly consistent inverse association of HDL-C with risk of occlusive vascular diseases in observational studies conducted in different populations,[13, 14] emerging evidence from genetic studies,[129, 283] and randomised clinical trials,[143, 144, 147] appears to suggest that HDL-C concentrations may not be an effective biomarker for the entire HDL system, which may consist of several subgroups of HDL with different functions.[78] Exposure to elevated HDL-C as estimated by combining genetic variants associated with HDL-C through several molecular pathways was not associated with risk of coronary events.[129] The apparent conflicting results might suggest that it is the pathways that raise HDL-C, rather than HDL-C per se, are protective for cardiovascular diseases.[78]

Previous genetic studies on *CETP* genetic variants have reported conflicting results.[78, 117] *CETP* genetic variants associated with higher plasma HDL-C concentrations have been associated with a lower risk of IS or of coronary events, as has been shown in populations of European,[132, 284, 285] or Asian ancestries,[286] with number of cases ranging from 800 to 27 000. In contrast, one study of about 3500 Japanese-American men with about 1000 cases, *CETP* genetic variants were related with higher HDL-C concentrations, and higher risk of CHD (p=0.008).[287] In contrast, a previous analysis of the *CETP* genetic variants in the CKB study, which were strongly associated with higher HDL-C, lower triglycerides, but were unrelated to LDL-C concentrations, were not significantly associated with any occlusive vascular disease, including IS.[288]

The five SNPs for *CETP* included in the present study differed slightly from those included in the previous CKB analyses, where the previous *CETP* loss-of-function variant rs2303790 was replaced by rs2033254 in the thesis, due to the relatively low LD information provided by the former. In aggregate, in the CKB study population with a relatively low LDL-C (the mean was about 2.3 mmol/L), there was no evidence of any association of a single loss-of-function *CETP* variant (rs2303790),[288] or the *CETP* gene score,[288] or the present common *CETP* haplotypes with risk of either stroke pathological types. Differences in the strength of associations of the above *CETP* genetic variants with concentrations of LDL-C, or other lipid fractions, may partially contribute to the inconsistent results observed in previous studies. In a recent genetic study of *CETP* variants, the estimated protective effects against CHD appeared to be wholly explained by the LDL-C differences conferred by the *CETP* variants.[134]

One study has reported *CETP* variants associated with both higher HDL-C concentrations and higher risk of ICH.[289] This study has performed a case-control candidate-gene association study using GWAS and genotyping data, involving a discovery set of 1149 ICH cases and 1238 controls, and a replication set of 1625 ICH cases and 1845 controls. In this study, the additive GRS of 7 SNPs for HDL-C was associated with higher HDL-C concentrations by 0.07 mmol/L (~0.19 SD) as estimated in the GLGC study, and was associated with higher risk of ICH (OR 1.86; 95% CI: 1.44-2.40) in the discovery set. In CKB, the absence of a significant association of genetically-determined HDL-C with risk of ICH could be due to the small proportion of HDL-C variation explained by the genetic instruments. However, it was still unable to establish a direct causal relationship of HDL-C with risk of ICH, as also noted by Anderson and colleagues in the report,[289] for, the

CETP variants selected could also be correlated with other vascular factors. In addition, in that study, the lead SNP (rs173539) was not captured in the replication set, and thus, the replication was conducted using other four *CETP* variants that were in high LD with the lead SNP. The replication was strongest for variants in slightly lower LD with the lead variant, and slightly higher heterogeneity was suggested between studies included. In comparison with the selection of genetic variants for replication, the GRS for HDL-C was tested in the discovery set, and the effects of the GRS had on ICH were largely driven by the three SNPs having high LD with the lead SNP (R^2 >0.95). Future studies involving large number of ICH cases together with the lead *CETP* genetic variants might be required to further validate such findings.

In comparison, another genetic variant in the *SCARB1* gene, which was significantly associated with higher HDL-C and particularly higher cholesteryl ester content in the large HDL particles, was also associated with significantly higher risks of CHD.[283] The rare *SCARB1* genetic variant was not included in the present analysis of CKB genetic data available for the thesis. The scavenger receptor class BI (SR-BI), encoded by the gene *SCARB1* has been reported to promote the selective uptake of HDL cholesteryl esters.[290, 291] However, it is unclear whether the *SCARB1* gene harbours new pharmaceutical targets, as in humans, the HDL cholesteryl esters are chiefly delivered to the liver via CETP-mediated exchange to Apo B-containing lipoproteins (i.e., the redistribution of cholesterol between lipoproteins) rather than by the direct uptake from HDL by the liver.[290]

6.4.3 Triglycerides

In the present study, a triglyceride GRS involving four genetic variants chiefly affecting the activity of lipoprotein lipase, was positively associated with plasma concentrations of triglycerides ($p=8.9X10^{-152}$), and as expected, inversely correlated with HDL-C ($p=4.6X10^{-74}$). The triglyceride GRS explained around 4.4% of the total variance in triglycerides (F-statistic=763), and 1.5% in HDL-C, respectively. There was no evidence of associations of the triglyceride GRS with risk of either stroke pathological types. Each 30% higher genetically-determined triglycerides was associated with ORs of 0.94 (95% CI: 0.81-1.09) for IS, and 0.91 (95% CI: 0.75-1.05) for ICH, respectively.

Genetic studies have also suggested the causal relevance of genetic variants within *LPL*,[292] *APOA5*,[124, 293] *APOC3*,[294, 295] *ANGPTL4*,[296] and *ANGPTL3* regions,[297, 298] which are correlated with higher plasma triglyceride concentrations, tended to be associated with higher risk of CHD. These studies were predominantly conducted within populations of European ancestry, and the triglyceride-raising genetic variants were also correlated with other lipid fractions, especially HDL-C, with inverse relationships shown in *APOA5*, *APOC3*, and *ANGPTL4*, and positive relationship in *ANGPTL3*. Previous genetic studies assessing the causal relevance of triglycerides were chiefly concerning coronary risk, while only one study examined the genetic relevance of *APOC3* for risk of IS, and ischaemic cerebrovascular events were included in ischaemic vascular events as a composite outcome, with over 70% being attributable to CHD.[295]

The links of *APOA5* or *APOC3* with atherogenesis may probably be the inhibition of the lipolysis activity of lipoprotein lipase, and thus cause higher triglycerides and

lower HDL-C concentrations.[83] In one genetic study of *APOA5* including 73 252 participants (in which, around 65% were of European descent), the genetic variant was strongly associated with triglycerides concentrations, moderately associated with HDL-C, but was not associated with other major vascular risk factors, including LDL-C. Each 0.25 mmol/L genetically-raised triglycerides (i.e., per effect allele) was associated with an 18% (95% CI: 11-26%) higher risk of CHD, which was concordant with, but relatively larger than, the observed 10% excess risk (95% CI: 8-12%) associated with equivalent triglyceride increment, adjusted for sex and age only in observational analyses. The stronger genetic relevance for CHD was suggested probably attributable to the life-long exposure to raised triglycerides, even though long-term usual concentrations of triglycerides being applied in the observational estimates. In a genetic study of 41 671 participants of European (~83%) or African, carriers of four rare protein coding-sequence variants of *APOC3,* had 39% lower plasma triglycerides concentrations, irrespective of the presence of *APOC5* variants. The identified *APOC3* coding variants were associated with a 40% (95% CI: 25-53%) lower risk of CHD, which was compatible with another genetic study involving around 12 000 ischaemic vascular events. Notably, in that study, the *APOC3* variants were also correlated with 22% higher HDL-C, and 16% lower LDL-C.

However, compared to *APOA5* and *APOC3,* the biological function of *ANGPTL4* and *ANGPTL3* in triglyceride metabolism is less clear. A recent genetic study of a genetic variant in *ANGPTL4*, reported that carriers tended to have lower triglycerides (13%), higher HDL-C (7%), and a 19% (95% CI: 8-30%) lower CHD risk, compared to non-carriers. Two recent genetic studies have reported that the loss-of-function variants in *ANGPTL3*, encoding angiopoietin-like 3, tended to

have lower plasma triglyceride concentrations and possibly lower risk of CHD.[297, 298]

The available evidence on triglyceride-related genetic variants appeared to suggest triglyceride-mediated pathways are probably causally associated with CHD risk.[83] The association of triglycerides with atherosclerotic disease has not been as widely studied as those for LDL-C. Hence, although studies involving large numbers of SNPs associated with triglycerides appear to support the causal role of triglyceride-mediated pathways in CHD, interpretation of the results is difficult, as the genetic variants included may possibly increase triglycerides concentrations through various ways.

Several lipid modifying agents, including fibrates, statins, and omega-3 fatty acids can lower plasma triglycerides concentrations, among which, fibrates appear to be the most powerful. However, there is no consensus in current guidelines on treatments of hypertriglyceridemia for prevention of cardiovascular diseases.[220-222] In a summary-level meta-analysis of 18 RCTs, involving around 45 000 cases, treatment with fibrates was associated with reductions in triglycerides by 0.48 mmol/L (95% CI: 0.28-0.69 mmol/L), and LDL-C by 0.36 mmol/L (95% CI: 0.13-0.59 mmol/L), respectively, and a modest increase in HDL-C by 0.05 mmol/L (95% CI: 0.01-0.10 mmol/L).[148] In this meta-analysis, beneficial effects of fibrates were chiefly shown for coronary events (RR 0.87; 95% CI: 0.81-0.93), but not for total stroke events (RR 1.03; 95% CI: 0.91-1.16).

6.4.4 Strengths and limitations of the genetic studies

The present study involving about 5567 IS cases, and 4911 ICH cases is the largest study recruited from a single population addressing this question, and has

thus greatly reduced the potential heterogeneity in the underlying genetic architecture between studies that typically occurs in meta-analysis of multiple studies. In addition, to avoid the potential population stratification, the present analyses were conducted separately by region, and then combined through inverse-variance-weighted fixed-effect methods. Although genetic markers selected in the present study were based on GWAS studies predominantly conducted in Western populations, the allele frequencies and the effect estimates on lipid fractions have shown consistent patterns in the CKB and GLGC studies. In CKB, all lipid assays were measured through direct methods, and thus has reduced the inter-correlation between lipid fractions due to the calculations.

The major limitation in the present study is the lack of statistical power. As lipid related genetic markers included in the present CKB genotyping panels were chiefly selected based on the underlying biological mechanisms, these genetic variants have captured a large proportion of the genetic variation in the lipid measurements, as suggested from the F-statistics. However, the relatively limited number of stroke cases has precluded the ability to reliably assess causality in individual studies, and highlighted the importance of large consortia to discover and validate hypotheses on the causal relevance of major blood lipids for vascular diseases. In addition, in the present genetic study, a GRS approach was adopted to assess the causal relevance of LDL-C and triglycerides for risk of stroke, to minimise weak instrument bias and to maximise statistical power. This may potentially include genetic variants that affect lipids through different (or unknown) pathways, and thus combining all the potential genetic markers together may have complicated the biological interpretation of the instrumental variables. Future studies using a specific gene region and larger numbers of stroke cases may

therefore provide more reliable evidence on these questions, particularly for triglycerides.

6.5 Summary

In the present study, the externally-weighted GRS for LDL-C was positively associated with IS, with risk estimates consistent with those estimated in observational studies, and randomised clinical trials. By contrast, there was no significant association of the GRS for LDL-C with ICH, but the estimates were statistically compatible with findings in both observational studies and clinical trials.

Regarding HDL-C, the present study had insufficient power to address the causal relevance of HDL-C determined by either the *LIPC* variant or the common *CETP* haplotypes constructed in this Chinese population. The triglyceride GRS of genetic variants studied here were chiefly involved in the lipolysis process of triglyceride-rich lipoproteins, which have demonstrated strong evidence of association with higher triglycerides and lower HDL-C concentrations. However, there was no evidence of association of the GRS for triglycerides with either stroke pathological types. Hence, better genetic instruments are needed to assess the causal relevance, if any, of HDL-C and of triglycerides for risk of stroke, observed in this thesis.

6.6 Tables

Table 6.1: Basic information on loci primarily associated with LDL cholesterol selected in the present China Kadoorie Biobank study

SNPID	Gene	Chr.	hg19 position	Effect allele	Other allele	1000Genome		GLGC		
						EAF_EUR	EAF_CHN	Beta	Se	P value
rs11206510	*PCSK9*	1	55496039	T	C	0.846	0.952	0.083	0.005	2.38E-53
rs602633	*SORT1*	1	109821511	G	T	0.792	0.954	0.159	0.004	1.50E-261
rs1713223	*APOB*	2	21271270	C	A	0.500	0.127	0.023	0.004	3.83E-09
rs12916	*HMGCR*	5	74656539	C	T	0.431	0.514	0.073	0.004	7.79E-78
rs579459	*ABO*	9	136154168	C	T	0.215	0.207	0.067	0.005	2.42E-44
rs2000999	*HPR*	16	72108093	A	G	0.185	0.296	0.065	0.005	4.22E-41
rs6511720	*LDLR*	19	11202306	G	T	0.902	0.993	0.221	0.006	3.85E-262
rs7412	*APOE*	19	45412079	C	T	0.934	0.909	0.590	0.010	0

Chr.=Chromosome; EAF=Effect allele frequency; Effect alleles are given as the LDL cholesterol raising allele, as reported in the GLGC data.[92] Effect allele frequencies were estimated from 1000Genome project (Phase 3), among European (EUR) and Chinese (CHN) populations, respectively.

Table 6.2: Linkage disequilibrium statistics of the genetic markers for LDL cholesterol, within the same chromosome*

	rs11206510 and rs602633	rs6511720 and rs7412
D prime	0.0601	1.0
R^2	0.0034	0.0007
P-value	0.2330	0.5815

*The linkage disequilibrium characteristics were calculated from 1000Genome Project (Phase 3), using data of Han Chinese in Beijing (CHB), and Southern Han Chinese (CHS).

Table 6.3: Linkage disequilibrium matrix (R^2) of the *CETP* genetic variants*

SNPID	rs3764261	rs1800775	rs708272	rs9939224	rs2033254	rs2303790
rs3764261	1	0.208	0.253	0.021	0.025	0.045
rs1800775		1	0.663	0.114	0.041	0.002
rs708272			1	0.108	0.071	0.007
rs9939224				1	0.441	0.005
rs2033254					1	0.005
rs2303790						1

* Calculated based on the 1000Genome project (Phase 3) Chinese populations (CHB: Han Chinese in Beijing, China; and CHS: Southern Han Chinese; n = 208).

Table 6.4: Basic information on loci primarily associated with triglycerides selected in the present China Kadoorie Biobank study

SNPID	Gene	Chr.	hg19 position	Triglycerides							HDL cholesterol				
				Effect allele	Other allele	1000Genome		GLGC			Effect allele	Other allele	GLGC		
						EAF_EUR	EAF_CHN	Beta	Se	P value			Beta	Se	P value
rs17145738	*MLXIPL*	7	72982874	C	T	0.883	0.892	0.115	0.005	9.42E-99	T	C	0.041	0.005	4.95E-13
rs12678919	*LPL*	8	19844222	A	G	0.879	0.894	0.170	0.006	1.82E-199	G	A	0.155	0.006	1.38E-149
rs2954029	*TRIB1*	8	126490972	A	T	0.532	0.433	0.076	0.003	1.02E-107	T	A	0.040	0.003	2.67E-29
rs662799	*APOA5 – 1131T>C*	11	116663707	G	A	0.080	0.264	0.247	0.007	1.57E-241	A	G	0.100	0.007	4.16E-37

Chr.=Chromosome; EAF=Effect allele frequency; Effect alleles are given as the LDL cholesterol raising allele, as reported in the GLGC data. Effect allele frequencies were estimated from 1000Genome project (Phase 3),[265] among European (EUR) and Chinese (CHN) populations, respectively.

Table 6.5: Selected baseline characteristics of the study population (n=20 220)

	Ischaemic stroke cases	Intracerebral haemorrhage cases	Controls
Number of participants	5567	4911	9742
Demographic factors			
Age at baseline, Mean (SD), years	54.6 (11.0)	59.0 (10.4)	57.9 (11.0)
Women, %	46.7	52.6	52.3
Urban, %	45.9	22.4	21.5
High school or above, %	14.5	12.1	14.8
Lifestyle factors			
Male ever smoker, %	62.7	60.2	60.5
Female ever smoker, %	3.8	3.8	3.3
Male ever drinker, %	27.4	30.2	29.4
Female ever drinker, %	2.8	2.3	2.2
Physical activity, Mean (SD), MET h/day	17.3 (14.5)	17.4 (12.4)	18.4 (11.6)
Regular consumption of certain foods, %			
Meat or poultry	37.9	35.7	37.9
Fish or other seafood	5.3	3.9	5.0
Egg	24.5	21.5	26.4
Fresh fruit	18.9	16.9	20.8
Dairy products	10.1	7.6	10.6
Physical and blood measurements, Mean (SD)			
SBP, mmHg	145 (35)	153 (28)	135 (21)
DBP, mmHg	83 (19)	87 (14)	78 (11)
BMI, kg/m^2	23.8 (4.5)	23.5 (3.6)	23.2 (3.3)
Random blood glucose, mmol/L	6.7 (4.8)	6.5 (3.4)	6.1 (2.5)
Lipid measurements, mmol/L			
Total cholesterol	4.8 (1.3)	4.6 (1.0)	4.6 (0.9)
LDL cholesterol	2.5 (0.9)	2.3 (0.7)	2.3 (0.7)
HDL cholesterol	1.2 (0.4)	1.2 (0.3)	1.3 (0.3)
Triglycerides	1.7 (1.4)	1.6 (1.0)	1.6 (1.0)
Self-reported conditions at baseline, %			
Diabetes	11.1	9.0	5.5
Hypertension	25.0	26.3	11.4
Poor health	15.2	15.5	9.9

SD=Standard deviation; MET=Metabolic equivalent; SBP=Systolic blood pressure; DBP=Diastolic blood pressure; BMI=Body mass index. Mean (SD) values were directly standardised to the age (at baseline, 10-year range), sex, and study area structure of the entire study population included, unless otherwise states.

Table 6.6: Associations of genetic variants for LDL cholesterol with ischaemic stroke (IS) and intracerebral haemorrhage (ICH) in the CKB study

Marker	Effect allele	EAF	LDL-C, mean (SE)	IS, OR (95% CI)	ICH, OR (95% CI)
rs2000999	A	0.258	0.020 (0.008)	1.07 (1.01-1.13)	1.00 (0.95-1.06)
rs11206510	T	0.944	0.030 (0.015)	1.10 (0.98-1.23)	1.10 (0.99-1.22)
rs12916	C	0.530	0.053 (0.007)	1.02 (0.97-1.07)	0.99 (0.95-1.04)
rs579459	C	0.213	0.053 (0.009)	1.08 (1.02-1.15)	1.02 (0.96-1.08)
rs1713223	C	0.131	0.057 (0.011)	0.99 (0.92-1.07)	0.95 (0.88-1.02)
rs6511720	G	0.991	0.112 (0.042)	1.05 (0.78-1.42)	0.98 (0.74-1.29)
rs602633	G	0.942	0.126 (0.015)	0.97 (0.87-1.09)	1.03 (0.92-1.14)
rs7412	C	0.920	0.463 (0.013)	1.08 (0.98-1.19)	1.00 (0.92-1.10)

EAF=Effect allele frequency; LDL-C=Low-density lipoprotein cholesterol; SE=Standard error; IS=Ischaemic stroke; OR=Odds ratio; CI=Confidence interval; ICH=Intracerebral haemorrhage. Both linear regression and the logistic regression were conducted separately by region, adjusted for sex, age at baseline, and age at baseline squared, and then combined through inverse-variance weighted fixed-effect meta-analysis to yield the overall estimate.

Table 6.7: Associations of common *CETP* haplotypes with major blood lipids, blood pressure, and adiposity measurements

Haplotypes*	Estimated frequency (SE)	Differences in phenotypes				
		HDLC	LDLC	Ln TG	BMI	SBP
h. CCGGT	0.35	-0.008 (0.004)	0.011 (0.008)	0.009 (0.007)	-0.035 (0.041)	-0.215 (0.302)
h. CAAGT	0.24	-0.001 (0.004)	-0.010 (0.009)	-0.010 (0.008)	-0.019 (0.045)	0.466 (0.329)
h. AAAGT	0.15	0.090 (0.005)	0.011 (0.011)	-0.023 (0.010)	0.065 (0.055)	-0.203 (0.406)
h. CAGGT	0.09	-0.034 (0.006)	-0.003 (0.014)	0.012 (0.012)	-0.038 (0.067)	0.279 (0.518)
h. CCGTC	0.07	-0.053 (0.007)	-0.016 (0.017)	0.022 (0.015)	-0.031 (0.081)	-0.492 (0.681)
Overall	*0.91*	*0.005 (0.002)*	*0.005 (0.005)*	*-0.001 (0.004)*	*-0.012 (0.024)*	*0.022 (0.176)*
P value		0.02	0.3	0.9	0.6	0.9

SE= Standard error; HDLC=High-density lipoprotein cholesterol; LDLC=Low-density lipoprotein cholesterol; Ln TG= natural logarithm of triglycerides; BMI=Body mass index; SBP=Systolic blood pressure; *Haplotypes included were ones with inferred frequency of 5% or above. Overall estimates were calculated through inverse-variance weighted average of each individual common haplotype.

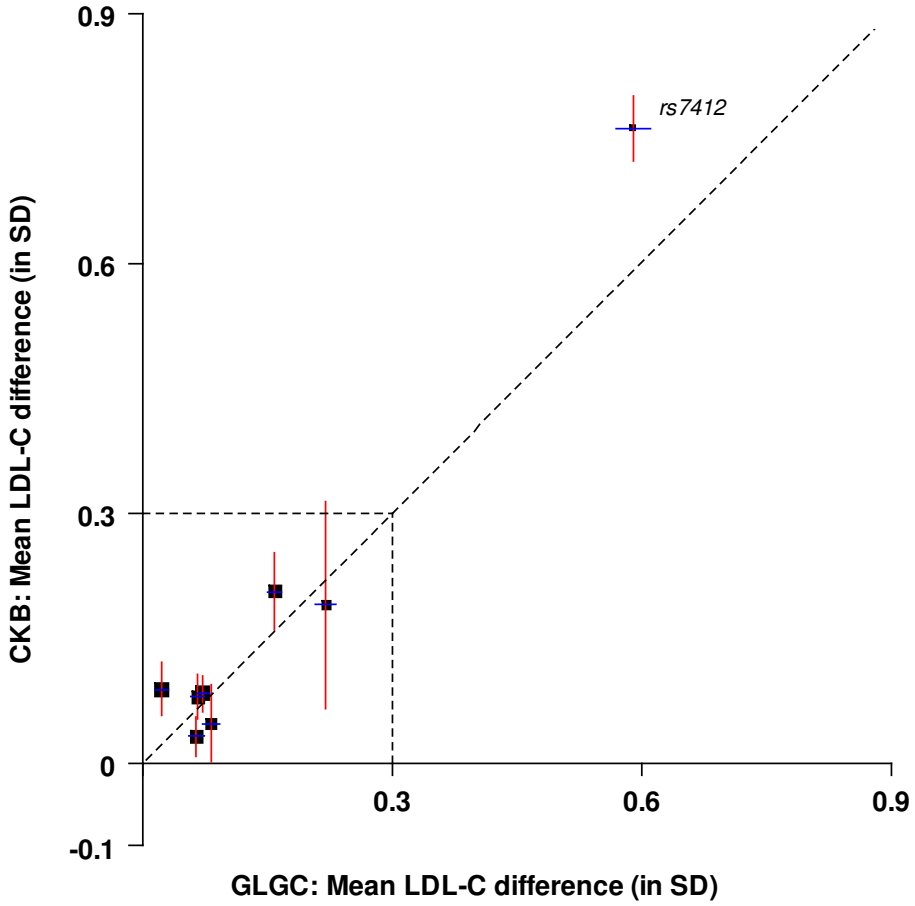
Table 6.8: Associations of common *CETP* haplotypes with HDL cholesterol, ischaemic stroke (IS), and intracerebral haemorrhage (ICH)

Haplotypes	Estimated frequency	F-statistic	HDLC, mean (SE)	IS, OR (95% CI)	ICH, OR (95% CI)
h. CCGGT	0.35	5.5	-0.008 (0.004)	1.02 (0.97-1.08)	1.01 (0.96-1.07)
h. CAAGT	0.24	0.1	-0.001 (0.004)	1.01 (0.95-1.07)	1.03 (0.97- 1.09)
h. AAAGT	0.15	349.6	0.090 (0.005)	1.01 (0.94-1.09)	0.97 (0.90-1.04)
h. CAGGT	0.09	32.1	-0.034 (0.006)	1.07 (0.98-1.17)	1.02 (0.93-1.11)
h. CCGTC	0.07	54.0	-0.053 (0.007)	0.90 (0.81-1.01)	0.98 (0.88-1.10)

HDLC=High-density lipoprotein cholesterol; SE=Standard error; IS=Ischaemic stroke; ICH=Intracerebral haemorrhage; OR=Odds ratio; CI=Confidence interval.

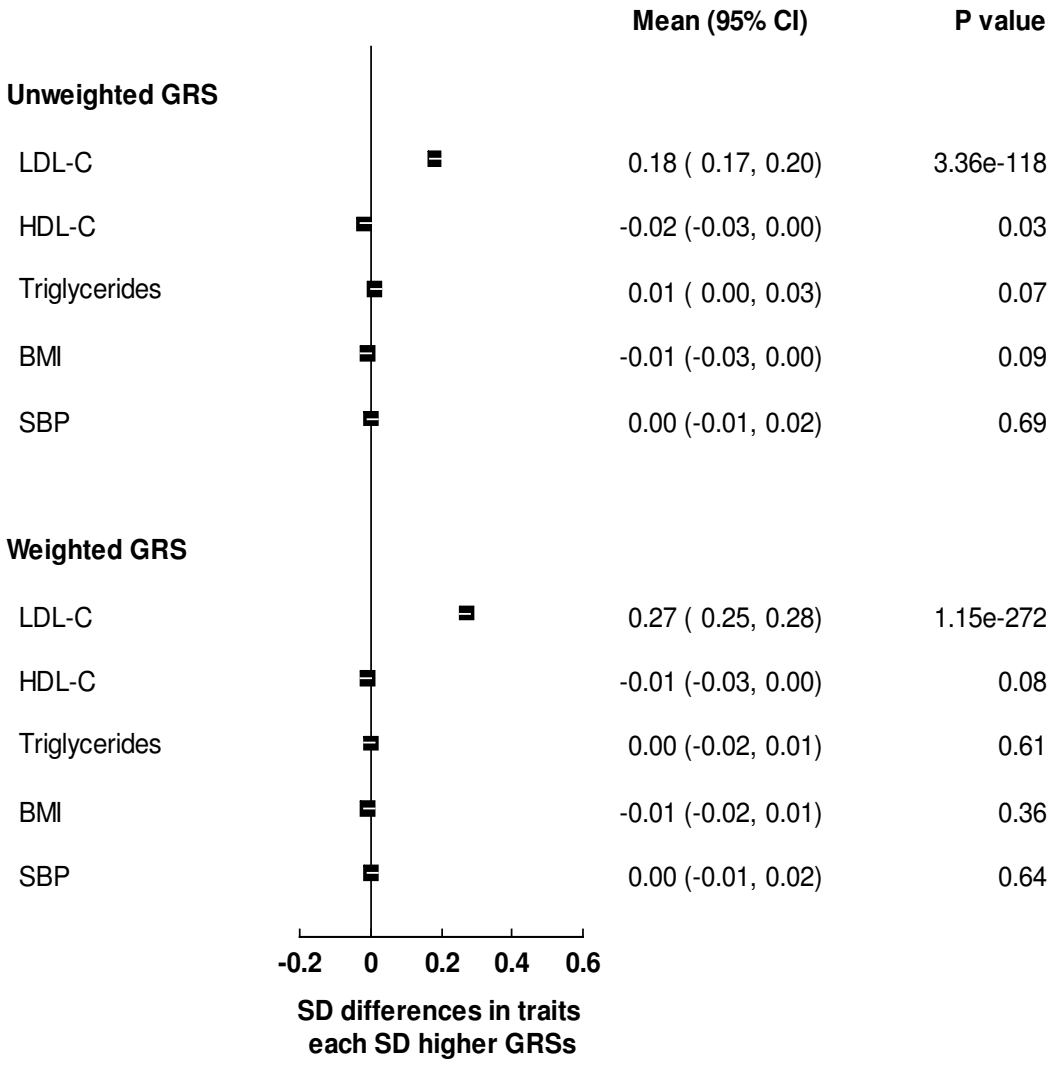
6.7 Figures

Figure 6.3: Comparison of effect sizes of LDL cholesterol related genetic variants on LDL cholesterol concentrations estimated in GLGC and CKB studies



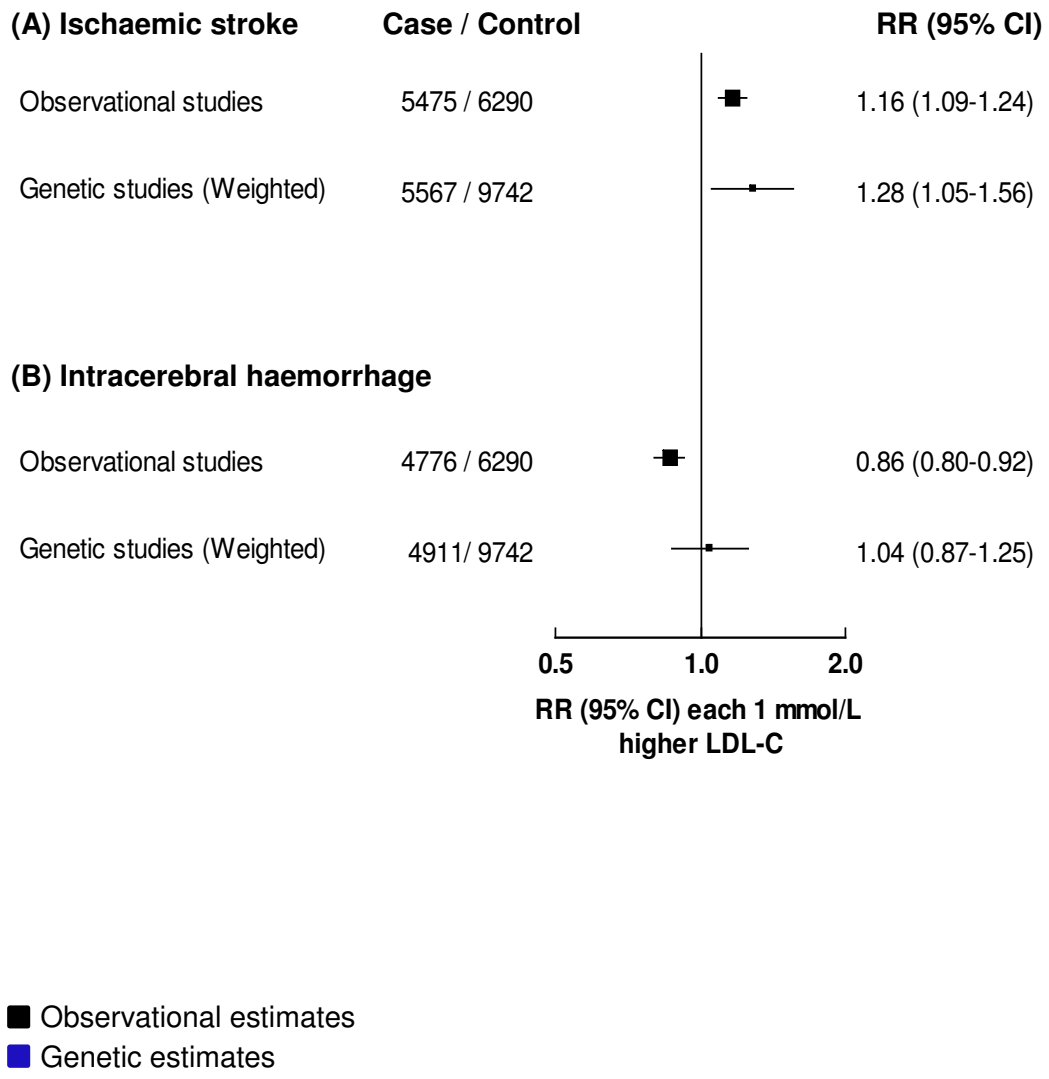
To facilitate comparison between the CKB study and the GLGC study, effect sizes were given with respect to differences of LDL-C in SD (i.e., rank-inverse-normal transformed) for each additional LDL-C raising allele, as reported in the GLGC study. Analyses were performed by region, separately, and adjusted for sex, age at baseline, age at baseline squared. Overall estimates were combined through inverse-variance weighted meta-analysis. Sizes of the squared box were in proportional to the variance estimated in the GLGC study. Horizontal lines represent the 95% confidence intervals of the effect estimates in the GLGC study, with the vertical lines representing the corresponding confidence intervals in the CKB study.

Figure 6.4: Associations of the genetic risk scores (GRSs) for LDL cholesterol with major blood lipids, systolic blood pressure (SBP), and body mass index (BMI)



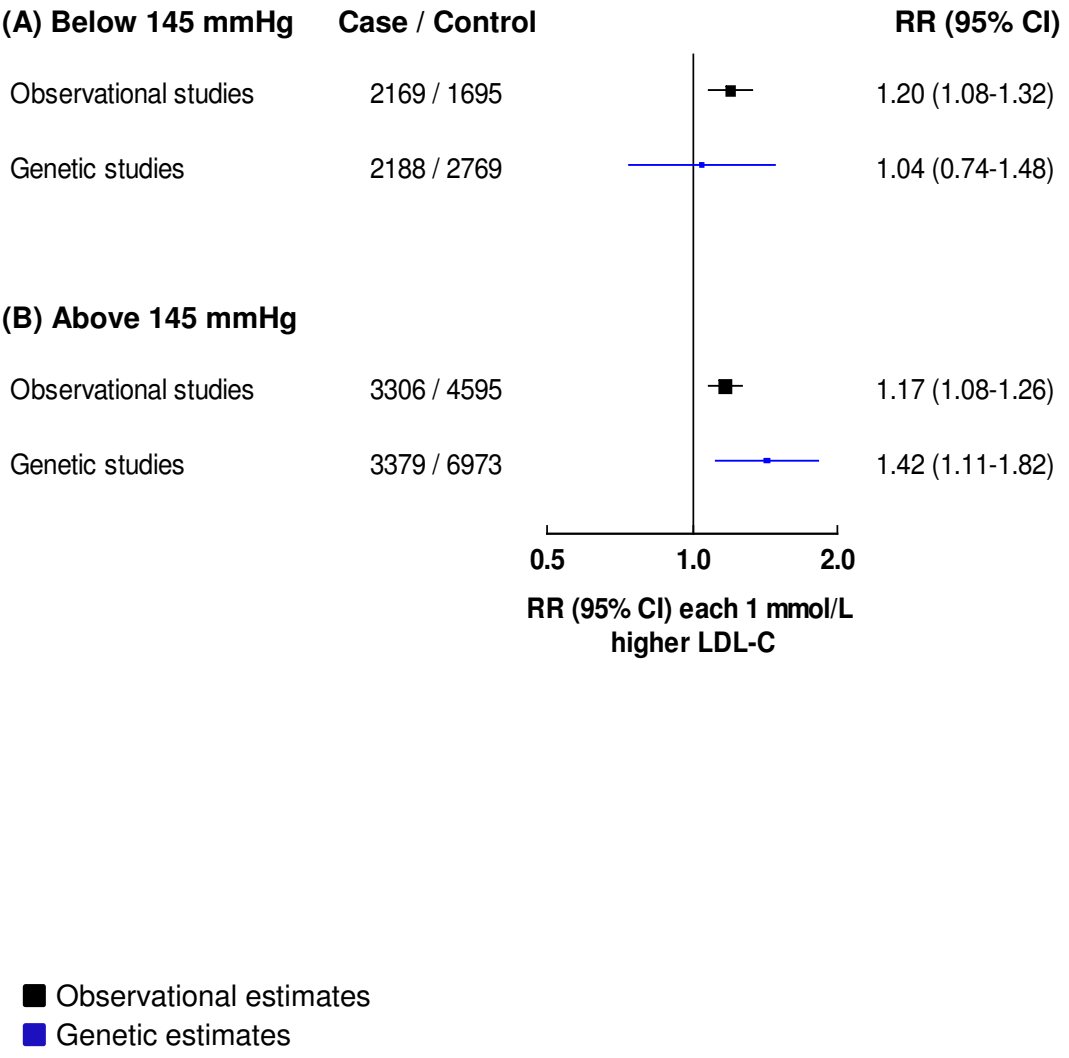
GRS=Genetic risk score; LDL-C=Low-density lipoprotein cholesterol; HDL-C=High-density lipoprotein cholesterol; BMI=Body mass index; SBP=Systolic blood pressure; SD=Standard deviation. The associations were analysed separately, by region, adjusted for sex, age at baseline, and age at baseline squared. Overall estimates were obtained using inverse-variance-weighted meta-analysis of region-specific estimates. The effect sizes were given with respect to SD changes in all traits (rank-inverse-normal transformed) for each SD higher of the GRSs. P values were not corrected for multiple testing, i.e., 10 times.

Figure 6.5: Comparison of associations of LDL cholesterol with (A) Ischaemic stroke, and (B) Intracerebral haemorrhage, in observational study, and genetic study



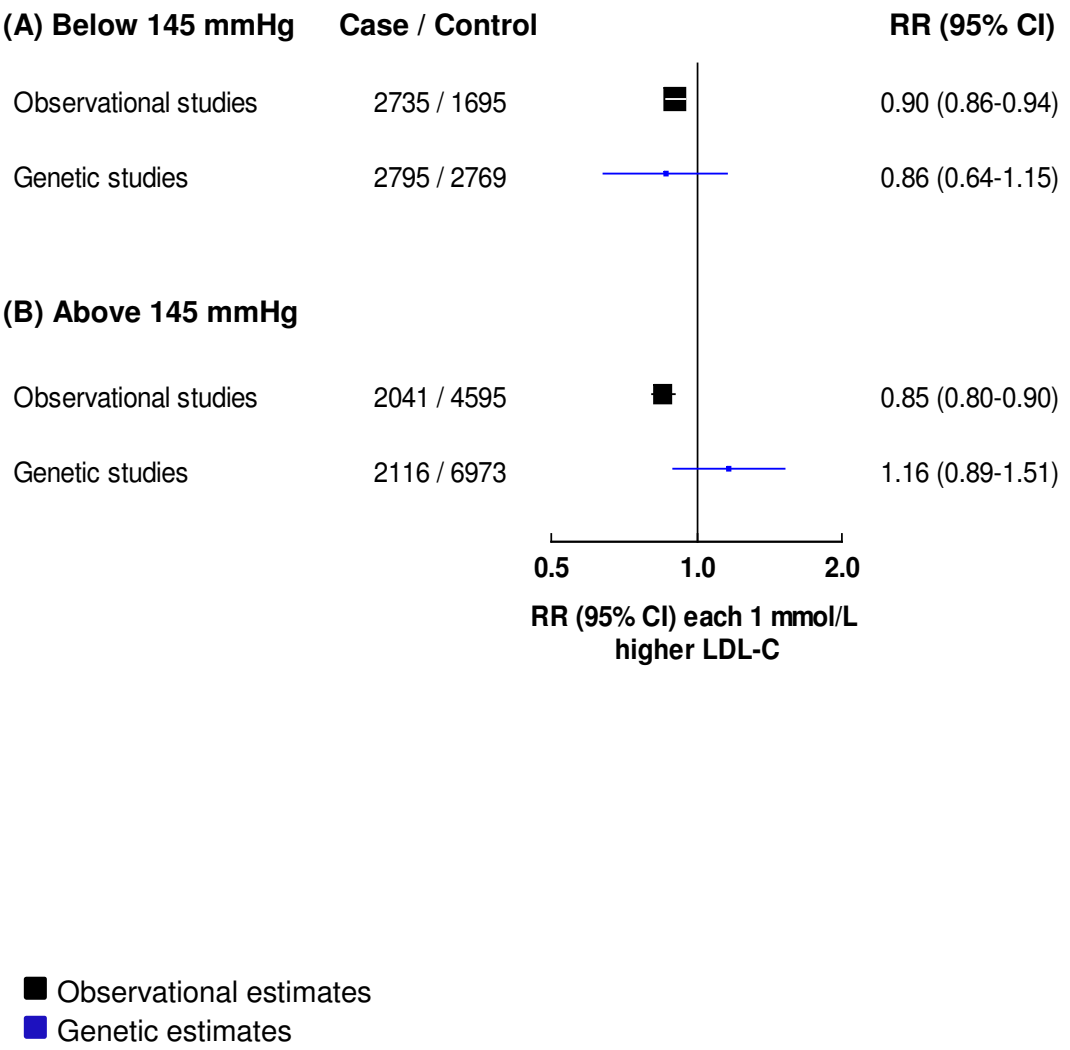
Observational estimates were obtained from Chapter 4-5, respectively. Genetic analyses were using LDL-C GRS as the instrumental variables, in which, all analyses were conducted by region, separately, adjusted for age at baseline, age at baseline squared, and sex. Overall estimates were yielded through inverse-variance-fixed-effect meta-analyses of the region-specific results, and were reported for 1 mmol/L higher LDL-C.

Figure 6.6: Comparison of associations of LDL cholesterol with Ischaemic stroke, by levels of baseline systolic blood pressure, in observational study, and genetic study



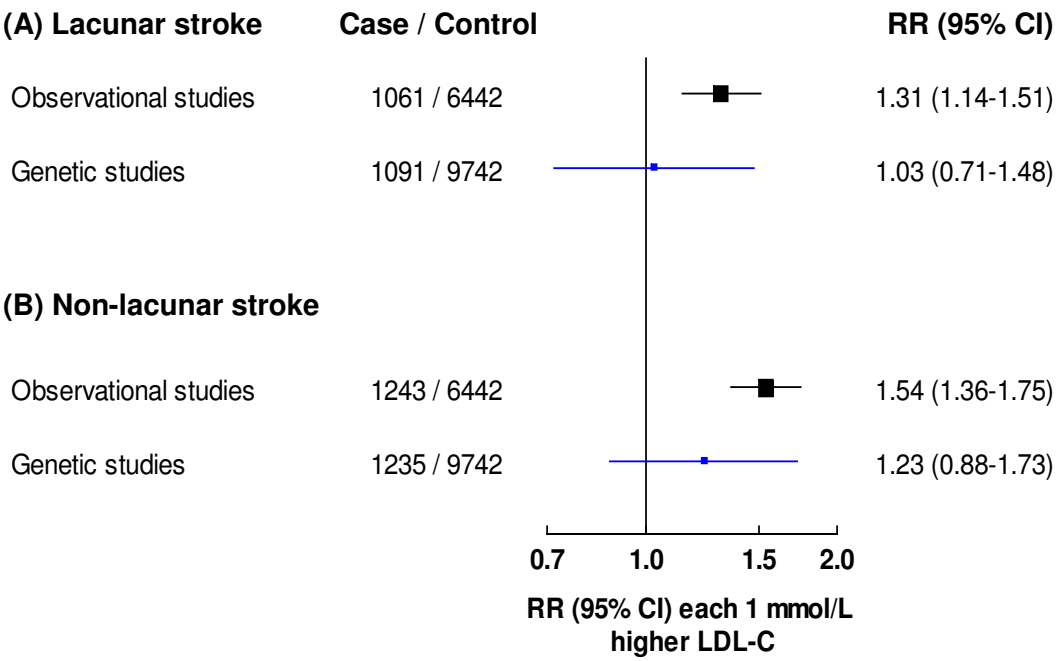
Observational estimates were taken from Chapter 4-5, respectively. Genetic analyses were using LDL-C GRS as the instrumental variables, in which, all analyses were conducted by region, separately, adjusted for age at baseline, age at baseline squared, and sex. Overall estimates were yielded through inverse-variance-fixed-effect meta-analyses of the region-specific results, and were reported for 1 mmol/L higher LDL-C.

Figure 6.7: Comparison of associations of LDL cholesterol with intracerebral haemorrhage, by levels of baseline systolic blood pressure, in observational study, and genetic study



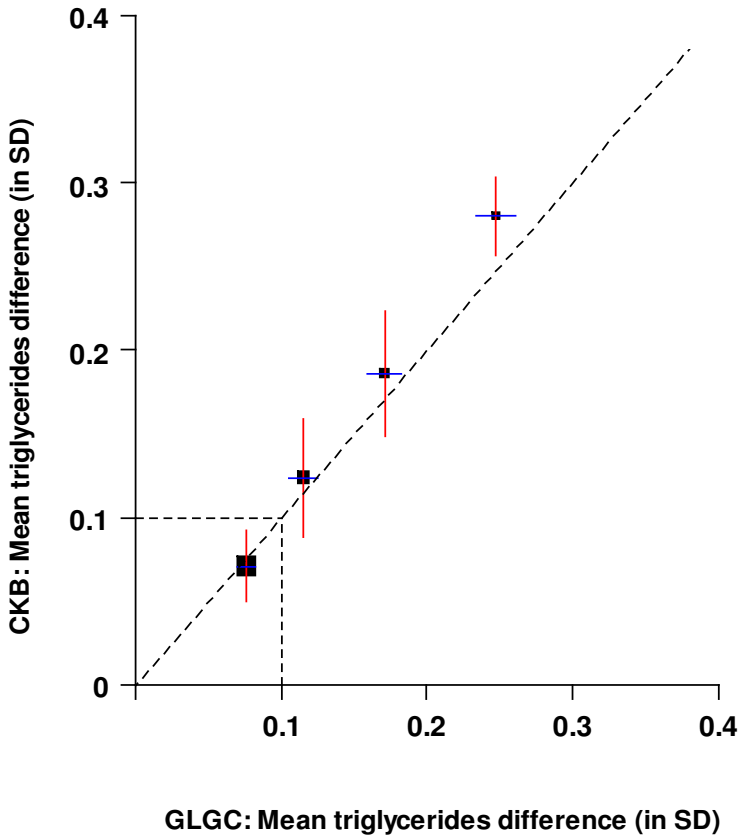
Observational estimates were taken from Chapter 4-5, respectively. Genetic analyses were using LDL-C GRS as the instrumental variables, in which, all analyses were conducted by region, separately, adjusted for age at baseline, age at baseline squared, and sex. Overall estimates were yielded through inverse-variance-fixed-effect meta-analyses of the region-specific results, and were reported for 1 mmol/L higher LDL-C.

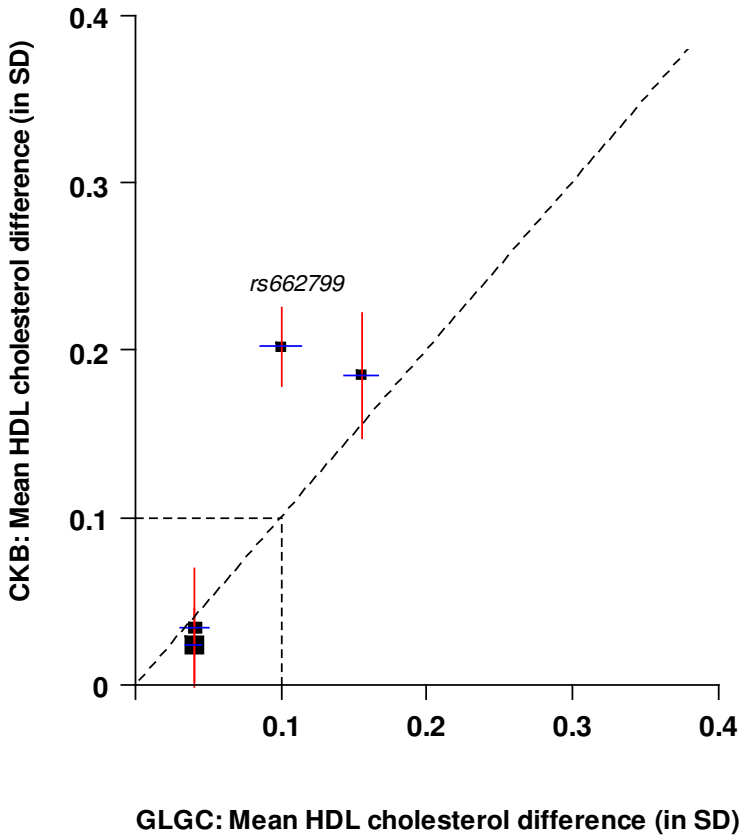
Figure 6.8: Comparison of associations of LDL cholesterol with (A) lacunar stroke, and (B) non-lacunar stroke, in observational study, and genetic study*



*Analyses were conducted among the adjudicated ischaemic stroke events (2304 in observational analysis, and 2326 in genetic analysis). Observational estimates were taken from Chapter 4-5, respectively. Genetic analyses were using LDL-C GRS as the instrumental variables, in which, all analyses were conducted by region, separately, adjusted for age at baseline, age at baseline squared, and sex. Overall estimates were yielded through inverse-variance-fixed-effect meta-analyses of the region-specific results, and were reported for 1 mmol/L higher LDL-C.

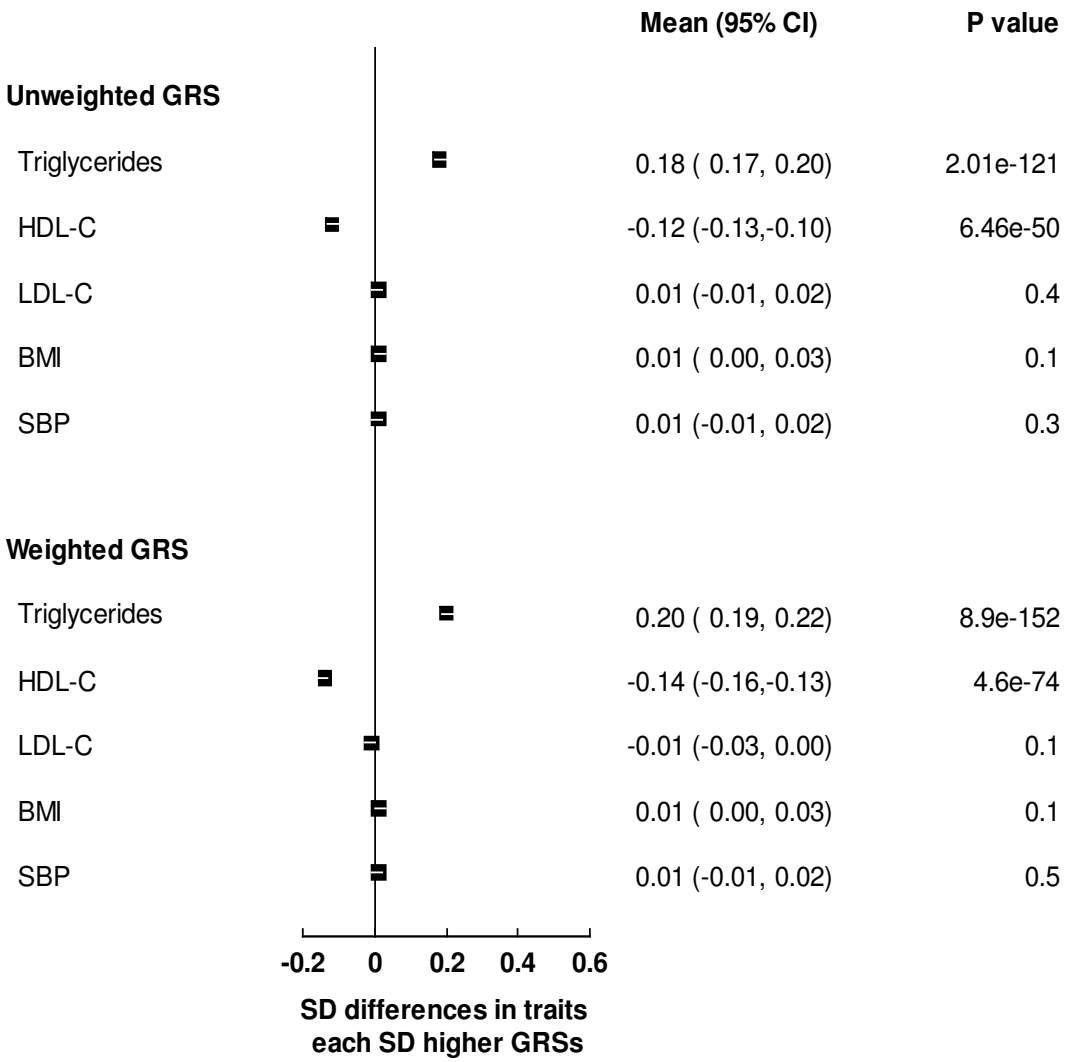
Figure 6.9: Comparison of effect sizes of genetic variants for triglycerides on plasma concentrations of triglycerides, and HDL cholesterol, estimated in GLGC and the CKB studies





To facilitate comparison between the CKB study and the GLGC study, effect sizes were given with respect to differences of Ln triglycerides, and HDL-C in SD (i.e., rank-inverse-normal transformed) for each additional effect allele (lipid-raising), as reported in the GLGC study. The analyses were performed by region, separately, and adjustments were made for sex, age at baseline, age at baseline squared. Overall estimates were combined through inverse-variance weighted meta-analysis. Sizes of the squared box were in proportional to the inverse-variance estimated in the GLGC study. Horizontal lines represent the 95% confidence intervals of the effect estimates in the GLGC study, with the vertical lines representing the corresponding confidence intervals in the CKB study.

Figure 6.10: Associations of the genetic risk scores (GRSs) for triglycerides with major blood lipid fractions, systolic blood pressure (SBP), and body mass index (BMI)



GRS=Genetic risk score; LDL-C=Low-density lipoprotein cholesterol; HDL-C=High-density lipoprotein cholesterol; BMI=Body mass index; SBP=Systolic blood pressure; SD=Standard deviation. The associations were analysed separately, by region, adjusted for sex, age at baseline, and age at baseline squared. Overall estimates were obtained using inverse-variance-weighted meta-analysis of region-specific estimates. The effect sizes were given with respect to SD changes in all traits (rank-inverse-normal transformed) for each SD higher of the GRSs. P values were not corrected for multiple testing, i.e., 10 times.

Chapter 7 Discussion

The final chapter summarises the main findings of this DPhil thesis that have been previously described in **Chapters 4-6**, and displays selected key results in figures and tables, where appropriate. The chapter compares the findings of this thesis with previous studies conducted mainly in Western populations (**Chapter 2**). The chapter includes a commentary on the strengths and limitations of the CKB study on major blood lipids with stroke pathological types. This chapter also provides discussions on implications for public health intervention and clinical practice, and suggests recommendations for future research.

7.1 Summary of main findings

7.1.1 Incidence rates of stroke in Chinese adults

After a median follow-up of 8.8-years (censoring at 1st January 2016) in CKB study population with no prior history of CHD or stroke or TIA at baseline (489 762 individuals; mean age: 52 years; women: 59%), there were 32 869 incident IS cases, and 8270 incident ICH cases recorded, with, approximately, one-fifth (20%) of all incident strokes were attributable to ICH. Incidence rates of IS, and ICH were 761 cases, and 187 cases per 100 000 person-years, respectively, with approximately doubling of incidence rates per decade older age (**Figure 7.1**). In each age group, stroke incidence rates were slightly higher in men than in women, with more extreme sex differences at older ages. After excluding a few individuals aged above 75 years, the incidence rates of IS, and ICH in this Chinese population aged 30-75 years were 658 cases, and 160 cases per 100 000

person-years, respectively. The incidence rates varied between the 10 CKB regions by 10-fold for IS, and 5-fold for ICH.

7.1.2 Prospective associations of major lipids with stroke risks

Baseline characteristics of the study participants

The nested case-control study of stroke pathological types included 5475 incident IS cases, 4776 incident ICH cases and 6290 common controls, among those reporting no current use of statin treatment. Overall, the distribution of most lifestyle characteristics and vascular risk factors at baseline were largely similar between stroke cases and controls. IS cases were slightly younger than controls (54.3 [10.7]) versus 56.7 [11.6] years), but ICH cases were slightly older (58.8 [10.7] years) than controls. The overall mean (SD) plasma concentrations of TC, LDL-C, and HDL-C were, 4.6 (0.9) mmol/L, 2.4 (0.6) mmol/L, 1.2 (0.3) mmol/L, respectively, and the median (IQR) of plasma triglyceride concentrations were 1.6 (1.3) mmol/L.

Regression dilution ratios of the major blood lipids

The RDRs of the major blood lipids were estimated in a subset of the nested case-control study among participants with repeat lipid measurements at the first and second resurveys compared with those measured at the baseline survey. The RDRs of major blood lipid fractions at around 3.7 years' follow-up (the mid-point of follow-up time) after the baseline survey were all about two-thirds (i.e., HDL-C: 0.72; LDL-C: 0.68; triglycerides: 0.62, respectively), consistent with the findings of previous prospective studies for a similar duration of follow-up, albeit estimated

predominantly in Western populations (in the PSC and ERFC meta-analyses of prospective studies[13, 14]).

Prospective associations of major blood lipids with stroke

In this Chinese population, throughout the range of the usual LDL-C concentrations examined (i.e., 1.7-3.2 mmol/L), there was a positive log-linear association of higher usual LDL-C concentrations with higher risk of IS (HR 1.17; 95% CI: 1.10-1.25 per 1 mmol/L higher), and an inverse association with ICH (HR 0.86; 95% CI: 0.80-0.92) (**Figure 7.2**). The proportional relevance of usual LDL-C for risk of stroke pathological types was similar in subgroups classified by age, sex, region, smoking, alcohol, and levels of SBP, at baseline (**Figure 7.3**).

Likewise, in the present study of Chinese adults, higher plasma triglyceride concentrations were also positively associated with higher risk of IS (HR 1.02; 95% CI: 1.00-1.04 per 30% higher triglyceride concentrations), albeit these associations were weaker than those observed for LDL-C. By contrast, plasma triglyceride concentrations were inversely associated with risk of ICH (HR 0.93; 95% CI: 0.91-0.95) (**Figure 7.4**). The inverse association of usual plasma triglyceride concentrations with risk of ICH was unaltered by further adjustment for LDL-C (HRs were attenuated from 0.93 to 0.94).

In this Chinese population with relatively low LDL-C concentrations, usual HDL-C concentrations were inversely associated with risk of IS (HR 0.93; 95% CI: 0.89-0.97 per 0.3 mmol/L higher HDL-C) (**Figure 7.5**). The inverse association of usual HDL-C with risk of incident IS persisted after further adjustment for LDL-C, triglycerides, and other major vascular risk factors. Hence, the present study has confirmed the predictive value of HDL-C for risk of IS in a population with relatively

low LDL-C concentrations. By contrast, there was no association of usual HDL-C concentrations with risk of ICH (HR 1.00; 95% CI: 0.96-1.05).

Table 7.1 compares the strengths of associations of major blood lipids (in SD units of measured lipid values) with IS and ICH, after adjustment for non-lipid vascular risk factors. Overall, both LDL-C and triglycerides were positively associated with risk of IS, and inversely associated with risk of ICH. By contrast, there was an inverse association of HDL-C with risk of IS, but no significant association with risk of ICH. The strength of associations with IS were weaker for triglycerides than for LDL-C, but LDL-C and triglycerides had approximately similar inverse associations with ICH. Importantly, the proportional relevance of LDL-C, triglycerides, and HDL-C for risk of IS differed substantially from those for ICH ($p_{heterogeneity}$: <0.00001, <0.00001, 0.0006, respectively).

7.1.3 Mendelian randomisation studies of major lipids with stroke risks

In the present study, a genetic instrument for LDL-C (i.e., LDL-C GRS) was constructed using eight candidate LDL-C genetic variants, by multiplying the number of LDL-C raising alleles by their effects on plasma LDL-C concentrations as reported in the GLGC study.[92] The effects of these eight SNPs on LDL-C concentrations in this Chinese population were consistent with those estimated in GLGC, where over 94% of the GLGC population were of European ancestry. This externally-weighted GRS was strongly associated with plasma LDL-C concentrations ($p=1.15x10^{-272}$), and explained around 7.2% of the total variance in LDL-C concentrations. By contrast, the GRS for LDL-C was not correlated with HDL-C, triglycerides, or other non-lipid major vascular risk factors (e.g., BMI and SBP). One of the key findings of this thesis was the demonstration that, each 1

mmol/L genetically-raised LDL-C was positively associated with IS (OR 1.28; 95% CI: 1.05-1.56), and not associated with ICH (OR 1.04; 95% CI: 0.87-1.25). Thus, the MR analyses in the CKB study confirmed the causal relevance of LDL-C for IS.

The genetic instrument for HDL-C was constructed by incorporating five SNPs within *CETP* region into haplotypes. Overall, the five common *CETP* haplotypes constituted approximately 91% of all the potential variation conveyed by the five *CETP* variants. There was a positive association of HDL-C with each additional common haplotype studied (p=0.02), and no association with other major vascular risk factors. However, there was no association of the *CETP* haplotypes determined HDL-C concentrations with risk of any stroke pathological type, with ORs of 1.05 (95% CI: 0.84-1.30) for IS, and 0.92 (95% CI: 0.74-1.13) for ICH, for each 0.3 mmol/L higher HDL-C, respectively.

Triglyceride-associated genetic variants included in the present study were chiefly involved in the lipolysis process of triglycerides, and consequently, all variants were also highly correlated with plasma HDL-C concentrations. In the present study, the effect estimates for the four SNPs selected for triglycerides were similar to those estimated in the GLGC study, and their effects on HDL-C concentrations were also similar. The externally-weighted GRS for triglycerides was strongly and positively associated with plasma triglyceride concentrations ($p=8.9x10^{-152}$), and as expected, inversely correlated with HDL-C ($p=4.6x10^{-74}$). However, there was no significant associations of the GRS for triglycerides with risk of stroke pathological types, with ORs of 0.94 (95% CI: 0.81-1.09) for IS, and 0.91 (95% CI: 0.79-1.05), per 30% higher genetically-determined plasma triglyceride concentrations.

Overall, the present study demonstrated that genetically-raised LDL-C was positively associated with risk of IS, but not with ICH. There were no significant associations of either genetically-raised HDL-C or triglycerides with risk of either stroke pathological types.

7.2 Comparison with previous studies

7.2.1 Epidemiology of stroke

The present study of Chinese adults (aged above 30 years) living in mainland China with no prior history of cardiovascular disease at recruitment, demonstrated that the incidence rates of IS and ICH, were 761 cases and 187 cases per 100 000 person-years. The present study has also demonstrated that approximately 20% of the total incident strokes (i.e., IS and ICH cases) were attributable to ICH. These results are important as there have been relatively few contemporary studies of incidence rates of stroke pathological types in China.

Compared with Western populations, Chinese populations have been reported to have high incidence rates of stroke overall or by pathological type, a higher proportion of ICH.[4, 10, 42, 52, 55] However, reliable comparisons in stroke studies have been complicated by a lack of methodologically robust studies of stroke incidence rates between populations, age standardization to different populations, and variable classification systems used for either stroke pathological types or subtypes.[43, 44, 299] The WHO MONICA project, as previously described, has been one of the most internationally comparable studies on stroke.[22, 42] However, in MONICA, most centres only ascertained events (first-ever and recurrent) in individuals younger than 65 years, whereas the present CKB study has recorded a high proportion of stroke cases that occurred in older age groups. Similarly, the

Oxford Vascular Study also had a high proportion of older individuals, where only around 25% of the stroke cases occurred before 65 years, but recorded a lower incidence rates of IS (141 per 100 000 person-years) and ICH (12) in 2005,[203] compared to the present study.

After directly standardised to the WHO standard population,[300] which represented a relatively younger age-structure than the CKB study population, and more generally, the Chinese Population Census (2010),[204] the age-standardised incidence rates of IS and ICH, in CKB, were 304 cases and 77 cases per 100 000 person-years, respectively, which were higher than those in HICs in 2000-2008 (70 and 22 cases, respectively, for IS and ICH).[39] Furthermore, incidence rates of the aggregate of IS and ICH in CKB were consistent with a review on stroke incidence of Chinese populations,[10] which included studies that using criteria for an "ideal" stroke incidence study.[43, 299] This methodologically-rigorous systematic review has found that for ages 45-74 years, incidence rates of overall stroke were higher in Chinese than those in Western populations (205-584 versus 170-335 cases per 100 000 person-years),[10] after directly standardised to the WHO standard population.[300]

There is limited evidence on incidence rates of stroke pathological types in China. In the Sino-MONICA-Beijing project (1984-2004), after age standardised to the Segi standard population, the incidence rates of first-ever IS and ICH ranged from 64 to 164 cases per 100 000 person-years, and from 26 to 57 cases per 100 000 person-years, respectively.[57] In another study (1991-2000), involving three communities of Beijing, Shanghai, and Changsha, the incidence rates of IS were 97, 47, and 78 cases per 100 000 person-years, and the incidence rates of ICH

were 45, 32, and 102 cases per 100 000 person-years, respectively, after directly standardised to the Seig standard population.[51] The corresponding incidence rates in CKB using the same external population were 261 cases per 100 000 person-years for IS, and 66 cases per 100 000 person-years for ICH, both of which were slightly higher than the averaging estimates in previous studies in China.

In China, over the last two decades, there has been a declining trend in the mortality rates of overall stroke,[301] and by type (**Figure 7.6**),[302] which reflects, probably, the improvement in medical care and techniques. However, the secular trends of stroke incidence have been varied, including an increase in IS and a slight decrease in ICH (**Figure 7.7**),[7, 9, 51] which has also been noted in some other low- and middle-income countries.[39] Some of the changes in incidence rates may reflect the increasing use of brain-imaging for diagnosis of stroke, whereas some of the increase in IS incidence rates could be expected as the result of an epidemiological transition occurring in China,[47, 52] as seen for CHD.[303, 304]

Notably, Changsha, a city in Hunan province, has been reported to have particularly high rates of ICH, and higher proportion of strokes attributable to ICH,[49, 51] which was also observed in CKB (region: Hunan). One study in Changsha of around 551 000 participants for about 15 years (from 1986 to 2000) reported that the age-adjusted incidence rates of ICH were about 73 cases per 100 000 person-years, and ICH accounted for 55% of total incident strokes.[49] However, reasons for this high ICH incidence rates in Changsha are still not fully understood.

In 2013-2014, a national door-to-door survey (The NESS-China) was conducted and involved an approximately similar number of participants to the CKB study,

but were younger (mean [SD] age: 46.6 [16.3] years) than the CKB (59.8 [10.4] years). In the NESS-China survey, the crude incidence rates of IS and ICH were 240 cases and 82 cases per 100 000 person-years, with around 80% of the stroke cases being confirmed by brain imaging. The age-specific incidence rates estimated in the CKB study were slightly lower than those in the NESS-China survey (160 and 50, respectively, for IS and ICH, in CKB). Compared with the CKB study, however, there were several limitations in the NESS-China study. The slightly higher incidence rates in the NESS-China study could be due to the fact it was a household-based survey using door-to-door active screening approach, where individuals within one family were more likely to share similar genetic and environmental risk factors for stroke, despite the participants being relatively young. Furthermore, it is unclear if the NESS-China study counted recurrent stroke events during the survey period as first-ever stroke cases, compared to the first-ever-in-a-lifetime stroke cases in the present analysis. Consistent with previous studies in China,[9, 10, 52] and elsewhere,[207, 305] the present study also recorded substantial variation in stroke incidence rates between the 10 study regions within the CKB study population.

It has been suggested that ICH accounts for a larger proportion of total strokes in China (17-55%),[49, 58, 205] compared with those observed in Western populations (9-15%).[10, 39] However, in the present study with around 93% of the stroke events confirmed with brain-imaging information, around 20% of total strokes were attributable to ICH, which was slightly higher than those observed in the Western populations, but were less extreme than those in previous reports in Chinese populations. Hence, methodological issues, such as the low rates of CT/MRI brain

imaging, cannot be excluded as a possible explanation for the differences in the proportions of HS estimated in previous studies.

7.2.2 Global trends in total cholesterol

Data on plasma lipid distributions in Chinese populations are also limited. Compared to Western populations, Chinese populations have been reported to have relatively lower plasma TC concentrations. The 1989 Ecological Survey of 69 Chinese counties, reported mean plasma concentrations of TC, non-HDL-C, HDL-C, and triglycerides, were 3.8 mmol/L, 2.6 mmol/L, 1.2 mmol/L and 1.0 mmol/L, respectively. Around the same time, mean TC concentrations in Shanghai were slightly higher (4.2 mmol/L).[75] The WHO MONICA Project (1980s-1990s) reported the mean TC concentrations in China were around 4.5 mmol/L, compared to that of 5.8 mmol/L in Western populations.[210] In 2002, the Chinese national survey reported the mean concentrations of TC, LDL-C, HDL-C, and triglycerides, were 4.5 mmol/L, 2.4 mmol/L, 1.3 mmol/L, and 1.1 mmol/L.[209] The present study (with baseline plasma samples collected in 2004-2008) was consistent with that in the national survey, with slightly higher mean plasma TC concentrations in urban than in rural areas (5.0 versus 4.5 mmol/L).

Worldwide, between 1980 and 2008, mean TC concentrations declined in HICs including Australasia, North America, and Western Europe, and in Central and Eastern Europe, while TC increased in East and Southeast Asia and Pacific by 0.08 mmol/L per decade.[161, 306] Such increase in TC was also observed in studies in China (from 4.2 to 5.3 mmol/L in 1984-2004).[211, 212] Despite these diverging trends worldwide, in 2008, mean TC concentrations were highest in HICs (mean: 5.2 mmol/L), of which, countries with the highest TC were chiefly in Western

Europe (mean: 5.5 mmol/L).[161, 213-216] Such differences in TC may reflect differences in prevalence of certain cardio-metabolic risk factors (e.g., adiposity and physical inactivity), genetic determinants between countries,[87, 88, 94] or use of lipid treatment. For examples, statins and other lipid-lowering medications have been more widely used in HICs (~2/3 of individuals of high risk for vascular disease),[214, 216, 307, 308] while the use of statins is still low in LMICs (~5%),[309] with even lower estimates observed in CKB (2%).

7.2.3 Low-density lipoprotein cholesterol

7.2.3.1 Observational associations

Several prospective studies, conducted in different populations, involving large number of IS cases, have provided reliable evidence of positive associations of higher plasma LDL-C concentrations (or, non-HDL-C, and TC, approximately) with higher risk of IS.[14, 111, 112, 116, 125] The present study confirmed the positive association of usual LDL-C concentrations with risk of IS, extending the LDL-C range further down to 1.5 mmol/L in general populations, without any suggestion within the range studied of a threshold below which a lower concentration of LDL-C was not associated with a lower risk of IS. The strength of the association observed in the CKB study, with more than 5000 incident IS cases, was consistent with the previously observed estimates in Western populations (i.e., the ERFC,[14] and the AMORIS studies[116]), but was slightly less extreme than the previously reported estimates in Asian populations (i.e., in the APCSC,[109] the KNHS,[111] or the Kailuan studies[112]). Among these studies, both the KNHS and Kailuan studies adjusted for hypertension status as a binary rather than a continuous trait. Given the substantial effect of blood pressure had on the association of LDL-C with IS,

as suggested in the PSC,[13] the ERFC,[14] and the present study, the estimates in the KNHS and Kailuan studies may be subject to residual confounding due to the inadequate adjustment for blood pressure. The APCSC study, with fewer IS cases, reported more extreme estimates than others, but nevertheless, the estimates were statistically compatible with the other studies.

By contrast, in the present Chinese population, with relatively lower LDL-C concentrations than those commonly seen in Western populations, throughout the range of usual LDL-C examined (1.7-3.2 mmol/L), LDL-C was inversely associated with risk of ICH (HR 0.86; 95% CI: 0.80-0.92 per 1 mmol/L higher). The inverse association of LDL-C with risk of ICH had been suggested in previous population studies, particularly were restricted to a specific type of participants (e.g., individuals with hypertension), with relatively small number of HS cases.[13, 108, 111, 125, 247] There have been no large prospective epidemiological studies investigating the relevance of LDL-C for risk of ICH in Chinese populations, where the rates of ICH are high. The present observational study of around 5000 first incident ICH cases demonstrated strong inverse associations of LDL-C with risk of ICH. The estimates in the present study of Chinese populations were consistent with, although were slightly more extreme than, the large KNHS study of 3947 HS cases identified through linkage to routinely collected data (HR 0.91; 95% CI: 0.87-0.95 per 1 mmol/L higher TC). The difference might be due to the fact that the KNHS study was unable to correct for regression dilution bias, and the potential “surrogate dilution bias” (using TC rather than LDL-C), and unable to adjust for blood pressure adequately. More extreme estimates of the proportional risks were observed in other studies with relatively small numbers of HS cases,[109,

[125] but the confidence intervals were largely compatible with those in the present study.

It has also been suggested, in previous observational studies, that the excess proportional risks of HS associated with lower LDL-C were confined to individuals who have hypertension,[13, 108] albeit the confidence intervals were largely overlapping across levels of blood pressure, and the information on mortality from HS in these studies was obtained chiefly from death certificates. In the present study, which has included a larger number of CT/MRI confirmed ICH cases than any of the previous studies, there was, however, no evidence of heterogeneity of the associations of LDL-C with risk of ICH by levels of baseline SBP (145 mmHg below versus above), although failure to detect any such difference may reflect insufficient statistical power. In previous studies, the presence of significant association among individuals with hypertension could reflect the underlying rates of HS were higher among individuals with hypertension, who were more likely to develop HS. Indeed, the lack of reliable evidence on the interaction between blood pressure and blood cholesterol on associations with ICH suggests such interaction is likely to be modest, if any, compared to their individual effects.

7.2.3.2 Mendelian randomisation evidence

The MR study of externally-weighted GRS for LDL-C confirmed the causal relevance of LDL-C for risk of IS in a Chinese population, with an OR of 1.28 (95% CI: 1.05-1.56) associated with each 1 mmol/L genetically-raised LDL-C. There was no significant association of genetically-determined LDL-C with ICH, however, the estimate appeared to be compatible with the observational estimates ($p_{heterogeneity}$ =0.06).

Traditionally, stroke research has been constrained by inability to reliably distinguish IS from ICH, by phenotypic heterogeneity in subtypes of IS. Consequently, there have been only a few large studies assessing the relevance of genetic variants for LDL-C for risk of stroke, in contrast to more reliable evidence on the causal relevance of LDL-C for risk of CHD. In the large METASTROKE study in which more than 95% of the recorded stroke were ascertained with brain imagining findings,[278] LDL-C, as determined by an externally-weighted GRS of 12 SNPs, was associated with a non-significant 8% higher risk of IS (12 580 cases versus 62 546 controls; OR 1.08; 95% CI: 0.93-1.25), for each 1 mmol/L higher, and there was no evidence of heterogeneity in the effects of the GRS for LDL-C with IS subtypes ($p_{heterogeneity}$ =0.9). In contrast, the same LDL-C GRS was associated with a 2-fold higher risk of CHD (OR 2.14; 95% CI: 1.67-2.73), among individuals included in the HPS and PROCARDIS studies (8491 cases versus 9275 controls).[130] A more recent METASTROKE publication reported that higher LDL-C concentrations conferred by a *PCSK9* genetic variant (R46L) appeared to have significantly weaker associations ($p_{heterogeneity}$ =0.02) with risk of IS (OR 0.92; 95% CI: 0.61-1.41 per 1 mmol/L genetically-raised LDL-C) than with risk of CHD (OR 1.69; 95% CI: 1.34-2.13).

In another recent study of 3657 presumed IS cases (see Ference's report), the association of genetically-determined LDL-C with risk of IS (OR 1.52; 95% CI: 0.90-2.58) was less extreme, but compatible ($p_{heterogeneity}$ =0.2), with the estimates for risk of CHD (OR 2.24; 95% CI: 1.57-3.19).[281] The differences between the METASTROKE and Ference's reports may be explained by the selection of LDL-C genetic variants, and the heterogeneity between study populations. The loss-of-function genetic variant (rs11591147) used in the METASTROKE study

was rare (effect allele frequency: 1.5%) and might well be subject to differences introduced in different populations, compared to the relatively more common *PCSK9* variants considered in Ference's report (effect allele frequencies ranged from 15% to 84%).[280, 281]

The results of the two studies from the METASTROKE Consortium meta-analysis of GWAS study of IS,[130, 280] and Ference's report,[281] were less extreme than those observed in the present CKB study, but were still statistically compatible with each other. Each of these studies had relatively high rates of brain-imaging for stroke diagnoses, but the numbers of IS cases included in both the present study, and Ference's report,[281] were only half those included in METASTROKE study.[130, 280] However, the more homogeneous populations studied may have contributed to the slightly stronger estimates, particularly in CKB.

Most observational studies,[13, 14] and genetic studies.[130, 280, 281] have demonstrated that the relevance of LDL-C for IS was substantially weaker than that for IHD. In contrast, large-scale randomised trials of lipid lowering therapy with statins,[15] or evolocumab (a monoclonal antibody that inhibits proprotein convertase subtilisin-kexin type 9 [PCSK9]),[282] have reported similar risk reductions for both IS and IHD for each 1 mmol/L lower LDL-C. Particularly, in METASTROKE study, the *PCSK9* variants had significantly weaker effects on risk of IS than for CHD,[280] whereas in the FOURIER trial of evolocumab (PCSK9 inhibitor), lowering LDL-C was associated with comparable risk reductions of both IS and IHD,[282] as also noted by Hopewell and colleagues in their report.[280] The discrepant results of genetic studies and the randomised trials illustrate the complexities of using MR approaches to anticipate results of randomised trials on different outcomes. Such

discrepancies could be resolved, where available, by comparison of the effects of the genetic variants with positive controls, such as, comparing *PCSK9* variants with *HMGCR* variants.[281, 310]

Such discrepancies also highlight the complexities of comparing the relevance of LDL-C for risk of IS and IHD between studies. Compared to the more clearly defined role of atherosclerosis in CHD, IS has very different subtypes, some of which are more amenable to LDL-C lowering treatments than others, which are more likely to be driven by non-lipid factors, like hypertension and atrial fibrillation.[33] Comparisons of the proportional relevance of LDL-C for risk IS and IHD are frequently made between studies in different populations, or using different datasets, where the participants have different levels of underlying atherosclerotic risk, and consequently, have different proportions of IS subtypes. In the large randomised trials of LDL-C lowering therapies on clinical vascular outcomes, participants are more likely to be at high-risk for atherosclerotic events. Furthermore, the random-allocation in clinical trials has made the participants in either treatment, or placebo groups, differ from each other only by the play of chance with respect to their absolute atherosclerotic risks of having IS and IHD.[18] As observed in the large statin trials, the relevance of LDL-C lowering for risk of presumed IS were similar to, but slightly lower than, that for coronary events, and the difference in the proportional relevance could reflect 10-20% IS resulting from non-atherosclerotic aetiology.

In the present population-based genetic study, there was no evidence of an association between the genetically-determined LDL-C and the risk of ICH. Nevertheless, the confidence intervals estimated in the present genetic study were

comparable with the estimates from the observational studies ($p_{heterogeneity}$ =0.06). To date, there were no available large studies reporting the associations of genetically-determined LDL-C with risk of HS, probably reflecting the rarity of HS in other populations. The absence of significant associations in the present study appears to suggest the inverse association of LDL-C with ICH in observational studies could be inflated, due to residual confounding or reverse causality from other chronic diseases, even though the present nested case-control study of stroke excluded participants who reported history of major vascular disease at baseline.

7.2.3.3 Comparison with randomised clinical trial evidence

The results of large-scale randomised trials of LDL-C lowering treatment for prevention of cardiovascular disease were reviewed to facilitate the comparison of estimates from different evidence sources. This thesis reviewed unpublished results from the CTT collaboration, which is an IPD meta-analysis of trials involving at least 1000 participants and at least 2 years' treatment duration of statin versus control (22 trials; 134 537 individuals; mean baseline LDL-C 3.70 [SD 0.7] mmol/L; mean difference at 1 year 1.08 mmol/L; median follow-up 4.8 years), or more versus less intensive statin regimens (5 trials; 39,612 individuals; mean baseline LDL-C 2.53 [SD 0.6] mmol/L; mean difference at 1 year 0.51 mmol/L; median follow-up 5.1 years).[15 16] Individual data were unavailable (before 2017) for only two eligible trials involving 6331 high-risk individuals with previous vascular disease (SPARCL,[141] and GREACE[311]). Results for effects on stroke pathological types are not publically available for the GREACE study, and the IPD data from the SPARCL study has recently been combined to the CTT study. In the

available CTT results of 27 trials,[16] allocation to statins was associated with a 15% (95% CI: 11-20%) lower risk of total stroke associated for each 1 mmol/L lower LDL-C, including of a definite reduction in risk of IS (RR 0.80; 95% CI: 0.74-0.85), and a non-significant increased risk of HS (RR 1.15; 95% CI: 0.97-1.38). Among the HS events, one-quarter were from the Heart Protection Study (HPS), where the participants had been recruited among individuals with prior history of CHD, cerebrovascular disease, other occlusive disease of non-coronary arteries, or diabetes, or were men aged 65 years or above.[77] In the randomised SPARCL trial among 4731 individuals with prior history of cerebrovascular disease, allocation to atorvastatin 80 mg daily reduced the risk of IS (218 [9.2%] versus 274 [11.6%]; p=0.008), but increased the risk of HS (55 [2.3%] versus 33 [1.4%]; p=0.02).[141] When the SPARCL results were combined with the results of the CTT meta-analysis, there was a 21% (95% CI: 5-14%, p=0.01) higher risk of HS for each 1 mmol/L lowering of LDL-C.[15]

Both the CTT and the SPARCL trials compared the effects on stroke of lowering LDL-C with statin treatment, but the recent FOURIER trial assessed the effect of LDL-C lowering with evolocumab (PCSK9 inhibitor). The FOURIER trial, involved 27 564 participants (including over 1000 individuals recruited from China) and demonstrated that further reduction in LDL-C yielded further reductions in the risk of having cardiovascular events.[282] Compared to the statin trials included in the CTT collaboration, the median duration of treatment in the FOURIER trial was much shorter (2.2 years), but the proportional reductions in risk were consistent with those in the CTT study at around 2-years of follow-up (**Figure 7.6**). Hence, in the thesis, the results from the FOURIER trial were subsequently combined with both the CTT and SPARCL studies (**Figure 7.7**). Overall, there was no statistically

significant heterogeneity regarding the estimates of stroke risks associated with LDL-C lowering in all three studies ($p_{heterogeneity}$ =0.7 for IS, and $p_{heterogeneity}$ =0.4 for HS, respectively). The summary estimates from all the available randomised trials indicated that for each 1 mmol/L LDL-C lowering, irrespective of which treatment was used (i.e., statins or evolocumab), was associated with a 19% (95% CI: 14-23%; p <0.00001) lower risk of IS, and a 19% (95% CI: 4-37%; p=0.01) higher risk of HS.

Overall, the present thesis demonstrated strong positive associations of LDL-C with IS, with highly consistent estimates between observational studies, genetic studies, and clinical trials (**Figure 7.8**). The strength of the inverse associations of LDL-C with ICH was also statistically compatible between the observational studies, genetic studies, and randomised trials. The meta-analysis of large-scale randomised trials provided conclusive evidence that lowering LDL-C was causally associated with a lower risk in IS, and with a higher risk of HS. The present nested case-control studies of stroke pathological types had the largest number of CT/MRI confirmed stroke cases, and it demonstrated consistent findings with those suggested in previous observational studies. Such observational findings have provided complementary evidence that the relevance of LDL-C for risk of IS, and ICH, differ qualitatively from each other in the general population. The present genetic study used externally-weighted GRS for LDL-C by combining eight candidate genetic variants for LDL-C, although had fewer cases than those included in previous meta-analyses (such as the METASTROKE study), confirmed the causal relevance of LDL-C for IS, but not for ICH. Compared to the findings from observational studies, and randomised trials, the less consistent findings from genetic studies of lipids with stroke pathological types highlighted the

importance of large meta-analyses for discovery and confirmation of certain research questions. The excess risk of HS associated with LDL-C lowering is probably causal, but the mechanism is not fully understood. The onset of HS may involve rupture of micro-aneurysms resulting from arterionecrosis in small intracerebral arteries, and the available evidence suggests that very low serum cholesterol concentrations may weaken the endothelium of intracerebral arteries, resulting in bleeding, accelerated by the presence of hypertension.[253, 254]

7.2.4 High-density lipoprotein cholesterol

The present observational study provided reliable evidence that the proportional relevance of HDL-C for stroke risk differs by stroke pathological types ($p_{heterogeneity}$ =0.0003). Among this Chinese population, with comparatively lower LDL-C concentrations than Western population, HDL-C was inversely associated with risk of IS. The inverse association of usual HDL-C concentrations with risk of IS remained significant after further adjustment for LDL-C, triglycerides, and other vascular risk factors, and thus has confirmed the predictive value of HDL-C for IS risk in this Chinese population, consistent with the previous estimates from Western populations.[14] By contrast, there was no evidence of any association between HDL-C concentrations and risk of incident ICH.

Despite the highly consistent inverse associations of HDL-C with risk of occlusive vascular events throughout different populations in observational studies,[13, 14] emerging evidence from genetic studies and randomised clinical trials tends to suggest increasing plasm HDL-C concentrations using certain biological mechanisms is not associated with a lower risk of major occlusive vascular events.[78, 129] Previous genetic studies of *CETP* genetic variants have reported

conflicting results for vascular risk. In some studies, *CETP* genetic variants were associated with higher HDL-C concentrations and lower risk of IS or coronary events, in populations of European,[132, 284, 285] or Asian ancestries,[134, 286] or mixed,[134] whereas one study reported *CETP* genetic variants were associated with higher HDL-C concentration, and higher risk of CHD (p =0.008).[287] However, in the CKB population, there was no evidence of any association of the single loss-of-function *CETP* variant (rs2303790),[288] or the *CETP* gene score,[288] or the present common *CETP* haplotypes with stroke risks. The apparent discrepant results observed in previous studies of *CETP* variants for occlusive vascular events might be accounted by the differences in the associations of the above *CETP* genetic variants with LDL-C concentrations,[134] or other lipid fractions (e.g., lipoprotein [a]). One study has found that the *CETP* genetic variants were associated with higher HDL-C concentrations, and higher risk of ICH.[289] Such findings require further validation with studies involving large number of ICH cases, as the potential heterogeneity effect indicated in that replication study, and high LD between the genetic variants included for HDL-C GRS.

By contrast, another genetic variant within the *SCARB1* gene, which is significantly associated with higher HDL-C concentrations and higher cholesteryl ester content in large HDL particles), was associated with higher risk of CHD.[283] It is unclear, however, whether the *SCARB1* gene harbours novel pharmaceutical targets, as in humans, the HDL cholesteryl esters are chiefly delivered to the liver via CETP-mediated exchange to Apo B-containing lipoproteins rather than by the direct uptake from HDL by the liver. The apparent conflicting findings between genetic studies involving individual genes,[283, 285] or multiple genes,[129] have

prompted the questions about the interpretation of the validity of conventional MR studies, involving biomarkers whose functions are not fully understood.

Randomised trials of raising HDL-C with niacin have failed to demonstrate any beneficial effects on reducing cardiovascular disease, including stroke of any type, alone, or in combination with other vascular disease outcomes used in such trials. In the CDP study,[142] which was conducted in the pre-statin era, the benefits of niacin treatment versus placebo in reducing the risk of vascular events were consistent with the LDL-C reduction observed.[15] Two further trials investigated the effects of the addition of niacin to statin. However, both the AIM-High and the large HPS2-THRIVE trials were unable to demonstrate any beneficial effects of niacin when added to statin treatments,[143, 144] but also reported an excess risk of serious adverse effects associated with niacin treatment. In the AIM-HIGH trial, the predicted 8% lower event rate was only one third the estimated 25% reduction on which the power calculations were initially based. Hence, the AIM-HIGH trial was unable to reliably test the hypothesis that raising HDL-C with niacin reduces cardiovascular risk in statin-treated patients.

CETP inhibition affords another pharmacological approach to test the effect of raising HDL-C concentrations on risk of major vascular disease. The HPS3/TIMI55-REVEAL trial of 30 449 men and women with a prior history of atherosclerotic vascular disease, demonstrated that the addition of the anacetrapib (CETP inhibitor) to intensive statin therapy for around four years resulted in a significant 9% (95% CI: 3-15%) lower risk of major coronary events (but had no significant effects on presumed IS), without any of previously hypothesized adverse effects during the follow-up period.[145] Allocation to

anacetrapib treatment, tested in the REVEAL study, was associated with 100% higher HDL-C concentrations, and 18% lower non-HDL-C concentrations. Consequently, the observed reductions in risk of vascular disease in the REVEAL trial was largely consistent with the estimated risk reduction for an equivalent difference in non-HDL-C lowering achieved with statin treatment.[17] This large trial casts doubt on differences in HDL-C, or other actions of anacetrapib played in modifying the risk of coronary events. Results from the REVEAL study also suggest that CETP inhibitors are chiefly involved in the redistribution of cholesterol between circulating lipoproteins, in contrast with the overall cholesterol reduction achieved by statins, ezetimibe, or evolocumab.

The beneficial effects with treatment of anacetrapib in the REVEAL trial appeared to differ from results of previous CETP inhibitor trials, although each of these trials had their own limitations. The ILLUMINATE trial of torcetrapib was stopped prematurely due the off-target drug effects (significantly elevated SBP and higher plasma aldosterone concentrations). Another two trials (the Dal-OUTCOME and ACCELERATE trials) were also stopped early (after approximately two years) for futility.[146, 147] Importantly, the dalcetrapib tested in the Dal-OUTCOME trial had relatively weak effects on lipid profiles, and had no effects on LDL-C. The evacetrapib tested in the ACCELERATE trial appeared to have similar potency on both HDL-C and non-HDLC to anacetrapib, increasing HDL-C by as much as 130% and decreasing LDL-C by nearly 35%. However, the ACCELARATE trial of evacetrapib treatment involved only around one-third of the number of participants included in the REVEAL trial, and was followed-up for only 2 years. Overall, the available evidence appears to cast doubt on the causal role of HDL-C as a risk factor for stroke or CHD. It is possible that HDL-C acts as a biomarker reflecting a

combination of metabolic and inflammatory processes, in addition to the reverse cholesterol transport system.

7.2.5 Triglycerides

In this Chinese population, plasma triglyceride concentrations were positively associated with risk of IS, albeit the risk estimates were smaller than, but compatible with, the estimates previously reported in Western populations, including the ERFC,[14] CCHS,[83, 107] and AMORIS studies.[116] In both the ERFC and the CKB studies, the associations of triglycerides with IS were chiefly attenuated (by about a half) after adjustment for SBP. By contrast, in the ERFC study,[14] the association of triglycerides with CHD was marginally attenuated after adjustment for SBP, probably reflecting the much more predominant effect of higher SBP on risk of IS, compared with that for CHD. Consequently, in both the ERFC, and the present study, the effect of further adjustment for HDL-C and non-HDL-C on the associations of triglycerides with IS may be less informative than those observed for CHD, if any. In the AMORIS study,[116] plasma triglyceride concentrations were positively associated with risk of IS (HR 1.17; 95% CI: 1.14-1.21 per 1 SD higher; calculated through the published group-specific estimates). As previously described, the AMORIS study was unable to adequately adjust for diabetes and blood pressure. In the present CKB study, similar results were observed without adjusting for diabetes, or systolic blood pressure (HR 1.14; 95% CI: 1.08-1.20 per SD [0.58] higher triglycerides [Log_e]).

The present study demonstrated that triglycerides were inversely associated with risk of ICH, independent of other cholesterol fractions, suggesting that the relevance of triglycerides to ICH risk was unexplained by its correlations with

LDL-C (r=0.17). Triglycerides were moderately correlated with BMI (r=0.34), and consistently, adiposity, given SBP, has also been observed to be inversely associated with ICH risk in different populations.[261, 262] The primary analyses in the present study did not adjust for adiposity, because adiposity acts partly through lipid pathways,[60, 61] so is of little independent relevance in these multivariable analyses. However, further adjustment for BMI in sensitivity analyses did not alter the strength of these associations. Interestingly, as an exploratory analysis, triglycerides appeared to account for the inverse association of BMI with risk of ICH (Data not shown). The potential mechanisms underlying the inverse association of triglycerides with risk of ICH observed among this Chinese population are not fully understood. Moreover, there has been no worldwide clinical trial data, or genetic evidence on this, in China and elsewhere.

In the present study, the triglycerides GRS of four genetic variants chiefly affecting the lipoprotein lipase activity, was positively associated with triglycerides, and as expected, inversely associated with HDL-C. However, there was no evidence of any association of the triglycerides GRS with risk of stroke pathological types, due probably to the lack of statistical power. Previous genetic studies assessing the causal relevance of triglycerides were chiefly concerning coronary risk, while only one existing study examined the genetic relevance of *APOC3* to risk of IS included in a composite outcome, with over 70% being attributable to IHD.[295] The triglycerides associated genetic variants, including those within *LPL*,[292] *APOA5*,[124, 293] *APOC3*,[294, 295] *ANGPTL4*,[296] *ANGPTL3* regions,[297, 298] and multiple genetic variants,[312, 313] each tended to suggest higher plasma triglyceride concentrations (or more accurately, triglyceride-mediated pathways) are associated with higher risk of occlusive vascular disease.

Those genetic variants associated with triglycerides tend to be correlated with other lipid fractions, in particular HDL-C. Interpretation of findings from MR studies involving multiple genetic variants may be difficult in certain situations, as the metabolic process of triglyceride-rich lipoproteins also have effects on other lipoproteins. For example, the resulting inverse correlation between triglycerides and HDL-C in the lipolysis process, as suggested in the present study, and many others,[124, 294, 296] and the inter-correlation between concentrations of triglycerides and non-HDL-C.

Several lipid modifying agents, including fibrates, statins, and omega-3 fatty acids can lower triglyceride concentrations, among which, fibrates appear to be the most potent. Due perhaps to a lack of reliable large-scale clinical trial evidence, there is no consensus in guidelines on treatment of hypertriglyceridemia for prevention of cardiovascular events.[83, 220-222] In a summary-level based meta-analysis of 18 RCT studies, involving around 45,000 participants, treatment with fibrates decreased plasma concentrations of triglycerides, and improved other lipid profiles (e.g., decreased LDL-C and increased HDL-C).[148] In this meta-analysis, beneficial effects of fibrates were chiefly found for the prevention of coronary events (RR 0.87; 95% CI: 0.81-0.93), and had no effect on stroke (RR 1.03; 95% CI: 0.91-1.16).[148] The absence of any beneficial effects on stroke may suggest a less predominant role of atherosclerosis in risk of stroke, and certainly in HS. Furthermore, the large on-going randomised trial of new generation of fibrate – Pemafibrate to Reduce Cardiovascular OutcoMes by Reducing Triglycerides IN patiENts With diabeTes (PROMINENT; NCT03071692) – of around 10 000 participants with diabetes and dyslipidaemia is scheduled to complete in 2022,

and may provide further insights into the role of triglycerides on cardiovascular disease outcomes.

7.3 Strengths and limitations of the study

In estimating the stroke incidence rates, the use of multiple overlapping surveillance systems – the mortality, disease surveillance system, and the health insurance systems – in identifying stroke events, should have substantially reduced the underreporting of stroke events. Collection of data on disease outcomes through a number of pre-existing data sources should have reduced the risks of selection and recall bias, as well as limiting the effects of biases inherent in any individual data source. The high rates of CT/MRI scanning for stroke diagnoses have enabled a more accurate estimate of the epidemiology of stroke pathological types, rather than the aggregate of stroke in general. Although the CKB study was not designed to be representative, the large scale and diversity of the population in the CKB study, in terms of geographical sites, socioeconomic status, and patterns of exposure and disease, would have not invalidated its generalisation to the Chinese population at large. In addition, the CKB study has covered a wide age range groups, including those older than 65 years, in whom the incidence (and caseload at least in the areas studied) of stroke is high.

For analyses of the associations of major blood lipids with risk of IS and ICH, the present study included individuals from a single population, and hence minimised the possible heterogeneity between cohort studies that may occur in large international consortia of studies from different populations. Furthermore, the large number of well-characterised IS cases (~93% of validated stroke cases had been confirmed with CT/MRI), is a remarkable strength of the present prospective study,

along with the use of directly-measured lipids, compared to the use of either TC, or calculated variables (e.g., non-HDL-C, or LDL-C calculated from Friedewald equation[179]). The diversity of the data collected in the CKB enabled a comprehensive adjustment for potential confounding factors, and thus, the residual confounding is likely to have been largely eliminated. Data collected in resurveys provided assessment of, and correction for, the regression dilution bias in major blood lipid values, enabling the assessment of prolonged exposure to differences in lipid values.[67, 177] Although genetic markers selected in the present study were based on GWAS studies predominantly conducted in Western populations, the allele frequencies and the effect estimates on lipid fractions observed in the CKB study are largely consistent with those observed in the GLGC studies of Western populations.

While the adjustment for regression dilution bias of major blood lipids is the strength of the present analyses, the RDR estimates in the present CKB study may still be subject to residual confounding, due to the relatively limited number of participants with re-measurements. Despite this, RDR estimates in the CKB study were largely comparable with those observed in previous studies with similar duration of follow-up.[13, 14] Furthermore, in regression analyses, confounding variables, may also be susceptible to regression dilution bias and should ideally be adjusted for regression dilution bias,[226, 227] but this again is limited by the relatively small number of individuals with repeat measurements, and thus, further consideration of the multivariate regression dilution in the present analyses may have added greater uncertainty to the estimates.[228-230, 314]

Despite the greater absolute number of definite stroke pathological types recorded in the CKB study, compared with most of previous studies,[13, 14, 111, 112, 315] the present study still had limited statistical power for several subgroup analyses (e.g., age-at-risk groups or by levels of SBP), and it was not feasible to routinely exclude cases occurring during the first few years of follow-up. Therefore, further studies involving much larger numbers of definite IS and ICH events, and also on subtypes of IS and ICH, would enable more reliable assessment of the associations of major blood lipids with stroke pathological types, and further subtypes of IS and ICH. Furthermore, the present genetic analyses were based on a subset of the genetic data (In CKB, large GWAS studies covering all the previously identified lipid-related loci are still on-going), which has somewhat limited the strength of the instrumental variables, especially for HDL-C and triglycerides. Future studies using a specific gene region, or multiple genetic variants involving certain biological pathways, with large numbers of stroke cases may therefore provide more reliable evidence on lipids, particularly for triglycerides. A further limitation of the present analyses were not possible in subtypes of IS or of ICH, which may have different pathophysiology reflecting variable relevance of atherosclerosis. The strength of the association of usual LDL-C with IS risk estimated in the present study may therefore be an averaging effect of LDL-C on different subtypes of IS, including both subtypes in which LDL-C may be more relevant (e.g., large arterial atherosclerotic and cardio-embolic IS), and others where LDL-C may be less relevant (e.g., small vessel IS).

7.4 Public health and clinical implications

7.4.1 Stroke and low-density lipoprotein cholesterol

Among this Chinese population with relatively low plasma LDL-C concentrations (mean: 2.3 mmol/L), throughout the range of LDL-C studied (i.e., 1.7-3.2 mmol/L), there was a continuous log-linear association of usual LDL-C concentration with IS risk, and the proportional relevance was similar to those observed in Western populations (mean: 3.4 mmol/L). In contrast, throughout the range of usual LDL-C concentration studied in CKB, there was a proportional excess risk of ICH associated with lower LDL-C concentration, and this observation in general population-based study was corroborated with evidence from clinical trials among high-risk individuals, and was comparable with findings from the genetic studies.

If the proportional relevance of LDL-C to stroke risk in different subgroups really is similar, then the net effects (benefits versus hazards) of differences in usual LDL-C on total stroke risk would depend on the underlying absolute rate ratios of ischaemic to haemorrhagic stroke in different settings. **Figure 7.9** provides a hypothetical calculation of what the absolute effects of LDL-C lowering on 5-year stroke events (IS: 20% risk reduction; versus ICH: 10-20% risk increment) would be by statin treatment (estimated through combining published CTT,[316] and SPARCL trial results), if the annual stroke rates were as estimated among the CKB participants (**Tables 7.2-3**), with and without previous vascular events (i.e., CHD, stroke, and TIA), or with and without hypertension, at the baseline survey (2004-2008). The absolute excess numbers of ICH were greater in individuals where the underlying ICH rates are higher, of the elderly, individuals with elevated blood pressure, or those with pre-existing cardiovascular events. **Figure 7.9**

suggests that, in the context of stroke, lowering LDL-C with statins should provide definite net benefits, among this Chinese population with averaging higher rates of strokes than the Western populations, irrespective of the pathological type.[6, 10] Moreover, the net beneficial effects at population levels are likely to increase, if the increasing incidence rates of IS along with the decreasing rates of HS observed over the last two decades continue in decades ahead.

7.4.2 Stroke and high-density lipoprotein cholesterol

Although the causal relevance of HDL-C in atherosclerosis has been challenged, it has been suggested that the HDL function hypothesis may be relevant.[78, 147, 283] In this Chinese population, HDL-C concentration was still inversely associated with risk of IS, but not with ICH. The differential role (or contrasting results) of HDL-C with risk of IS and ICH has, however, confirmed the predictive role of HDL-C in atherosclerosis in a general Chinese population, independent of LDL-C. Despite the increasing number of studies questioning about the causal role of HDL-C in atherosclerosis, measurements of HDL-C concentrations may still have some relevance for prediction of occlusive vascular disease in a general population. By contrast, therapeutic approaches to raising HDL-C concentrations beyond those commonly observed in general populations have not demonstrated any cardio-protective effects that are attributable to HDL-C,[144-147] and uncertainty persists about the long-term effects on clinical outcomes.

7.4.3 Stroke and triglycerides

Albeit much weaker than for LDL-C, the positive association of plasma triglycerides with risk of IS is suggestive of a causal role of triglycerides in atherosclerosis. However, it is unclear if triglycerides act as a biomarker of the

cholesterol content of the triglyceride-rich lipoproteins or a measurement of the triglyceride-mediated pathways. This has, to some extent, justified the use of non-HDL-C to replace LDL-C as a predictive biomarker for atherosclerosis, as non-HDL-C includes the cholesterol measurements in all atherogenic lipoproteins, including triglyceride-rich lipoproteins. Furthermore, measurement of non-HDL-C can also overcome the inaccurate measurements of LDL-C, when multiple lipid-lowering therapies are used (e.g., anacetrapib). There is no consensus in current guidelines on whether triglycerides should be treated for the prevention of cardiovascular disease outcomes, although there is definite agreement on treating elevated triglycerides for prevention and treatment of pancreatitis.[220, 221, 317, 318]

The inverse association of triglycerides with risk of ICH has not been previously reported, and cannot be explained by the correlation between triglycerides and LDL-C. Hence, the public health and clinical implications remain unclear, unless this can be further validated in other populations or studies, or confirmed with better genetic evidence.

7.4.4 Stroke pathological types and major blood lipids

In this population, the associations of major blood lipid fractions, including LDL-C, HDL-C, and triglycerides, with risk of stroke differ qualitatively by stroke pathological type, although findings in this thesis cannot confirm or refute the causal nature of the associations of HDL-C, and triglycerides with stroke. The contrasting associations of major blood lipids with stroke pathological types are probably real. For, it is unlikely that some unknown confounders are correlated with three major lipid fractions, and risk of two types of stroke, in a highly consistent way. The definite qualitative heterogeneity of the proportional relevance

of major blood lipids for risk of IS and ICH suggests that a differential role of atherosclerosis in the aetiology of stroke pathological types. It also suggests, when an intervention, has relevance to different outcomes that differ in direction, then their combination in a composite outcome will reduce the ability to detect any outcome-specific effects and limit the generalisability of the analyses. Furthermore, if the intervention produces similar proportional reductions in IS risk, and increases in ICH risk, then the effect on the aggregate of stroke would depend on the ratio of strokes attributable to either IS or ICH in a particular setting. For example, in the large PSC meta-analysis of observational studies,[13] where half of the stroke deaths were of unknown pathological types, and of those with known stroke pathological types, half were attributable to haemorrhage, the relatively similar proportions of stroke attributable to either IS or ICH may have partially explained the weaker (or even null) association between prolonged exposure to higher TC and risk of total stroke.

7.4.5 Stroke prevention in China and worldwide

Overall, Chinese populations have been reported to have a higher incidence rates of stroke, and higher proportions of total stroke attributable to haemorrhage, than Western populations.[7, 10, 52, 58] In the present study, incidence rates of two stroke pathological types were higher than those observed in Western populations, but the ratios of strokes attributable to each pathological type appeared to be consistent with that observed in Western populations. In the setting of stroke prevention, the chief risk factors for IS were also risk factors for HS (e.g., age, blood pressure, smoking, and the presence of cerebrovascular or other occlusive vascular disease),[12] of which, elevated blood pressure appeared to be more

important for risk of haemorrhagic than ischaemic stroke.[67, 319, 320] Consistent with the previous studies, the present study has demonstrated discrepant effects of major blood lipids on aetiology of strokes. This study confirmed the causal relevance of LDL-C for risk of IS, and demonstrated an excess risk of ICH associated with lower LDL-C. The findings of the present study would inform the debate about the net utility of lipid-lowering therapy, particularly LDL-C lowering, for the primary and secondary prevention of stroke, in China and elsewhere.

In the setting of stroke prevention and treatment, the net benefit would depend on the ratio of ischaemic to haemorrhagic stroke in a population (i.e., benefits versus hazards), and the underlying absolute rates of stroke (i.e., the overall number of cases). Major blood lipid fractions studied in the thesis appeared to affect the type of stroke more than the overall stroke numbers, compared to other major risk factors for stroke, such as higher levels of blood pressure. The likely 10-20% higher risk of ICH associated with each 1 mmol/L lower LDL-C will be outweighed by the definite 40% lower risk associated with 10 mm Hg lower usual SBP or 5 mm Hg lower usual DBP (as has been observed in previous randomised trials of just a few years of blood pressure lowering[321-324]).

In general setting of preventing cardiovascular events attributable to one strategy or another in a particular population is, therefore, dependent on the rates of stroke and other occlusive vascular events. In China and other Asian countries where stroke rates are higher than that in Western countries, greater benefits at population levels are likely to accrue from blood pressure-lowering treatment. For example, a 10 mm Hg lower usual SBP (as has been typically assessed in previous randomised trials of just a few years of blood pressure lowering[321, 322, 324])

would, in the long term, be associated with about 40% lower risk of stroke deaths and 30% lower risk of cardiac deaths. Conversely, in Northern Europe and the USA, where CHD is more common than stroke, greater overall benefits at population levels are likely to accrue from lipid lowering than from blood pressure lowering.

7.5 Future research

Several analyses of CKB conducted in the present thesis would benefit from having a larger sample size, involving a larger number of participants with both baseline lipid measurements and of stroke events. Larger sample size would allow for estimation of effect modification for a wider range of age groups at risk, and of baseline blood pressure. Further research could also address some specific limitations of the present study. The effects of reverse causality could be assessed by repeating the main analyses of the thesis after excluding strokes occurring during the first few years of follow-up. Important future analyses might investigate the associations of usual blood lipids with subtypes of IS, important for understanding the mechanisms by which blood lipids cause IS, and could help to clarify the reason for the substantial weaker proportional relevance of major blood lipids to IS than that to IHD, in observational studies.

Specifically, for LDL-C, concerns regarding the excess risk of HS, especially among populations with relatively high rates of HS (e.g., hypertensive, or cerebrovascular patients), would benefit from clarifications with evidence from randomised trials in Chinese populations of lowering LDL-C treatment on the background of antihypertensive treatment. For the relevance of HDL to IS, or more generally, atherosclerosis, data with further classification of different subgroups of

HDL particles (e.g., large HDL particles) would permit investigation of the consistent inverse associations of HDL-C concentrations with risk of ischaemic vascular events. Regarding the relevance of triglycerides, in population-based studies, an inverse relationship has been observed between plasma triglycerides concentrations and HDL-C concentrations, as shown in the present genetic study. It is still unclear if low HDL-C concentrations (and thus increased risk) act as a marker for some defect in the metabolism of the triglyceride-rich lipoproteins. Given the relatively complex nature the triglyceride-mediated pathways, further research on the benefits of reducing residual risks of atherosclerotic vascular events, in addition to LDL-C lowering, would offer insights into the clinical relevance for individuals having "normal" concentrations of LDL-C (probably elevated proportion of small-dense LDL particles), and high concentrations of triglycerides, low HDL-C concentrations, particularly among individuals with diabetes. Future studies involved improved biomarker assays, including the nuclear magnetic resonance spectroscopy, together with genetic markers, should provide a better understanding of the relevance of biomarkers at a molecular level.

Future analyses of other vascular disease outcomes in the CKB study, (including, IHD), should be informative, as would analyses comparing the effects observed in other large prospective studies in diverse populations. For example, the UK Biobank study has over 0.5 million participants, and the GWAS data for the entire study population has been recently released,[167, 169] and other large prospective studies (e.g., the Million Women Study,[171] and the Mexico City Prospective Study[172]). The published reports from the METASTROKE consortium have assembled GWAS data on almost 12 000 IS cases and includes Western populations, whilst much larger numbers of IS cases (~16 000) from both Asian

and Western populations have been accrued. Any differences in the strength of associations between studies should provide additional insights into the aetiology, prevention, and treatment of stroke, and how this may differ in different parts of the world. Future meta-analyses of GWAS studies involving CKB, together with expanded METASTROKE, should afford opportunities to re-examine the causal relevance of HDL-C and triglyceride-related variants for risk of IS subtypes. The participation of the CKB study in other large genetic consortia (the CHARGE Consortium,[325] and the International Stroke Genetics Consortium[326]) should contribute importantly to an improved understanding of the stroke pathological types worldwide.[73]

7.6 Conclusions

In conclusion, the associations of major blood lipids with risk of stroke differ qualitatively by stroke pathological types. Lower LDL-C was causally associated with lower risk of IS, and probably, higher risk of ICH. In this Chinese population with high rates of stroke, irrespective of prior history of cardiovascular disease status, and hypertensive status, lower LDL-C was associated with net benefits, with the excess risk of ICH being outweighed by reduction in risk of IS. Likewise, plasma triglyceride concentrations were also positively associated with risk of IS, albeit weaker than for LDL-C. Triglycerides were inversely associated with risk of ICH, independent of LDL-C. By contrast, in this Chinese population, with relatively low LDL-C concentrations, there was an inverse association of usual HDL-C with IS risk, but no significant association of usual HDL-C with risk of ICH. Future research using better genetic instruments are needed to confirm or refute the

causal relevance, if any, of HDL-C, and of triglycerides for risk of stroke pathological types, observed in CKB.

7.7 Tables

Table 7.1: Comparison of associations of major blood lipids with ischaemic stroke and intracerebral haemorrhage*

Major lipids (SD)	RR (95% CI)		P for heterogeneity
	Ischaemic stroke	Intracerebral haemorrhage	
LDL cholesterol (0.68 mmol/L)	1.12 (1.07-1.16)	0.90 (0.86-0.95)	<0.00001
Triglycerides (0.58)	1.05 (1.00-1.10)	0.87 (0.83-0.91)	<0.00001
HDL cholesterol (0.29 mmol/L)	0.93 (0.89-0.97)	1.00 (0.96-1.05)	P=0.0006

SD=Standard deviation; * Rate ratios (RRs) are presented for differences that are in approximate proportion to the SD of the measured lipid values. Values for triglycerides are natural logarithm of triglyceride measurements.

Table 7.2: Age-specific rates of ischaemic stroke and haemorrhagic stroke in the China Kadoorie Biobank study, by presence or absence of cardiovascular disease at baseline

Age at risk	Ischaemic stroke			Haemorrhagic stroke		
	Person-years	Number of events	Rates (% per annum)	Person-years	Number of events	Rates (% per annum)
Participants with no prior cardiovascular disease (n=489 762)						
<50	1 486 878	2316	0.16	1 491 375	783	0.05
50-60	1 387 956	7829	0.56	1 407 239	1776	0.13
60-70	962 264	11 063	1.15	994 178	2599	0.26
≥70	484 144	11 661	2.41	521 980	3112	0.60
Overall	4 321 242	32 869	0.76	4 414 773	8270	0.19
Participants with prior cardiovascular disease (n=23 129)						
<50	9394	140	1.49	9666	58	0.60
50-60	37 871	1025	2.71	40 387	262	0.65
60-70	64 497	2486	3.85	72 127	534	0.74
≥70	57 237	3129	5.47	69 696	604	0.87
Overall	169 000	6780	4.01	191 876	1458	0.76
Total	***4 490 241***	***39 649***	***0.88***	***4 606 649***	***9728***	***0.21***

Prior cardiovascular disease was defined as physician-diagnosed ischaemic heart disease, stroke, or transient ischaemic attack at baseline.

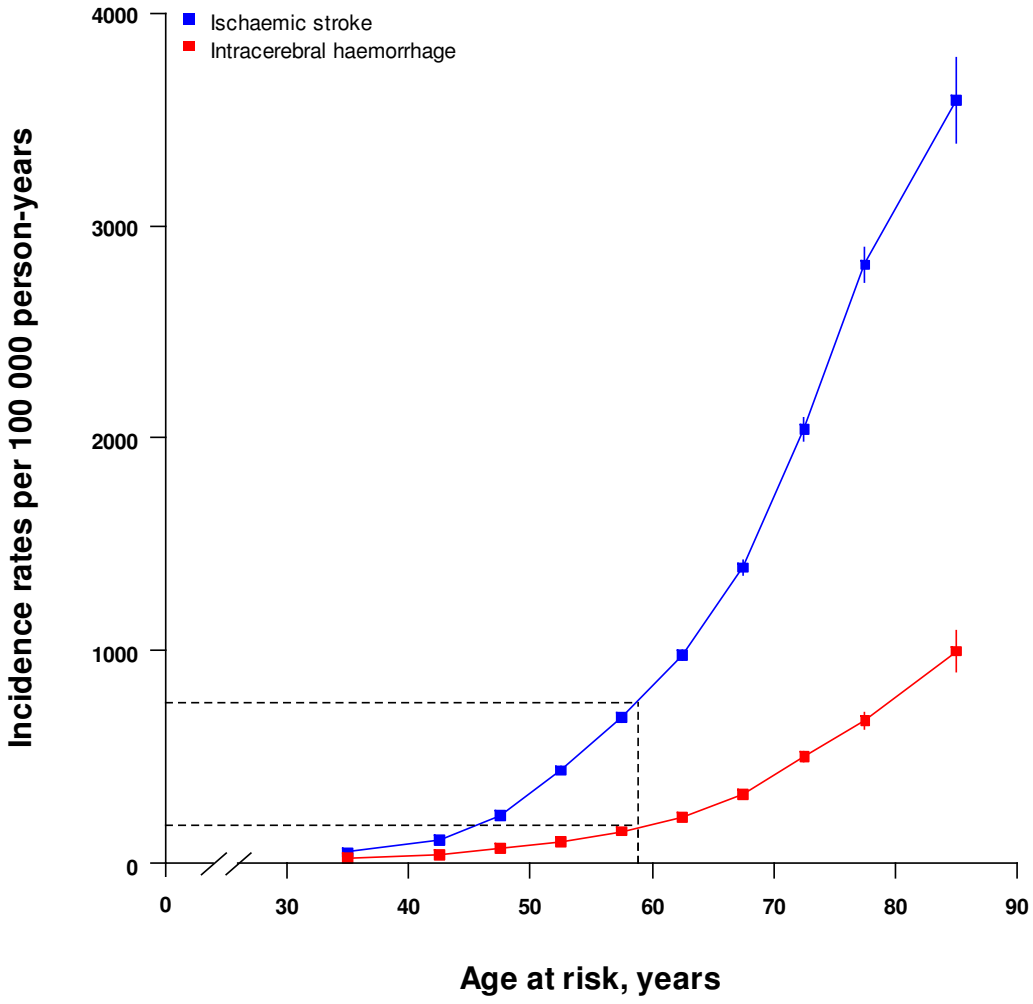
Table 7.3: Age-specific rates of ischaemic stroke and haemorrhagic stroke in the China Kadoorie Biobank study, by baseline hypertensive status

Age at risk	Ischaemic stroke			Haemorrhagic stroke		
	Person-years	Number of events	Rates (% per annum)	Person-years	Number of events	Rates (% per annum)
Participants with no hypertension (n=345 500)						
<50	1 264 355	1395	0.11	1 267 389	331	0.03
50-60	998 076	4342	0.44	1 009 143	580	0.06
60-70	583 703	5353	0.92	599 632	823	0.14
≥70	245 723	5395	2.20	263 855	1054	0.40
Overall	3 091 857	16 485	0.53	3 140 020	2788	0.09
Participants with hypertension (n=167 391)						
<50	231 918	1061	0.46	233 652	510	0.22
50-60	427 750	4512	1.05	438 483	1458	0.33
60-70	443 059	8196	1.85	466 673	2310	0.49
≥70	295 659	9395	3.18	327 820	2662	0.81
Overall	1 398 385	23 164	1.66	1 466 629	6940	0.47
Total	***4 490 241***	***39 649***	***0.88***	***4 606 649***	***9728***	***0.21***

Participants were considered to have hypertension if they had a measured systolic blood pressure of at least 140 mm Hg or a measured diastolic blood pressure of at least 90 mm Hg or were receiving treatment for hypertension. The latter was defined as those who reported a diagnosis of hypertension by a physician and use of anti-hypertensives at baseline.[242]

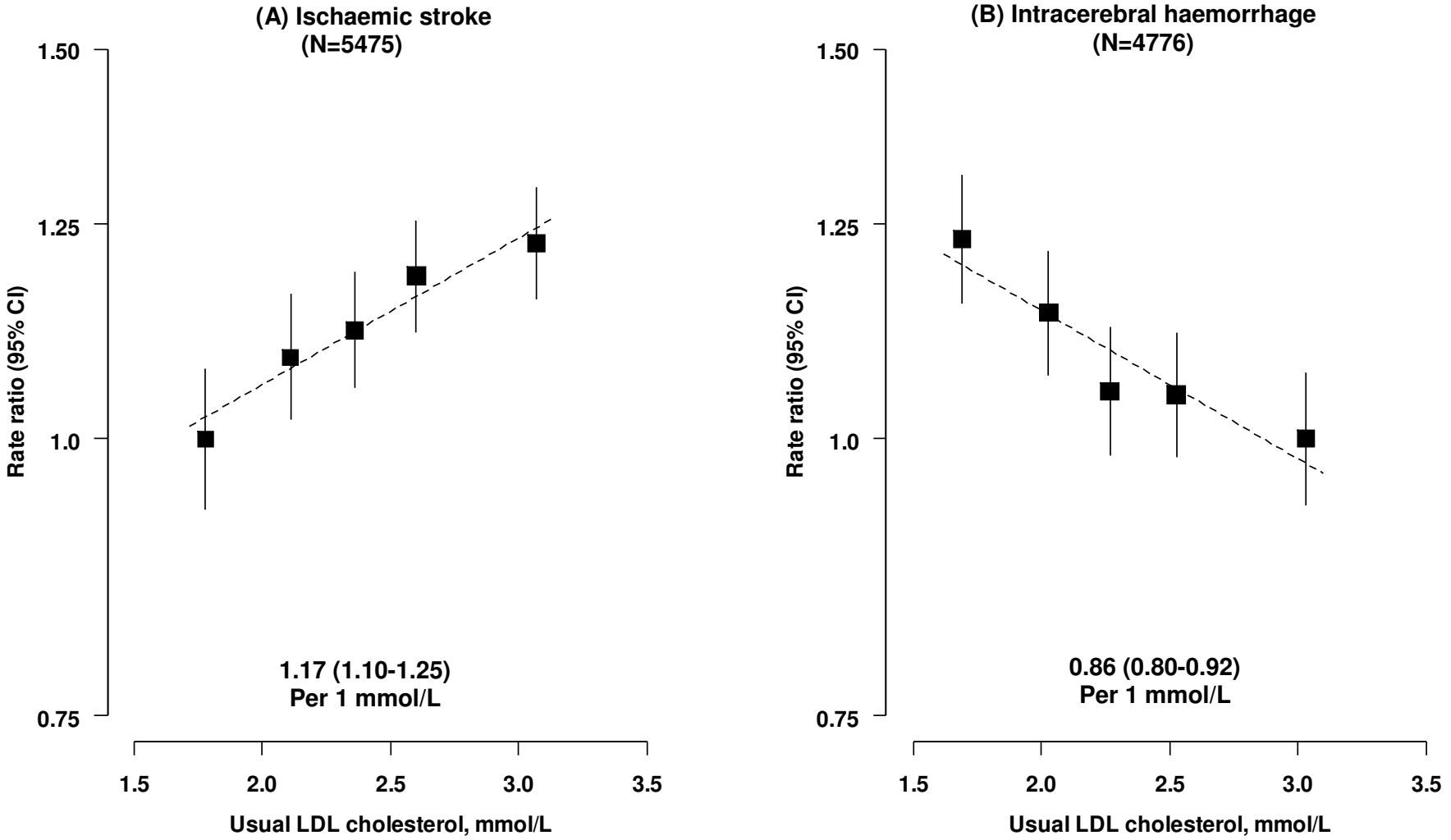
7.8 Figures

Figure 7.1: Incidence rates of ischaemic stroke and intracerebral haemorrhage in the China Kadoorie Biobank participants (n=489 762)*



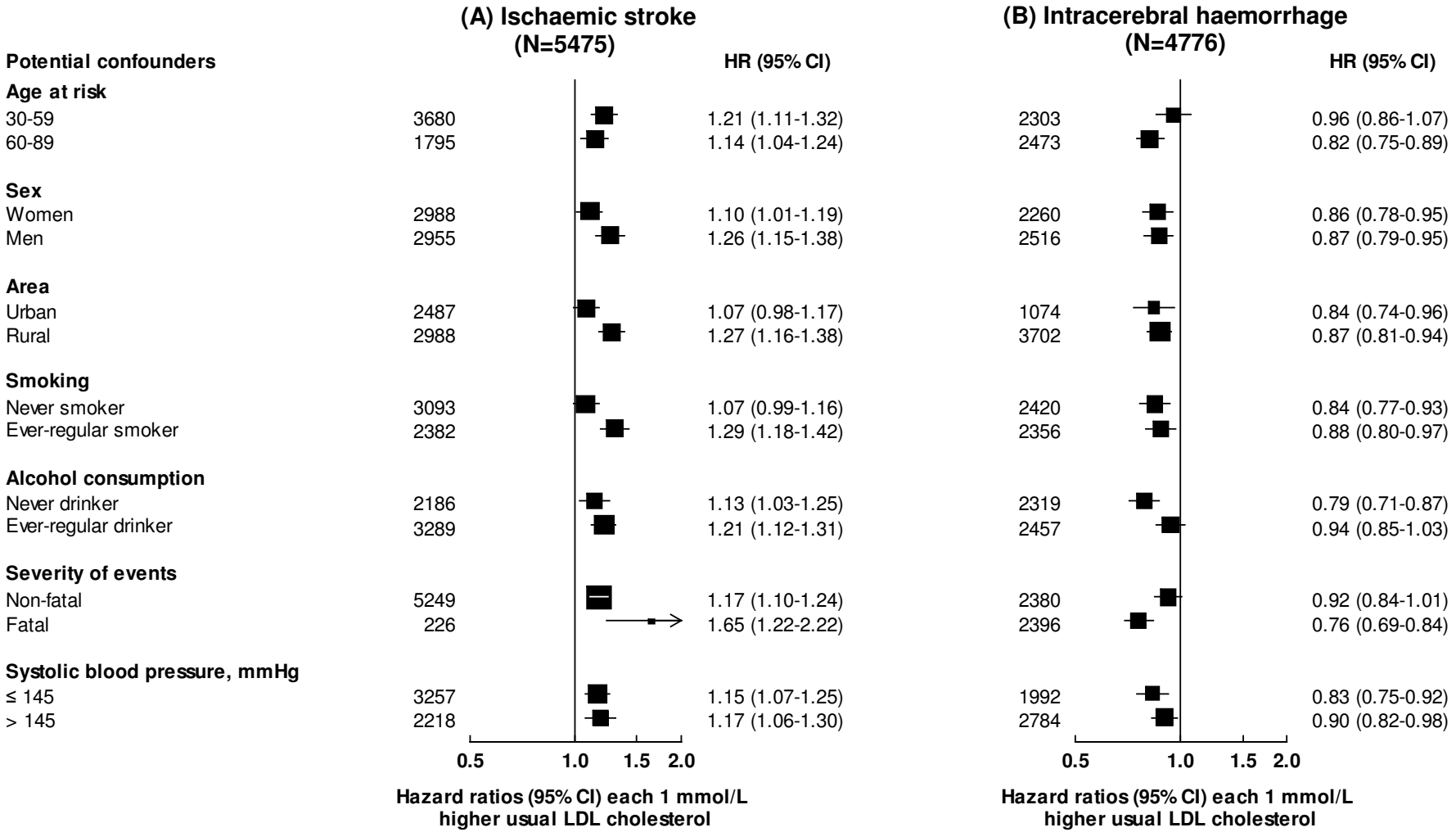
*Excluding individuals who reported doctor-diagnosed ischaemic heart disease, stroke or transient ischaemic attack at baseline. The vertical lines around each square represent the 95% confidence intervals, calculated using Poisson distribution. The dashed lines indicated the age at which the incidence rates of ischaemic stroke, and intracerebral haemorrhage in the overall study population were 761 and 187 per 100 000 person-years, respectively.

Figure 7.2: Observational associations of usual LDL cholesterol with risk of (A) ischaemic stroke and (B) intracerebral haemorrhage



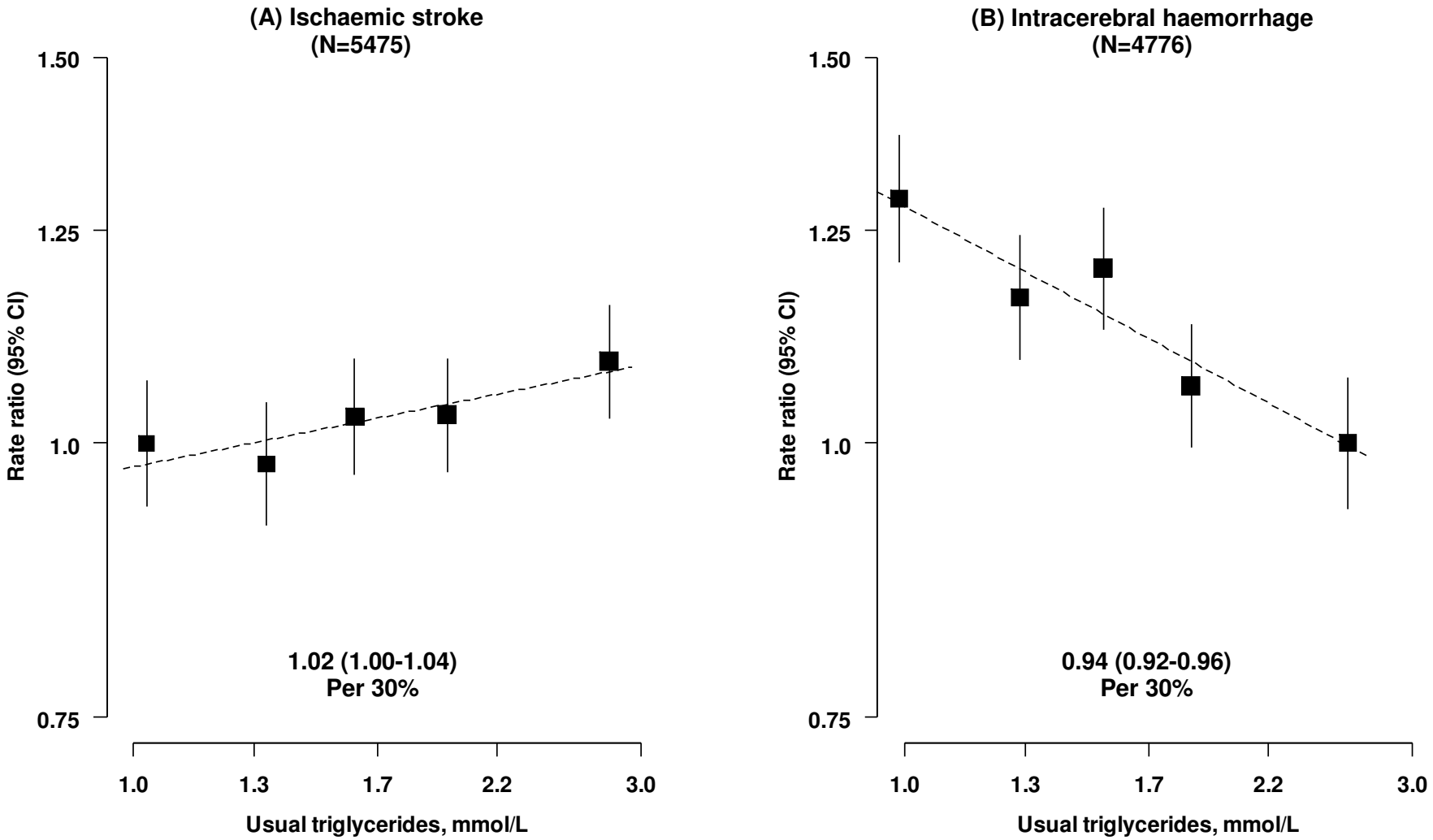
Analyses were stratified by age-at-risk (5-year), sex, and region, adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure. Each square has an area inversely proportional to the effective variance of the log risk in that specific group, with vertical line indicating the variance derived only from that category.

Figure 7.3: Observational associations of usual LDL cholesterol with risk of (A) ischaemic stroke (B) intracerebral haemorrhage, in population subgroups of other risk factors



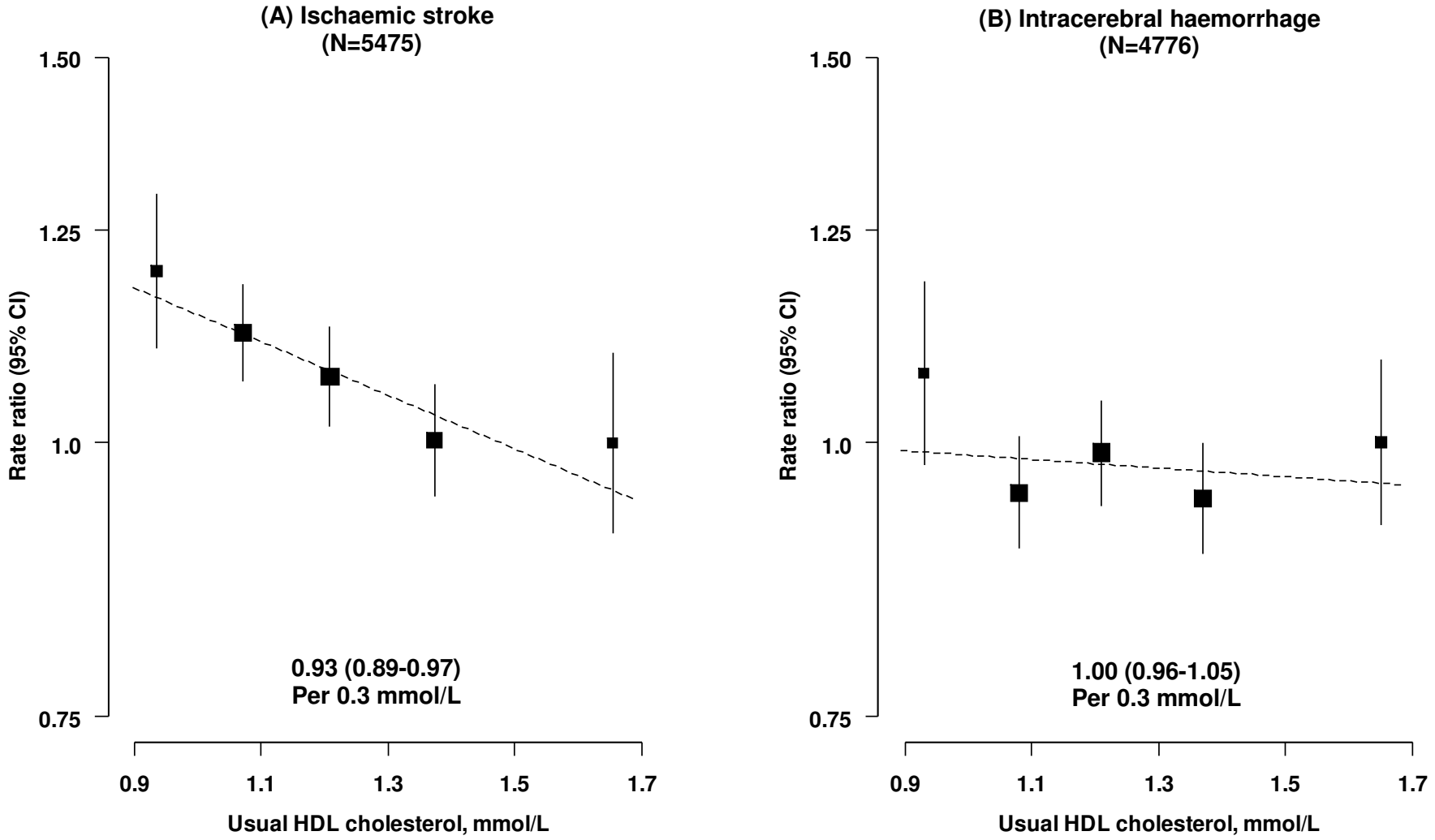
Analyses were stratified by age-at-risk (5-year), sex, and region, adjusted for education, smoking, alcohol consumption, physical activity, baseline diabetes, and systolic blood pressure, where appropriate. Reported associations were with usual LDL-C (i.e., corrected for regression dilution bias).

Figure 7.4: Observational associations of usual triglycerides with risk of (A) ischaemic stroke and (B) intracerebral haemorrhage



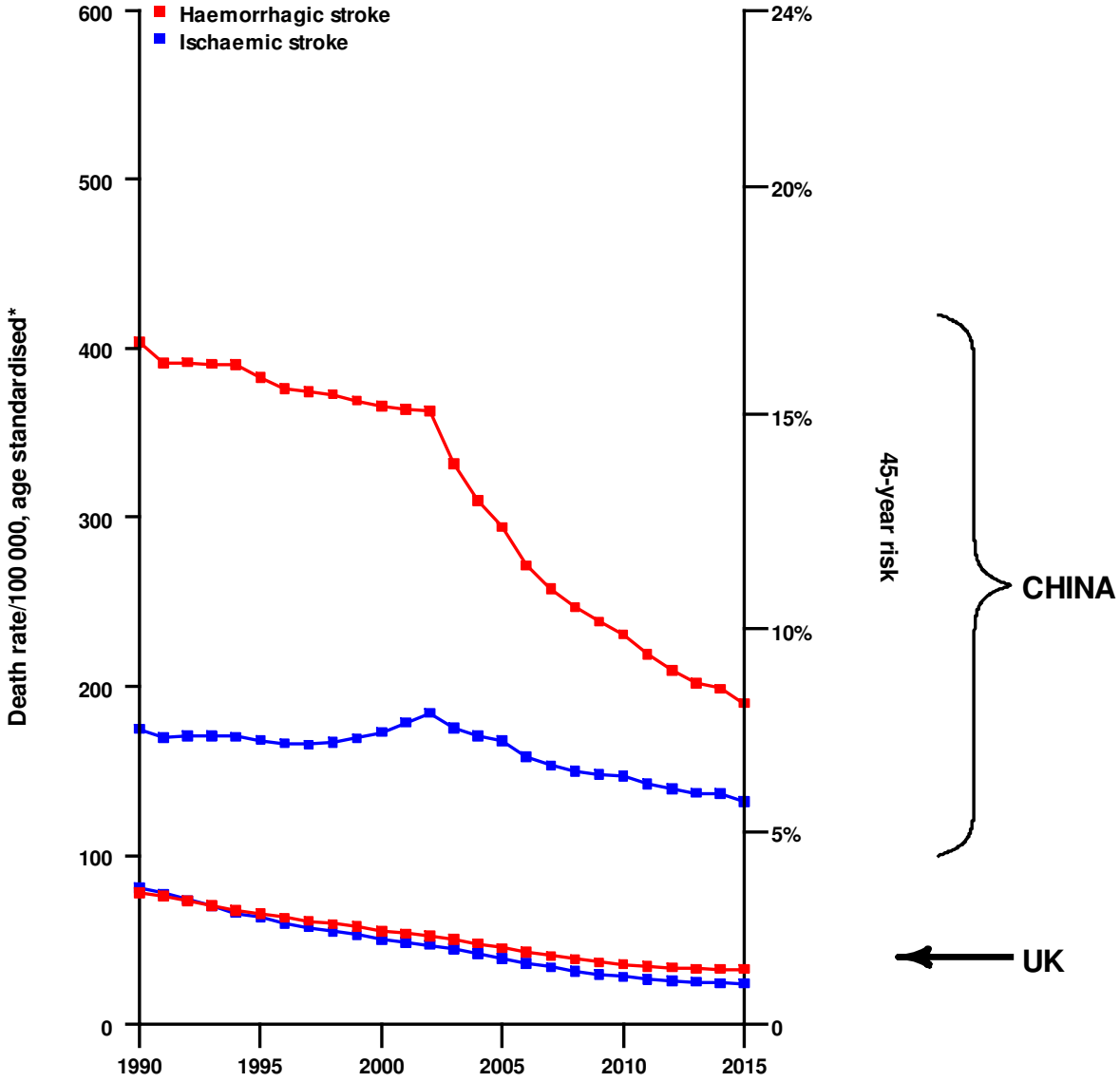
Conventions as in Figure 7.2

Figure 7.5: Observational associations of usual HDL-C with risk of (A) ischaemic stroke, and (B) intracerebral haemorrhage



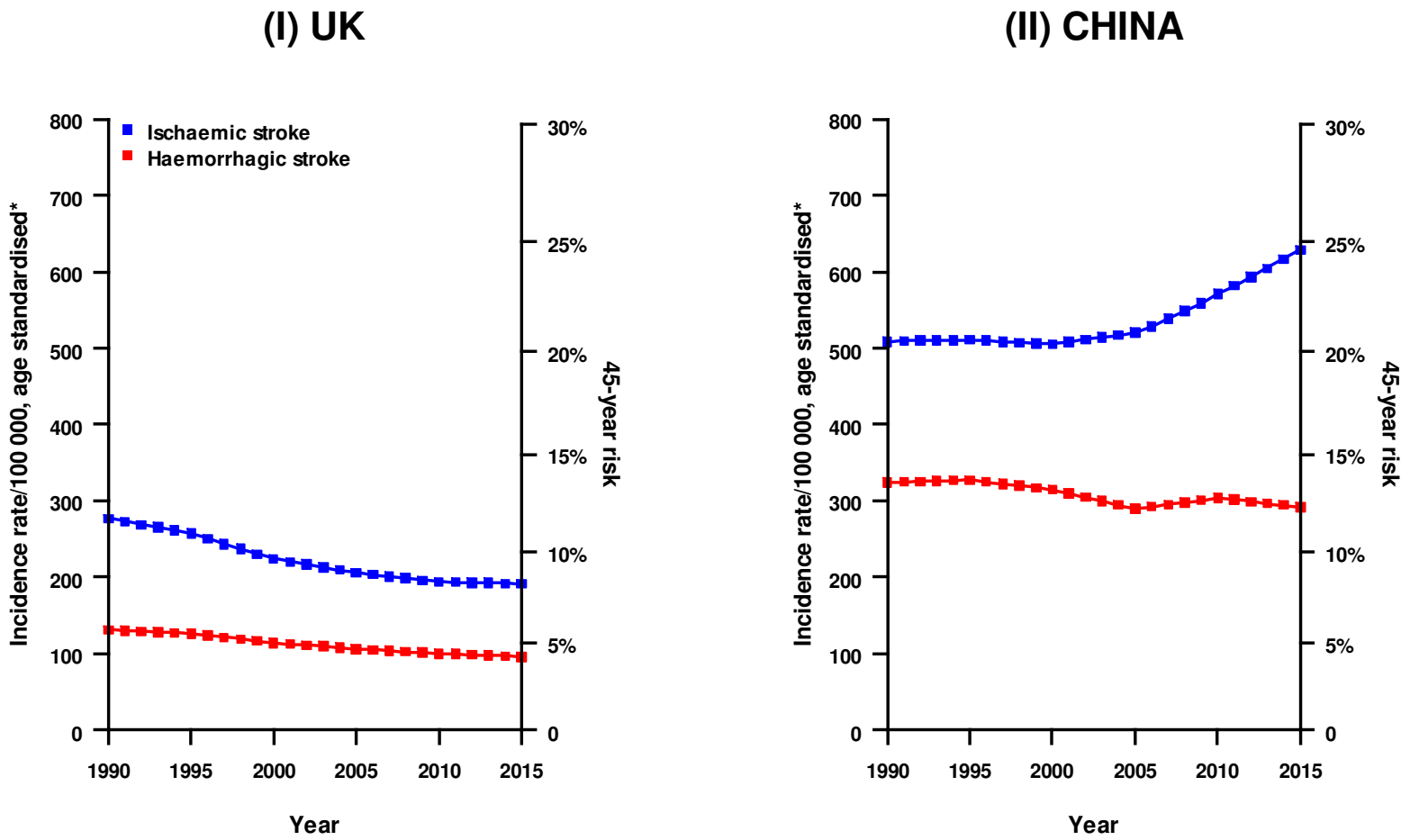
Conventions as in Figure 7.2

Figure 7.6: Mortality rates at ages 35-70, by stroke pathological type, between 1990-2015, for China and the UK



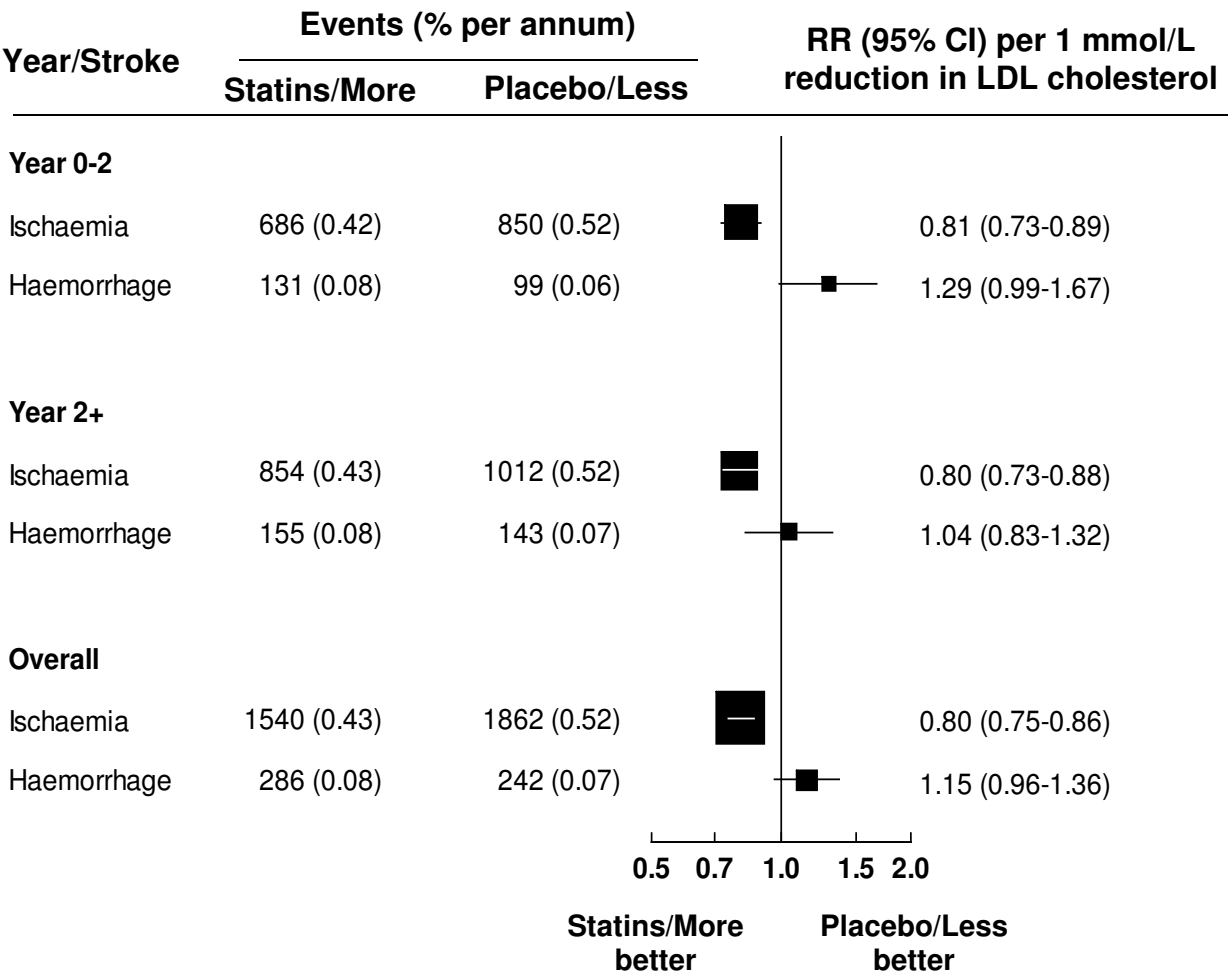
Source: Global Burden of Disease Study 2016; Mean of annual rates in the 9 component 5-year age groups (35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, and 75-79).

Figure 7.7: Incidence rates at ages 35-70, by stroke pathological type, between 1990-2015, for China and the UK



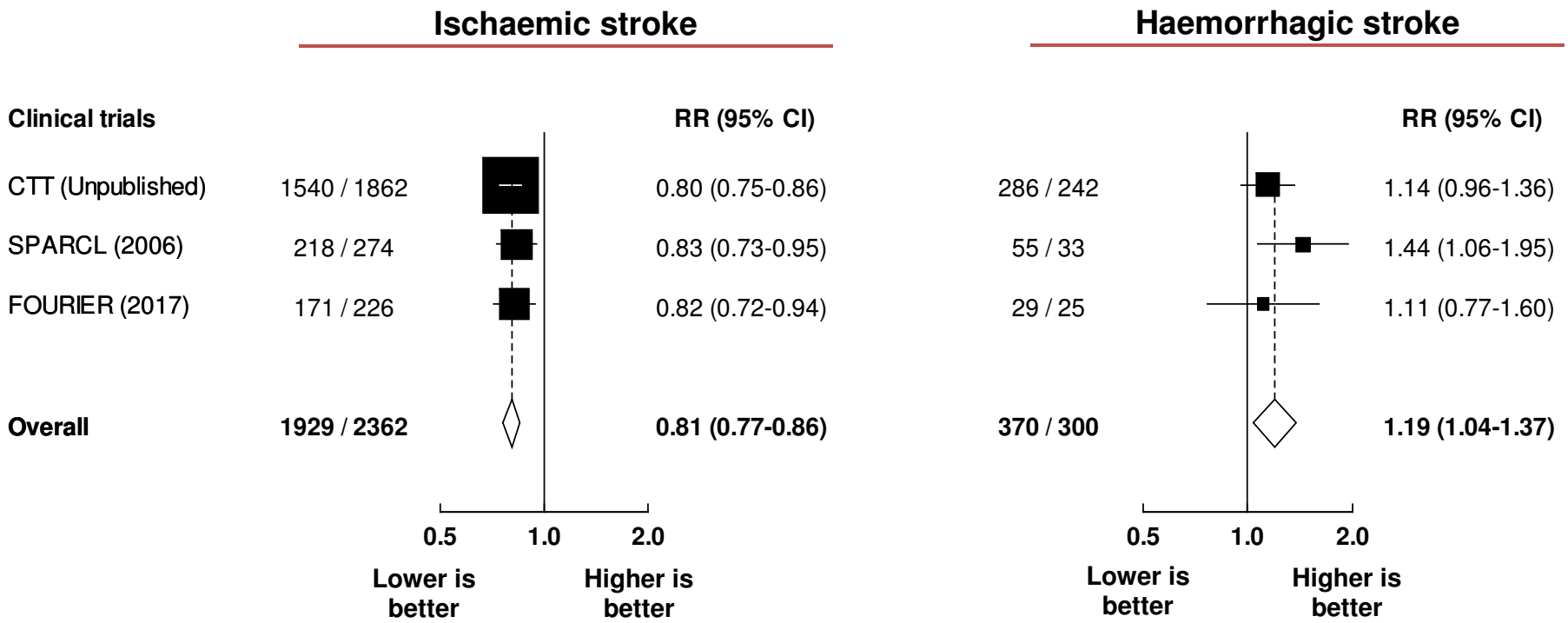
Source: Global Burden of Disease Study 2016; Mean of annual rates in the 9 component 5-year age groups (35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, and 75-79).

Figure 7.8: Effects of lowering LDL cholesterol with statin treatment on risk of stroke pathological types in the Cholesterol Treatment Trialists' Collaboration, by years of treatment duration



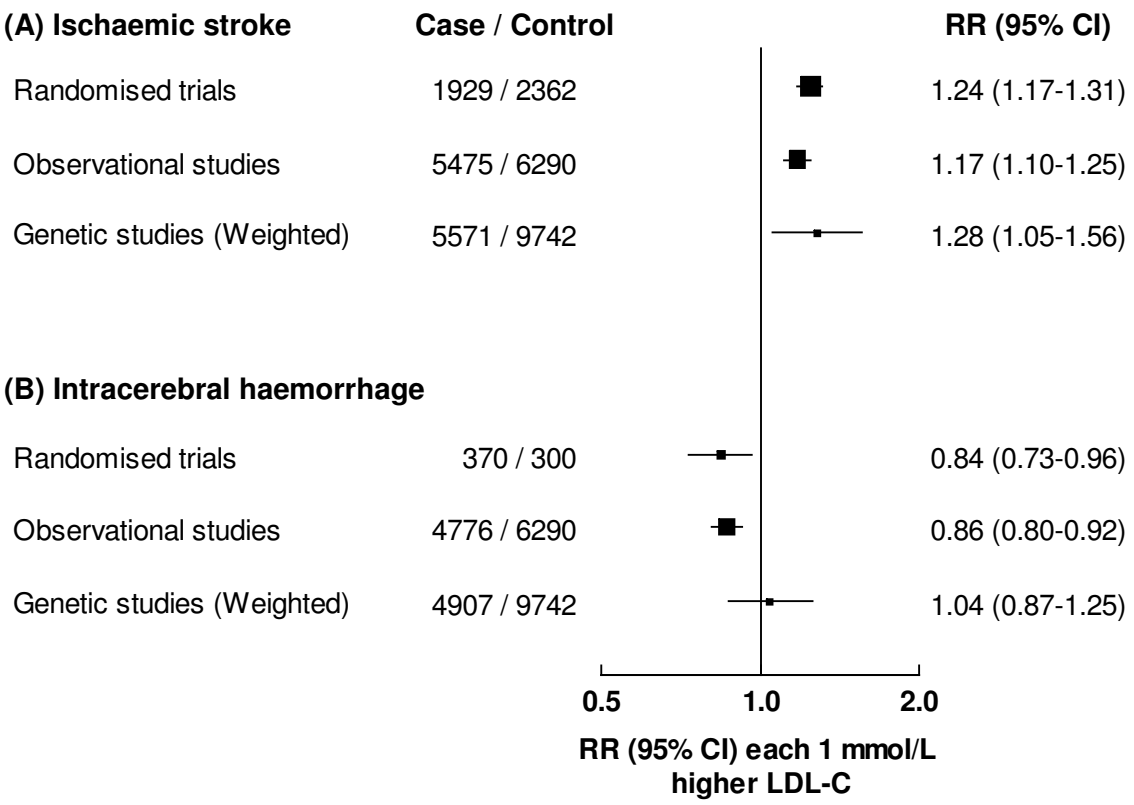
In each trial, the effects on stroke rates by year of follow-up duration were calculated as the logrank (o-e), and its variance (v) for first events weighted by the absolute LDL cholesterol difference in the trial (d mmol/L). In the meta-analysis of trials, the log of the rate ratio per 1 mmol/L (log RR) was calculated as S/V with variance $1/V$ (and hence with 95% CI of $S/V \pm 1.96/\sqrt{V}$), where S is the sum over all trials of $d(o\text{-}e)$ and V is the sum over all trials of d^2v.

Figure 7.9: Effects of lowering LDL cholesterol on ischaemic stroke and haemorrhagic stroke from meta-analysis of the CTT, SPARCL, and FOURIER trials



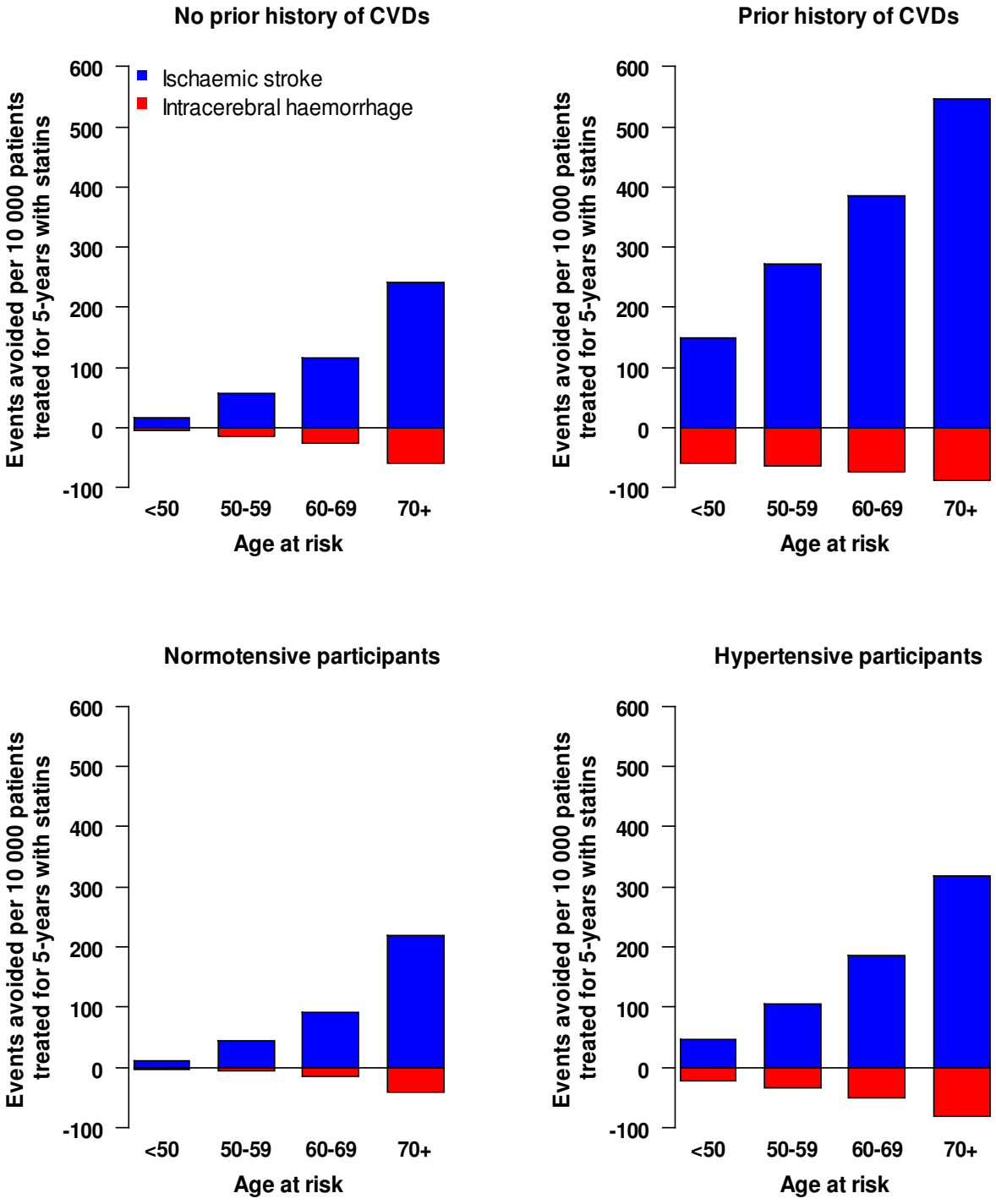
CTT=Cholesterol Treatment Trialist collaboration; SPARCL=The Stroke Prevention by Aggressive Reduction in Cholesterol Levels; FOURIER=The Further Cardiovascular Outcomes Research with PCSK9 Inhibition in Subjects with Elevated Risk (FOURIER). Results of each specific trial were reported for each 1 mmol/L lower LDL cholesterol. Overall estimates were combined using inverse-variance weighted average methodologies.

Figure 7.10: Comparison of the associations of LDL cholesterol with (A) ischaemic stroke and (B) intracerebral haemorrhage, from randomised trials, observational studies, and genetic studies



Estimates for randomised trials were obtained through meta-analysis of CTT, SPARCL, and FOURIER studies (see Figure 7.7 for details). Estimates for observational studies were obtained from Chapters 4-5. Estimates for genetic studies were taken from Chapter 6 using externally-weighted genetic risk score for LDL cholesterol.

Figure 7.11: Predicted absolute effects on ischaemic stroke (avoidance) and haemorrhagic stroke (excess) by lowering LDL cholesterol with statin therapy (per 10 000 patients treated for 5 years) by different characteristics in the Chinese population



CVD=Cardiovascular diseases, including doctor-diagnosed coronary heart disease, stroke, or transient ischaemic attack. The risk at each age group are equivalent to annual rates of ischaemic stroke and intracerebral haemorrhage among individuals with different baseline characteristics estimated in the China Kadoorie Biobank prospective study (**Tables 7.2-3**).

References

1. GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2013: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016; **388**(10053): 1603-58.

2. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016; **388**(10053): 1545-602.

3. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2013: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016; **388**(10053): 1459-544.

4. Global Burden of Disease Study 2015. Global Burden of Disease Study 2015 (GBD 2015) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME); 2015.

5. Krishnamurthi RV, Feigin VL, Forouzanfar MH, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet Global Health. 2013; **1**(5): e259-e81.

6. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of stroke during 1990 - 2010: findings from the Global Burden of Disease Study 2010. Lancet. 2014; **383**(9913): 245-55.

7. Zhou M, Wang H, Zhu J, Chen W, Wang L, Liu S, et al. Cause-specific mortality for 240 causes in China during 1990--2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. Lancet. 2015; **387**(10015): 251-72.

8. Feigin VL, Lawes CMM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. Lancet Neurology. 2003; **2**(1): 43-53.

9. Fang XH, Wang CX, Mei LP, Liu M, Ji XM, Li LM. [Progress in epidemiology study on stroke]. Zhonghua Liu Xing Bing Xue Za Zhi. 2011; **32**(9): 847-53.

10. Tsai CF, Thomas B, Sudlow CL. Epidemiology of stroke and its subtypes in Chinese vs white populations: a systematic review. Neurology. 2013; **81**(3): 264-72.

11. Huang CY, Chan Fl Fau - Yu YL, Yu Yl Fau - Woo E, Woo E Fau - Chin D, Chin D. Cerebrovascular disease in Hong Kong Chinese. Stroke. 1990; **21**(2): 230-5.

12. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet. 2016; **376**(9735): 112-23.

13. Lewington S, Whitlock G, Clarke R, Sherliker P, Emberson J, Halsey J, et al. 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. 2007; **370**(9602): 1829-39.

14. Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, Thompson A, et al. Major lipids, apolipoproteins, and risk of vascular disease. Journal of the American Medical Association. 2009; **302**(18): 1993-2000.

15. Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet. 2010; **376**(9753): 1670-81.

16. Cholesterol Treatment Trialists' (CTT) Collaborators. 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. Lancet. 2012; **380**(9841): 581-90.

17. Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174,000 participants in 27 randomised trials. Lancet. 2015; **385**(9976): 1397-405.

18. 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. Lancet. 2016; **388**(10059): 2532-61.

19. Muir KW. Stroke. Medicine; 2009; 37(2). p. 109-14.

20. Aho K, Harmsen P, Hatano S, Marquardsen J, Smirnov VE, Strasser T. Cerebrovascular disease in the community: results of a WHO collaborative study. Bulletin of the World Health Organization. 1980; **58**(1): 113-30.

21. Hatano S. Experience from a multicentre stroke register: a preliminary report. Bulletin of the World Health Organization. 1976; **54**(5): 541-53.

22. WHO MONICA Project Principal Investigators. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. Journal of Clinical Epidemiology. 1988; **41**(2): 105-14.

23. Easton JD, Saver JL, Albers GW, Alberts MJ, Chaturvedi S, Feldmann E, et al. Definition and Evaluation of Transient Ischemic Attack. Stroke. 2009; **40**(6): 2276.

24. Adams HP, Del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with ischemic stroke. Circulation. 2007; **115**(20): e478-e534.

25. Goldstein LB, Bushnell CD, Adams RJ, Appel LJ, Braun LT, Chaturvedi S, et al. Guidelines for the primary prevention of stroke. A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2010.

26. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993; **24**(1): 35-41.

27. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet. 1991; **337**(8756): 1521-6.

28. Brunser AM, Hoppe A, Illanes S, Diaz V, Munoz P, Carcamo D, et al. Accuracy of diffusion-weighted imaging in the diagnosis of stroke in patients with suspected cerebral infarct. Stroke. 2013; **44**(4): 1169-71.

29. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. Stroke. 2007; **38**(11): 2979-84.

30. Chen ZM, Sandercock P, Pan HC, Counsell C, Collins R, Liu LS, et al. Indications for early aspirin use in acute ischemic stroke : A combined analysis of 40 000 randomized patients from the chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups. Stroke. 2000; **31**(6): 1240-9.

31. Antithrombotic Trialists (ATT) Collaboration. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. Lancet. 2009; **373**(9678): 1849-60.

32. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJ, Culebras A, et al. An Updated Definition of Stroke for the 21st Century. Stroke. 2013; **44**(7): 2064.

33. Hankey GJ. Stroke. Lancet. 2017; **389**(10069): 641-54.

34. Wardlaw J, Keir S, Dennis M. The impact of delays in computed tomography of the brain on the accuracy of diagnosis and subsequent management in patients with minor stroke. Journal of Neurology, Neurosurgery & Psychiatry. 2003; **74**(1): 77-81.

35. Brazzelli M, Chappell FM, Miranda H, Shuler K, Dennis M, Sandercock PA, et al. Diffusion-weighted imaging and diagnosis of transient ischemic attack. Ann Neurol. 2014; **75**(1): 67-76.

36. Hill MD. Diagnostic biomarkers for stroke: a stroke neurologist's perspective. Clinical Chemistry. 2005; **51**(11): 2001-2.

37. Markus HS. Stroke genetics. Human Molecular Genetics. 2011; **20**(R2): R124-R31.

38. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age- and sex-specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014; **385**(9963): 117-71.

39. Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurology. 2009; **8**(4): 355-69.

40. Feigin VL, Roth GA, Naghavi M, Parmar P, Krishnamurthi R, Chugh S, et al. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Neurology. 2016; **15**(9): 913-24.

41. Thrift AG, Cadilhac DA, Thayabaranathan T, Howard G, Howard VJ, Rothwell PM, et al. Global stroke statistics. International Journal of Stroke. 2014; **9**(1): 6-18.

42. Thorvaldsen P, Asplund K Fau - Kuulasmaa K, Kuulasmaa K Fau - Rajakangas AM, Rajakangas Am Fau - Schroll M, Schroll M. Stroke incidence, case fatality, and

mortality in the WHO MONICA project. World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease. Stroke. 1995; **26**(3): 361-7.

43. Sudlow CLM, Warlow CP. Comparing stroke incidence worldwide what makes studies comparable? Stroke. 1996; **27**(3): 550-8.

44. Sudlow. CLM, Warlow. CP, International Stroke Incidence Collaboration. Comparable studies of the incidence of stroke and its pathological types results from an international collaboration. Stroke. 1997; **28**(3): 491-9.

45. Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA, et al. Update on the Global Burden of Ischemic and Hemorrhagic Stroke in 1990-2013: The GBD 2013 Study. Neuroepidemiology. 2015; **45**(3): 161-76.

46. Bennett DA, Krishnamurthi RV, Barker-Collo S, Forouzanfar MH, Naghavi M, Connor M, et al. The global burden of ischemic stroke: findings of the GBD 2010 study. Global Heart. 2014; **9**(1): 107-12.

47. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: Part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. Circulation. 2001; **104**(23): 2855-64.

48. Roth GA, Forouzanfar MH, Moran AE, Barber R, Nguyen G, Feigin VL, et al. Demographic and Epidemiologic Drivers of Global Cardiovascular Mortality. N Engl J Med. 2015; **372**(14): 1333-41.

49. Yang QD, Niu Q, Zhou YH, Liu YH, Xu HW, Gu WP, et al. Incidence of cerebral hemorrhage in the Changsha community. A prospective study from 1986 to 2000. Cerebrovascular Disease. 2004; **17**(4): 303-13.

50. Chau PH, Woo J, Goggins WB, Tse YK, Chan KC, Lo SV, et al. Trends in stroke incidence in Hong Kong differ by stroke subtype. Cerebrovascular Diseases. 2011; **31**(2): 138-46.

51. Jiang B, Wang W-z, Chen H, Hong Z, Yang Q-d, Wu S-p, et al. Incidence and trends of stroke and its subtypes in China results from three large cities. Stroke. 2006; **37**(1): 63-5.

52. Liu M, Wu B, Wang W-Z, Lee L-M, Zhang S-H, Kong L-Z. Stroke in China: epidemiology, prevention, and management strategies. Lancet Neurology. 2007; **6**(5): 456-64.

53. Wu Z, Yao C, Zhao D, Wu G, Wang W, Liu J, et al. Sino-MONICA project: a collaborative study on trends and determinants in cardiovascular diseases in China, Part i: morbidity and mortality monitoring. Circulation. 2001; **103**(3): 462-8.

54. Fang X, Kronmal RA, Li S, Longstreth WT, Cheng X, Wang W, et al. Prevention of Stroke in Urban China. Stroke. 1999; **30**(3): 495.

55. Thorvaldsen P, Kuulasmaa K Fau - Rajakangas AM, Rajakangas Am Fau - Rastenyte D, Rastenyte D Fau - Sarti C, Sarti C Fau - Wilhelmsen L, Wilhelmsen L. Stroke trends in the WHO MONICA project. Stroke. 1997; **28**(3): 500-6.

56. Kjell Asplund M, Rajakangas A-M, Thorvaldsen P, Tuomilehto J. Stroke Incidence and Mortality Correlated to Stroke Risk Factors in the WHO MONICA Project An Ecological Study of 18 Populations. Stroke. 1997; **28**(7): 1367-74.

57. Zhao D, Liu J, Wang W, Zeng Z, Cheng J, Sun J, et al. Epidemiological transition of stroke in China: twenty-one-year observational study from the Sino-MONICA-Beijing Project. Stroke. 2008; **39**(6): 1668-74.

58. Wang W, Jiang B, Sun H, Ru X, Sun D, Wang L, et al. Prevalence, Incidence and Mortality of Stroke in China: Results from a Nationwide Population-Based Survey of 480,687 Adults. Circulation. 2017; **135**(8): 759-71.

59. Carandang R, Seshadri S Fau - Beiser A, Beiser A Fau - Kelly-Hayes M, Kelly-Hayes M Fau - Kase CS, Kase Cs Fau - Kannel WB, Kannel Wb Fau - Wolf PA, et al. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. Journal of the American Medical Association. 2006; **296**(24): 2939-46.

60. Prospective Studies Collaboration. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. Lancet. 2009; **373**(9669): 1083-96.

61. The Global BMI Mortality Collaboration. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet. 2016; **388**(10046): 776-86.

62. Kroll ME, Green J, Beral V, Sudlow CL, Brown A, Kirichek O, et al. Adiposity and ischemic and hemorrhagic stroke Prospective study in women and meta-analysis. Neurology. 2016; **87**(14): 1473-81.

63. Lee CD, Folsom AR, Blair SN. Physical activity and stroke risk: a meta-analysis. Stroke. 2003; **34**(10): 2475-81.

64. Hart RG. Cardiogenic embolism to the brain. Lancet. 1992; **339**(8793): 589-94.

65. Lubitz SA, Parsons OE, Anderson CD, Benjamin EJ, Malik R, Weng L-C, et al. Atrial Fibrillation Genetic Risk and Ischemic Stroke Mechanisms. Stroke. 2017; **48**(6): 1451-6.

66. Asymptomatic Carotid Surgery Trial Collaborative Group. Prevention of disabling and fatal strokes by successful carotid endarterectomy in patients without recent neurological symptoms: randomised controlled trial. Lancet. 2004; **363**(9420): 1491-502.

67. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002; **360**(9349): 1903-13.

68. Kurth T, Kase CS, Berger K, Schaeffner ES, Buring JE, Gaziano JM. Smoking and the risk of hemorrhagic stroke in men. Stroke. 2003; **34**(5): 1151-5.

69. Reynolds K, Lewis B, Nolen JD, Kinney GL, Sathya B, He J. Alcohol consumption and risk of stroke: a meta-analysis. Journal of the American Medical Association. 2003; **289**(5): 579-88.

70. Sutherland GR, Auer RN. Primary intracerebral hemorrhage. Journal of Clinical Nueroscience. 2006; **13**(5): 511-7.

71. NINDS Stroke Genetics Network (SiGN), International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurology. 2015; **15**(2): 174-84.

72. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide Association Studies of Stroke. N Engl J Med. 2009; **360**(17): 1718-28.

73. Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium tSGNS, and the International Stroke Genetics Consortium (ISGC),. Identification of additional risk loci for stroke and small

vessel disease: a meta-analysis of genome-wide association studies. Lancet Neurology. 2016; **15**(7): 695-707.

74. Frayn KN. Metabolic regulation: a human perspective: John Wiley & Sons; 2009.

75. Chen Z, Peto R, Collins R, MacMahon S, Lu J, Li W. Serum cholesterol concentration and coronary heart disease in population with low cholesterol concentrations. BMJ. 1991; **303**(6797): 276-82.

76. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet. 1994; **344**(8934): 1383-9.

77. Collins R, Armitage J, Parish S, Sleight P, Peto R. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: A randomised placebo-controlled trial. Lancet. 2002; **360**(9326): 7-22.

78. Rader DJ, Hovingh GK. HDL and cardiovascular disease. Lancet. 2014; **384**(9943): 618-25.

79. Rye KA, Barter PJ. Cardioprotective functions of HDLs. J Lipid Res. 2014; **55**(2): 168-79.

80. Rye KA, Barter PJ. Regulation of high-density lipoprotein metabolism. Circ Res. 2014; **114**(1): 143-56.

81. Salonen JT, Salonen R, Seppänen K, Rauramaa R, Tuomilehto J. HDL, HDL2, and HDL3 subfractions, and the risk of acute myocardial infarction. A prospective population study in eastern Finnish men. Circulation. 1991; **84**(1): 129-39.

82. Karpe F. Postprandial lipoprotein metabolism and atherosclerosis. J Intern Med. 1999; **246**(4): 341-55.

83. Nordestgaard BG, Varbo A. Triglycerides and cardiovascular disease. Lancet. 2014; **384**(9943): 626-35.

84. Goldberg IJ, Eckel RH, McPherson R. Triglycerides and Heart Disease, Still a Hypothesis? Arteriosclerosis, Thrombosis, and Vascular Biology. 2011; **31**(8): 1716-25.

85. Zilversmit DB. Atherogenesis: a postprandial phenomenon. Circulation. 1979; **60**(3): 473.

86. Hegsted DM, Ausman LM, Johnson JA, Dallal GE. Dietary fat and serum lipids: an evaluation of the experimental data. American Journal of Clinical Nutrition. 1993; **57**(6): 875-83.

87. Clarke R, Frost C, Collins R, Appleby P, Peto R. Dietary lipids and blood cholesterol: quantitative meta-analysis of metabolic ward studies. BMJ. 1997; **314**(7074): 112-7.

88. Keys A, Anderson JT, Grande F. Prediction of serum-cholesterol responses of man to changes in fats in the diet. Lancet. 1957; **273**(7003): 959-66.

89. Mente A, Dehghan M, Rangarajan S, McQueen M, Dagenais G, Wielgosz A, et al. Association of dietary nutrients with blood lipids and blood pressure in 18 countries: a cross-sectional analysis from the PURE study. The Lancet Diabetes & Endocrinology. 2017; **5**(10): 774-87.

90. Dattilo AM, Kris-Etherton PM. Effects of weight reduction on blood lipids and lipoproteins: a meta-analysis. American Journal of Clinical Nutrition. 1992; **56**(2): 320-8.

91. Lakka T, Salonen J. Physical activity and serum lipids: a cross-sectional population study in eastern Finnish men. American Journal of Epidemiology. 1992; **136**(7): 806-18.

92. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013; **45**(11): 1274-83.

93. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V, et al. The impact of low-frequency and rare variants on lipid levels. Nat Genet. 2015; **47**(6): 589-97.

94. Keys A. Seven countries: a multivariate analysis of death and coronary heart disease: Harvard University Press; 1980.

95. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine. 2008; **27**(8): 1133-63.

96. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet. 2005; **366**(9489): 941-51.

97. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature Protocols. 2010; **5**(9): 1564-73.

98. Asselbergs FW, Lovering RC, Drenos F. Progress in genetic association studies of plasma lipids. Current Opinion in Lipidology. 2013; **24**(2): 123-8.

99. MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. Lancet. 2001; **357**(9254): 455-62.

100. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003; **32**(1): 1-22.

101. 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.

102. Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Med. 2007; **4**(12): e352.

103. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol. 2004; **33**(1): 30-42.

104. Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005; **366**(9501): 1906-8.

105. Burgess S, Thompson SG. Mendelian Randomization. Methods for Using Genetic Variants in Causal Estimation: Boca Raton, Florida: CRC Press; 2015.

106. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. Lancet. 2001; **357**(9253): 373-80.

107. Freiberg JJ, Tybjaerg-Hansen A, Jensen JS, Nordestgaard BG. Nonfasting triglycerides and risk of ischemic stroke in the general population. JAMA. 2008; **300**(18): 2142-52.

108. Iso H, Jacobs DR, Jr., Wentworth D, Neaton JD, Cohen JD. Serum cholesterol levels and six-year mortality from stroke in 350,977 men screened for the multiple risk factor intervention trial. N Engl J Med. 1989; **320**(14): 904-10.

109. Zhang X, Patel A, Horibe H, Wu Z, Barzi F, Rodgers A, et al. Cholesterol, coronary heart disease, and stroke in the Asia Pacific region. International journal of epidemiology. 2003; **32**(4): 563-72.

110. Woodward M, Huxley R, Ueshima H, Fang X, Kim HC, Lam TH. The Asia pacific cohort studies collaboration: a decade of achievements. Glob Heart. 2012; **7**(4): 343-51.

111. Ebrahim S, Sung J, Song YM, Ferrer RL, Lawlor DA, Davey Smith G. Serum cholesterol, haemorrhagic stroke, ischaemic stroke, and myocardial infarction: Korean national health system prospective cohort study. BMJ. 2006; **333**(7557): 22.

112. Wu J, Chen S, Zhou Y, Wang C, Wang A, Zhang Q, et al. Non-high-density lipoprotein cholesterol on the risks of stroke: a result from the Kailuan study. PloS one. 2013; **8**(9): e74634.

113. Jiang B, Fang XH, Liu YH, Dai XY, Lin ZG, Su FZ, et al. [Prospective study on associations between levels of total cholesterol, triglyceride and risk of ischemic and hemorrhagic strokes]. Zhonghua xin xue guan bing za zhi. 2010; **38**(3): 268-71.

114. Cui R, Iso H, Yamagishi K, Saito I, Kokubo Y, Inoue M, et al. High serum total cholesterol levels is a risk factor of ischemic stroke for general Japanese population: the JPHC study. Atherosclerosis. 2012; **221**(2): 565-9.

115. Wang W, Zhao D, Liu J, Zeng ZC, Sun JY, Qin LP, et al. [Serum total cholesterol and 10-year cardiovascular risk in a Chinese cohort aged 35-64 years]. Zhonghua xin xue guan bing za zhi. 2006; **34**(2): 169-73.

116. Holme I, Aastveit AH, Hammar N, Jungner I, Walldius G. Relationships between lipoprotein components and risk of ischaemic and haemorrhagic stroke in the Apolipoprotein MOrtality RISk study (AMORIS). Journal of internal medicine. 2009; **265**(2): 275-87.

117. Woodward M, Barzi F, Feigin V, Gu D, Huxley R, Nakamura K, et al. Associations between high-density lipoprotein cholesterol and both stroke and coronary heart disease in the Asia Pacific region. European heart journal. 2007; **28**(21): 2653-60.

118. Pikula A, Beiser AS, Wang J, Himali JJ, Kelly-Hayes M, Kase CS, et al. Lipid and lipoprotein measurements and the risk of ischemic vascular events: Framingham Study. Neurology. 2015; **84**(5): 472-9.

119. Madsen CM, Varbo A, Nordestgaard BG. Extreme high high-density lipoprotein cholesterol is paradoxically associated with high mortality in men and women: two prospective cohort studies. European heart journal. 2017; **38**(32): 2478-86.

120. Peto R. Current misconception 3: that subgroup-specific trial mortality results often provide a good basis for individualising patient care. Br J Cancer. 2011; **104**(7): 1057-8.

121. Bots ML, Elwood PC, Nikitin Y, Salonen JT, Freire de Concalves A, Inzitari D, et al. Total and HDL cholesterol and risk of stroke. EUROSTROKE: a collaborative study among research centres in Europe. Journal of epidemiology and community health. 2002; **56 Suppl 1**: i19-24.

122. Patel A, Barzi F, Jamrozik K, Lam TH, Ueshima H, Whitlock G, et al. Serum triglycerides as a risk factor for cardiovascular diseases in the Asia-Pacific region. Circulation. 2004; **110**(17): 2678-86.

123. Barter PJ, Rye KA. Is there a role for fibrates in the management of dyslipidemia in the metabolic syndrome? Arterioscler Thromb Vasc Biol. 2008; **28**(1): 39-46.

124. Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. Lancet. 2010; **375**(9726): 1634-9.

125. Hyun KK, Huxley RR, Arima H, Woo J, Lam TH, Ueshima H, et al. A comparative analysis of risk factors and stroke risk for Asian and non-Asian men: the Asia Pacific cohort studies collaboration. International journal of stroke : official journal of the International Stroke Society. 2013; **8**(8): 606-11.

126. Noda H, Iso H, Irie F, Sairenchi T, Ohtaka E, Doi M, et al. Low-density lipoprotein cholesterol concentrations and death due to intraparenchymal hemorrhage: the Ibaraki Prefectural Health Study. Circulation. 2009; **119**(16): 2136-45.

127. Nagasawa SY, Okamura T, Iso H, Tamakoshi A, Yamada M, Watanabe M, et al. Relation between serum total cholesterol level and cardiovascular disease stratified by sex and age group: a pooled analysis of 65 594 individuals from 10 cohort studies in Japan. Journal of the American Heart Association. 2012; **1**(5): e001974.

128. Iribarren C, Jacobs DR, Sadler M, Claxton AJ, Sidney S. Low total serum cholesterol and intracerebral hemorrhagic stroke: is the association confined to elderly men? The Kaiser Permanente Medical Care Program. Stroke. 1996; **27**(11): 1993-8.

129. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012; **380**(9841): 572-80.

130. Hopewell J, Stari T, Parish S, al. E. The impact of genetic variants related to LDL-cholesterol on risk of ischemic stroke and coronary heart disease. Circulation. 2012; **126**(Suppl 21): A11959-A.

131. Kent ST, Rosenson RS, Avery CL, Chen YI, Correa A, Cummings SR, et al. PCSK9 Loss-of-Function Variants, Low-Density Lipoprotein Cholesterol, and Risk of Coronary Heart Disease and Stroke: Data From 9 Studies of Blacks and Whites. Circulation Cardiovascular genetics. 2017; **10**(4): e001632.

132. Johannsen TH, Frikke-Schmidt R, Schou J, Nordestgaard BG, Tybjaerg-Hansen A. Genetic inhibition of CETP, ischemic vascular disease and mortality, and possible adverse effects. Journal of the American College of Cardiology. 2012; **60**(20): 2041-8.

133. Cannon CP, Shah S, Dansky HM, Davidson M, Brinton EA, Gotto AM, et al. Safety of Anacetrapib in Patients with or at High Risk for Coronary Heart Disease. N Engl J Med. 2010; **363**(25): 2406-15.

134. Nomura A, Won H-H, Khera AV, Takeuchi F, Ito K, McCarthy S, et al. Protein Truncating Variants at the Cholesteryl Ester Transfer Protein Gene and Risk for Coronary Heart Disease. Circulation research. 2017; **121**(8): 81-8.

135. Ference BA, Kastelein JP, Ginsberg HN, et al. Association of genetic variants related to cetp inhibitors and statins with lipoprotein levels and cardiovascular risk. JAMA. 2017; **318**(10): 947-56.

136. Sniderman AD, Peterson ED. Genetic studies help clarify the complexities of lipid biology and treatment. JAMA. 2017; **318**(10): 915-7.

137. Cholesterol Treatment Trialists' (CTT) Collaborators. Protocol for a prospective collaborative overview of all current and planned randomized trials of cholesterol treatment regimens. Cholesterol Treatment Trialists' (CTT) Collaboration. Am J Cardiol. 1995; **75**(16): 1130-4.

138. Steinberg D. Thematic review series: the pathogenesis of atherosclerosis. An interpretive history of the cholesterol controversy: part I. J Lipid Res. 2004; **45**(9): 1583-93.

139. Nordestgaard BG, Chapman MJ, Humphries SE, Ginsberg HN, Masana L, Descamps OS, et al. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. Eur Heart J. 2013; **34**(45): 3478-90a.

140. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ. 2003; **326**(7404): 1423.

141. The Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) Investigators. High-Dose Atorvastatin after Stroke or Transient Ischemic Attack. N Engl J Med. 2006; **355**(6): 549-59.

142. Canner PL, Berge KG, Wenger NK, Stamler J, Friedman L, Prineas RJ, et al. Fifteen year mortality in Coronary Drug Project patients: Long-term benefit with niacin. Journal of the American College of Cardiology. 1986; **8**(6): 1245-55.

143. The AIM-HIGH Investigators. Niacin in Patients with Low HDL Cholesterol Levels Receiving Intensive Statin Therapy. N Engl J Med. 2011; **365**(24): 2255-67.

144. The HPS2-THRIVE Collaborative Group. Effects of Extended-Release Niacin with Laropiprant in High-Risk Patients. N Engl J Med. 2014; **371**(3): 203-12.

145. The HPS3/TIMI55-REVEAL Collaborative Group. Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. New England Journal of Medicine. 2017.

146. Schwartz GG, Olsson AG, Abt M, Ballantyne CM, Barter PJ, Brumm J, et al. Effects of Dalcetrapib in Patients with a Recent Acute Coronary Syndrome. N Engl J Med. 2012; **367**(22): 2089-99.

147. Lincoff AM, Nicholls SJ, Riesmeyer JS, Barter PJ, Brewer HB, Fox KAA, et al. Evacetrapib and Cardiovascular Outcomes in High-Risk Vascular Disease. N Engl J Med. 2017; **376**(20): 1933-42.

148. Jun M, Foote C, Lv J, Neal B, Patel A, Nicholls SJ, et al. Effects of fibrates on cardiovascular outcomes: a systematic review and meta-analysis. Lancet. 2010; **375**(9729): 1875-84.

149. Horenstein RB, Smith DE, Mosca L. Cholesterol predicts stroke mortality in the Women's Pooling Project. Stroke. 2002; **33**(7): 1863-8.

150. Tanabe N, Iso H, Okada K, Nakamura Y, Harada A, Ohashi Y, et al. Serum total and non-high-density lipoprotein cholesterol and the risk prediction of cardiovascular events - the JALS-ECC. Circulation journal : official journal of the Japanese Circulation Society. 2010; **74**(7): 1346-56.

151. Wang X, Dong Y, Qi X, Huang C, Hou L. Cholesterol levels and risk of hemorrhagic stroke: a systematic review and meta-analysis. Stroke. 2013; **44**(7): 1833-9.

152. Liu Q, Zhao D, Wang W, Liu J, Sun JY. [A prospective study of serum high density lipoprotein cholesterol, cardiovascular and cerebrovascular risk in a Chinese muti-provinces cohort]. Zhonghua nei ke za zhi. 2008; **47**(4): 272-6.

153. Ren J, Zhao D, Liu J, Wang W, Wang M, Sun JY, et al. [Relationship between serum non-high-density lipoprotein cholesterol and incidence of cardiovascular disease]. Zhonghua xin xue guan bing za zhi. 2010; **38**(10): 934-8.

154. Chei CL, Yamagishi K, Kitamura A, Kiyama M, Imano H, Ohira T, et al. High-density lipoprotein subclasses and risk of stroke and its subtypes in Japanese population: the Circulatory Risk in Communities Study. Stroke. 2013; **44**(2): 327-33.

155. Bonaventure A, Kurth T, Pico F, Barberger-Gateau P, Ritchie K, Stapf C, et al. Triglycerides and risk of hemorrhagic stroke vs. ischemic vascular events: The Three-City Study. Atherosclerosis. 2010; **210**(1): 243-8.

156. Chen Z, Lee L, Chen J, Collins R, Wu F, Guo Y, et al. Cohort Profile: The Kadoorie Study of Chronic Disease in China (KSCDC). Int J Epidemiol. 2005; **34**(6): 1243-9.

157. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. 2011; **40**(6): 1652-66.

158. Chen J, Campbell T, Li J, Peto R, Boreham J, Feng Z, et al. Diet, Lifestyle and Mortality in China: A Study of the Characteristics of 65 Chinese Counties. Oxford, England: Oxford University Press; 1990.

159. Murray CJ, Yang G, Qiao X. Adult mortality: levels, patterns and causes. The health of adults in the developing world. 1992: 23-111.

160. Chen J, Peto R, Pan W-H, Liu B, Campbell TC, Boreham J, et al. Geographic Study of Mortality, Biochemistry, Diet and Lifestyle in Rural China. Oxford, England: Oxford University Press; 2006.

161. Farzadfar F, Finucane MM, Danaei G, Pelizzari PM, Cowan MJ, Paciorek CJ, et al. National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3.0 million participants. Lancet. 2011; **377**(9765): 578-86.

162. Li JY, Liu BQ, Li GY, Chen ZJ, Sun XI, Rong SD. Atlas of cancer mortality in the People's Republic of China. An aid for cancer control and research. Int J Epidemiol. 1981; **10**(2): 127-33.

163. He J, Gu D, Wu X, Reynolds K, Duan X, Yao C, et al. Major Causes of Death among Men and Women in China. N Engl J Med. 2005; **353**(11): 1124-34.

164. Zheng W, Chow WH, Yang G, Jin F, Rothman N, Blair A, et al. The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. American Journal of Epidemiology. 2005; **162**(11): 1123-31.

165. Jiang C, Thomas GN, Lam TH, Schooling CM, Zhang W, Lao X, et al. Cohort Profile: The Guangzhou Biobank Cohort Study, a Guangzhou–Hong Kong–Birmingham collaboration. Int J Epidemiol. 2006; **35**(4): 844-52.

166. Peto R. The marked differences between carotenoids and retinoids - methodological implications for biochemical epidemiology. Cancer Surveys. 1983; **2**(2): 327-40.

167. UK Biobank. Protocol for a Large-Scale Prospective Epidemiological Resource. 2006 [cited 03 September 2017]; Available from: http://www.ukbiobank.ac.uk/resources/

168. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PloS Med. 2015; **12**(3): e1001779.

169. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology. 2012; **1**(3): 123-6.

170. The Million Women Study. 1996 [cited 3 September 2017]; Available from: http://www.millionwomenstudy.org/million_women_study_protocols/

171. The Million Women Study: design and characteristics of the study population. Breast Cancer Research. 1999; **1**(1): 73.

172. Tapia-Conyer R, Kuri-Morales P, Alegre-Díaz J, Whitlock G, Emberson J, Clark S, et al. Cohort Profile: The Mexico City Prospective Study. International journal of epidemiology. 2006; **35**(2): 243-9.

173. Elwood JM. Commentary: On representativeness. Int J Epidemiol. 2013; **42**(4): 1014-5.

174. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013; **42**(4): 1012-4.

175. Chen Z, Smith M, Du H, Guo Y, Clarke R, Zheng B, et al. Blood pressure in relation to general and central adiposity among 500 000 adult Chinese men and women. Int J Epidemiol. 2015; **44**(4): 1305-19.

176. China Kadoorie Biobank study collaboration. China Kadoorie Biobank: Standard Operating Procedures. 2016 [cited 23 October 2016]; Available from: http://www.ckbiobank.org/site/Study+Resources/SOPs

177. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. American Journal of Epidemiology. 1999; **150**(4): 341-53.

178. Frost C, Thompson SG. Correcting for Regression Dilution Bias: Comparison of Methods for a Single Predictor Variable. Journal of the Royal Statistical Society Series A (Statistics in Society). 2000; **163**(2): 173-89.

179. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972; **18**(6): 499-502.

180. Yang G, Hu J, Rao KQ, Ma J, Rao C, Lopez AD. Mortality registration and surveillance in China: History, current situation and challenges. Population health metrics. 2005; **3**(1): 3.

181. Yang G, Rao C, Ma J, Wang L, Wan X, Dubrovsky G, et al. Validation of verbal autopsy procedures for adult deaths in China. Int J Epidemiol. 2006; **35**(3): 741-8.

182. Bachorik PS, Ross JW. National Cholesterol Education Program recommendations for measurement of low-density lipoprotein cholesterol: executive summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement. Clin Chem. 1995; **41**(10): 1414-20.

183. Nauck M, Friedrich I, Wieland H. Measurement of LDL and VLDL cholesterol with precipitation techniques. A comparison with the ultracentrifugation method. Clin Lab. 1994; **40**: 167-76.

184. Pisani T, Gebski CP, Leary ET, Warnick GR, Ollington JF. Accurate direct determination of low-density lipoprotein cholesterol using an immunoseparation reagent and enzymatic cholesterol assay. Archives of Pathology and Laboratory Medicine. 1995; **119**(12): 1127-35.

185. Warnick GR, Nauck M, Rifai N. Evolution of methods for measurement of HDL-cholesterol: from ultracentrifugation to homogeneous assays. Clin Chem. 2001; **47**(9): 1579-96.

186. Davidson M, Liu SX, Barter P, Brinton EA, Cannon CP, Gotto AM, et al. Measurement of LDL-C after treatment with the CETP inhibitor anacetrapib. Journal of Lipid Research. 2013; **54**(2): 467-72.

187. Miller WG, Myers GL, Sakurabayashi I, Bachmann LM, Caudill SP, Dziekonski A, et al. Seven direct methods for measuring HDL and LDL cholesterol compared with ultracentrifugation reference measurement procedures. Clin Chem. 2010; **56**(6): 977-86.

188. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine. 2009; **150**(9): 604-12.

189. World Health Organisation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. 2006 [cited 09 December 2016]; Available from: http://www.who.int/diabetes/publications/diagnosis_diabetes2006/en/

190. Bragg F, Li L, Smith M, Guo Y, Chen Y, Millwood I, et al. Associations of blood glucose and prevalent diabetes with risk of cardiovascular disease in 500 000 adult Chinese: the China Kadoorie Biobank. Diabetic Medicine. 2014; **31**(5): 540-51.

191. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and science in sports and exercise. 2000; **32**(9; SUPP/1): S498-S504.

192. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet. 1990; **335**(8692): 765-74.

193. Allison P. Survival analysis using sas: a practical guide: SAS Publishing; 1995.

194. Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. Stat Med. 1991; **10**(7): 1025-35.

195. Chen Z, Peto R, Zhou M, Iona A, Smith M, Yang L, et al. Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. Lancet. 2015; **386**(10002): 1447-56.

196. Alegre-Díaz J, Herrington W, López-Cervantes M, Gnatiuc L, Ramirez R, Hill M, et al. Diabetes and Cause-Specific Mortality in Mexico City. N Engl J Med. 2016; **375**(20): 1961-71.

197. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. American Journal of Epidemiology. 1985; **122**(1): 51-65.

198. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. American Journal of Epidemiology. 1992; **135**(10): 1114-26.

199. Mora S, Buring JE, Ridker PM. Discordance of LDL cholesterol with alternative LDL-related measures and future coronary events. Circulation. 2013; **136**(14): CIRCULATIONAHA. 113.005873.

200. Sniderman AD, Islam S, Yusuf S, McQueen MJ. Discordance analysis of apolipoprotein B and non-high density lipoprotein cholesterol as markers of cardiovascular risk in the INTERHEART study. Atherosclerosis. 2012; **225**(2): 444-9.

201. Thrift AG, Dewey HM, Macdonell RA, McNeil JJ, Donnan GA. Incidence of the major stroke subtypes: initial findings from the North East Melbourne stroke incidence study (NEMESIS). Stroke. 2001; **32**(8): 1732-8.

202. University of Oxford. Nuffield Department of Population Health: Wolfson Epidemiological Laboratory. 2014 [cited 23 October 2016]; Available from: https://www.ctsu.ox.ac.uk/research/laboratories

203. Rothwell PM, Coull AJ, Silver LE, Fairhead JF, Giles MF, Lovelock CE, et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). Lancet. 2005; **366**(9499): 1773-83.

204. National Bureau of Statistics of China. 2010 Population Census. 2012 05/01/2017 [cited 23 August 2017]; Available from: http://www.stats.gov.cn/english/Statisticaldata/CensusData/

205. Zhang LF, Yang J, Hong Z, Yuan GG, Zhou BF, Zhao LC, et al. Proportion of different subtypes of stroke in China. Stroke. 2003; **34**(9): 2091-6.

206. Rothwell PM, Coull AJ, Giles MF, Howard SC, Silver LE, Bull LM, et al. Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). Lancet. 2004; **363**(9425): 1925-33.

207. Morris RW, Whincup PH, Emberson JR, Lampe FC, Walker M, Shaper AG. North-south gradients in Britain for stroke and CHD: are they explained by the same factors? Stroke. 2003; **34**(11): 2604-9.

208. He J, Klag MJ, Wu Z, Whelton PK. Stroke in the People's Republic of China. I. Geographic variations in incidence and risk factors. Stroke. 1995; **26**(12): 2222-7.

209. National Health and Family Planning Commission of the People's Republic of China. Report on Chinese Chronic Diseases and Autritional Status (2015). Beijing People Health Publishing House; 2016.

210. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, Fortmann S, Sans S, Tolonen H, et al. Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations. Lancet. 2000; **355**(9205): 675-87.

211. Zhaosu Wu, Chonghua Yao. Multiprovincial monitoring of trends and determinants in cardiovascular diseases (Sino-MONICA project) - Ii: Results of risk factor monitoring. Chinese Journal of Cardiology. 1997; **25**(4): 255-9.

212. Sun JY, Zhao D, Wang W, Liu J, Li Y, Liu S, et al. [The changing trend of serum total cholesterol in Beijing population aged 25 - 64 years during 1984 - 1999]. Zhonghua Nei Ke Za Zhi. 2006; **45**(12): 980-4.

213. Carroll MD, Kit BK, Lacher DA, Shero ST, Mussolino ME. Trends in lipids and lipoproteins in US adults, 1988-2010. JAMA. 2012; **308**(15): 1545-54.

214. Cohen JD, Cziraky MJ, Cai Q, Wallace A, Wasser T, Crouse JR, et al. 30-year trends in serum lipids among United States adults: results from the National Health and Nutrition Examination Surveys II, III, and 1999-2006. Am J Cardiol. 2010; **106**(7): 969-75.

215. Ferrieres J, Bongard V, Dallongeville J, Arveiler D, Cottel D, Haas B, et al. Trends in plasma lipids, lipoproteins and dyslipidaemias in French adults, 1996-2007. Arch Cardiovasc Dis. 2009; **102**(4): 293-301.

216. Eliasson M, Janlert U, Jansson JH, Stegmayr B. Time trends in population cholesterol levels 1986-2004: influence of lipid-lowering drugs, obesity, smoking and educational level. The northern Sweden MONICA study. J Intern Med. 2006; **260**(6): 551-9.

217. University of North Carolina at Chapel Hill and Chinese Center for Disease Control and Prevention. China Health and Nutrition Survey (2002). [cited 03 September 2016]; Available from: http://www.cpc.unc.edu/projects/china/data

218. Bragg F, Holmes MV, Iona A, et al. Association between diabetes and cause-specific mortality in rural and urban areas of china. JAMA. 2017; **317**(3): 280-9.

219. Bansal S, Buring JE, Rifai N, Mora S, Sacks FM, Ridker PM. Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. Journal of the American Medical Association. 2007; **298**(3): 309-16.

220. Chapman MJ, Ginsberg HN, Amarenco P, Andreotti F, Boren J, Catapano AL, et al. Triglyceride-rich lipoproteins and high-density lipoprotein cholesterol in patients at high risk of cardiovascular disease: evidence and guidance for management. Eur Heart J. 2011; **32**(11): 1345-61.

221. Miller M, Stone NJ, Ballantyne C, Bittner V, Criqui MH, Ginsberg HN, et al. Triglycerides and cardiovascular disease a scientific statement from the American Heart Association. Circulation. 2011; **123**(20): 2292-333.

222. Hegele RA, Ginsberg HN, Chapman MJ, Nordestgaard BG, Kuivenhoven JA, Averna M, et al. The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. The Lancet Diabetes & Endocrinology. 2014; **2**(8): 655-66.

223. Catapano AL, Graham I, De Backer G, Wiklund O, Chapman MJ, Drexel H, et al. 2016 ESC/EAS Guidelines for the Management of Dyslipidaemias. European heart journal. 2016; **37**(39): 2999-3058.

224. Jorgensen AB, Frikke-Schmidt R, West AS, Grande P, Nordestgaard BG, Tybjaerg-Hansen A. Genetically elevated non-fasting triglycerides and calculated remnant cholesterol as causal risk factors for myocardial infarction. Eur Heart J. 2013; **34**(24): 1826-33.

225. Varbo A, Benn M, Tybjærg-Hansen A, Nordestgaard BG. Elevated Remnant Cholesterol Causes Both Low-Grade Inflammation and Ischemic Heart Disease, Whereas Elevated Low-Density Lipoprotein Cholesterol Causes Ischemic Heart Disease Without Inflammation. Circulation. 2013; **128**(12): 1298.

226. Fibrinogen Studies Collaboration. Regression dilution methods for meta-analysis: assessing long-term variability in plasma fibrinogen among 27 247 adults in 15 prospective studies. Int J Epidemiol. 2006; **35**(6): 1570-8.

227. Fibrinogen Studies Collaboration. Correcting for multivariate measurement error by regression calibration in meta-analyses of epidemiological studies. Statistics in medicine. 2009; **28**(7): 1067-92.

228. Rosner B, Spiegelman D, Willett W. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. American Journal of Epidemiology. 1990; **132**(4): 734-45.

229. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. The American Journal of Clinical Nutrition. 1997; **65**(4): 1179S-86S.

230. Rosner B, Willett W, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within‐person measurement error. Statistics in medicine. 1989; **8**(9): 1051-69.

231. Rosner B. Fundamentals of biostatistics: Nelson Education; 2015.

232. van Gijn J, Kerr RS, Rinkel GJE. Subarachnoid haemorrhage. Lancet. 2007; **369**(9558): 306-18.

233. Krishnamurthi RV, Moran AE, Forouzanfar MH, Bennett DA, Mensah GA, Lawes CM, et al. The global burden of hemorrhagic stroke: a summary of findings from the GBD 2010 study. Global Heart. 2014; **9**(1): 101-6.

234. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project--1981-86. 2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral and subarachnoid haemorrhage. Journal of neurology, neurosurgery, and psychiatry. 1990; **53**(1): 16-22.

235. Heuschmann PU, Di Carlo A, Bejot Y, Rastenyte D, Ryglewicz D, Sarti C, et al. Incidence of stroke in Europe at the beginning of the 21st century. Stroke. 2009; **40**(5): 1557-63.

236. Wang W, Jiang B, Wu S, Hong Z, Yang Q, Sander JW, et al. Change in Stroke Incidence from a Population-Based Intervention Trial in Three Urban Communities in China. Neuroepidemiology. 2007; **28**(3): 155-61.

237. Sarti C, Stegmayr B, Tolonen H, Mähönen M, Tuomilehto J, Asplund K. Are Changes in Mortality From Stroke Caused by Changes in Stroke Event Rates or Case Fatality? Stroke. 2003; **34**(8): 1833.

238. Wang Y, Cui L, Ji X, Dong Q, Zeng J, Zhou Y, et al. The China National Stroke Registry for patients with acute cerebrovascular events: design, rationale, and baseline patient characteristics. International Journal of Stroke. 2011; **6**(4): 355-61.

239. Prospective Studies Collaboration. Cholesterol, diastolic blood pressure, and stroke: 13,000 strokes in 450,000 people in 45 prospective cohorts. Prospective studies collaboration. Lancet. 1995; **346**(8991-8992): 1647-53.

240. He J, Gu D, Chen J, Wu X, Kelly TN, Huang JF, et al. Premature deaths attributable to blood pressure in China: a prospective cohort study. Lancet. 2009; **374**(9703): 1765-72.

241. Joffres M, Falaschetti E, Gillespie C, Robitaille C, Loustalot F, Poulter N, et al. Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. BMJ open. 2013; **3**(8).

242. Lewington S, Lacey B, Clarke R, Guo Y, Kong XL, Yang L, et al. The Burden of Hypertension and Associated Risk for Cardiovascular Mortality in China. JAMA internal medicine. 2016; **176**(4): 524-32.

243. Wu Y, Huxley R, Li L, Anna V, Xie G, Yao C, et al. Prevalence, awareness, treatment, and control of hypertension in China: data from the China National Nutrition and Health Survey 2002. Circulation. 2008; **118**(25): 2679-86.

244. Qureshi AI, Ezzeddine MA, Nasar A, Suri MF, Kirmani JF, Hussein HM, et al. Prevalence of elevated blood pressure in 563,704 adult patients with stroke presenting to the ED in the United States. Am J Emerg Med. 2007; **25**(1): 32-8.

245. Gu D, Gupta A, Muntner P, Hu S, Duan X, Chen J, et al. Prevalence of Cardiovascular Disease Risk Factor Clustering Among the Adult Population of China. Circulation. 2005; **112**(5): 658.

246. Lacey B. Blood pressure and stroke pathological types in China: an analysis of 500 000 men and women in the China Kadoorie Biobank study: University of Oxford; 2013.

247. Yano K, Reed DM, MacLean CJ. Serum cholesterol and hemorrhagic stroke in the Honolulu Heart Program. Stroke. 1989; **20**(11): 1460-5.

248. Law MR, Wald NJ, Wu T, Hackshaw A, Bailey A. Systematic underestimation of association between serum cholesterol concentration and ischaemic heart disease in observational studies: data from the BUPA study. BMJ. 1994; **308**(6925): 363-6.

249. Ikram MA, Wieberdink RG, Koudstaal PJ. International Epidemiology of Intracerebral Hemorrhage. Curr Atheroscler Rep. 2012; **14**(4): 300-6.

250. MacKenzie JM. Intracerebral haemorrhage. J Clin Pathol. 1996; **49**(5): 360-4.

251. Cholesterol Treatment Trialists' collaboration. Protocol for analyses of adverse event data from randomized controlled trials of statin therapy. Am Heart J. 2016; **176**: 63-9.

252. Cole FM, Yates PO. The occurrence and significance of intracerebral micro-aneurysms. The Journal of pathology and bacteriology. 1967; **93**(2): 393-411.

253. Anah CO. How does blood-pressure cause stroke? Lancet. 1976; **307**(7970): 1185-6.

254. Ooneda G, Yoshida Y, Suzuki K, Shinkai H, Hori S, Kobori K, et al. Smooth muscle cells in the development of plasmatic arterionecrosis, arteriosclerosis, and arterial contraction. Blood vessels. 1978; **15**(1-3): 148-56.

255. Sever PS, Dahlof B, Poulter NR, Wedel H, Beevers G, Caulfield M, et al. Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial--Lipid Lowering Arm (ASCOT-LLA): a multicentre randomised controlled trial. Lancet. 2003; **361**(9364): 1149-58.

256. Willett WC, Stampfer MJ, Manson JE, Colditz GA, Speizer FE, Rosner BA, et al. Intake of trans fatty acids and risk of coronary heart disease among women. Lancet. 1993; **341**(8845): 581-5.

257. Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, et al. Dietary Fat Intake and the Risk of Coronary Heart Disease in Women. N Engl J Med. 1997; **337**(21): 1491-9.

258. Iso H, Stampfer MJ, Manson JE, Rexrode K, Hu FB, Hennekens CH, et al. Prospective Study of Fat and Protein Intake and Risk of Intraparenchymal Hemorrhage in Women. Circulation. 2001; **103**(6): 856.

259. Okamura T, Kokubo Y, Watanabe M, Higashiyama A, Ono Y, Miyamoto Y, et al. Triglycerides and non-high-density lipoprotein cholesterol and the incidence of cardiovascular disease in an urban Japanese cohort: the Suita study. Atherosclerosis. 2010; **209**(1): 290-4.

260. Sturgeon JD, Folsom AR, Longstreth WT, Jr., Shahar E, Rosamond WD, Cushman M. Risk factors for intracerebral hemorrhage in a pooled prospective study. Stroke. 2007; **38**(10): 2718-25.

261. Kroll ME, Green J, Beral V, Sudlow CL, Brown A, Kirichek O, et al. Adiposity and ischemic and hemorrhagic stroke: Prospective study in women and meta-analysis. Neurology. 2016; **87**(14): 1473-81.

262. Chen Z, Peto R. Personal Communication. 2017.

263. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ : British Medical Journal. 2013; **346**.

264. Wei W-Q, Denny JC. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome medicine. 2015; **7**(1): 41.

265. The Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015; **526**(7571): 68-74.

266. Dahabreh IJ, Kent DM. Index event bias: an explanation for the paradoxes of recurrence risk research. JAMA. 2011; **305**(8): 822-3.

267. Yaghootkar H, Bancks MP, Jones SE, McDaid A, Beaumont R, Donnelly L, et al. Quantifying the extent to which index event biases influence large genetic association studies. Human Molecular Genetics. 2017; **26**(5): 1018-30.

268. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International journal of epidemiology. 2013; **42**(4): 1134-44.

269. Davidson MH, Robinson JG. Safety of aggressive lipid management. Journal of the American College of Cardiology. 2007; **49**(17): 1753-62.

270. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Molecular biology and evolution. 1995; **12**(5): 921-7.

271. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. Nature Protocols. 2011; **6**(2): 121-33.

272. Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. The Annals of Mathematical Statistics. 1940; **11**(3): 284-300.

273. Thomas DC, Lawlor DA, Thompson JR. Re: Estimation of Bias in Nongenetic Observational Studies Using “Mendelian Triangulation” by Bautista et al. Ann Epidemiol. 2007; **17**(7): 511-3.

274. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. Statistical Science. 2010: 22-40.

275. Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental Variable Estimation of Causal Risk Ratios and Causal Odds Ratios in Mendelian Randomization Analyses. American Journal of Epidemiology. 2011; **173**(12): 1392-403.

276. The International Consortium for Blood Pressure Genome-Wide Association S. Genetic Variants in Novel Pathways Influence Blood Pressure and Cardiovascular Disease Risk. Nature. 2011; **478**(7367): 103-9.

277. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet. 2017; **18**(6): 331-44.

278. Dichgans M, Malik R, König IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease – a genome-wide analysis of common variants. Stroke; a journal of cerebral circulation. 2014; **45**(1): 24-36.

279. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng Y-C, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. The Lancet Neurology. 2012; **11**(11): 951-62.

280. Hopewell JC, Malik R, Valdés-Márquez E, Worrall BB, Collins R. Differential effects of PCSK9 variants on risk of coronary disease and ischaemic stroke. Eur Heart J. 2017; **ehx373**(10.1093/eurheartj/ehx373).

281. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. N Engl J Med. 2016; **375**(22): 2144-53.

282. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N Engl J Med. 2017; **376**(18): 1713-22.

283. Zanoni P, Khetarpal SA, Larach DB, Hancock-Cerutti WF, Millar JS, Cuchel M, et al. Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. Science (New York, NY). 2016; **351**(6278): 1166-71.

284. Thompson A, Di Angelantonio E, Sarwar N, Erqou S, Saleheen D, Dullaart RP, et al. Association of cholesteryl ester transfer protein genotypes with CETP mass and activity, lipid levels, and coronary risk. JAMA. 2008; **299**(23): 2777-88.

285. Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NG, Jansen H, et al. Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease. Journal of the American College of Cardiology. 2017; **69**(7): 823-36.

286. Takeuchi F, Isono M, Katsuya T, Yokota M, Yamamoto K, Nabika T, et al. Association of genetic variants influencing lipid levels with coronary artery disease in Japanese individuals. PLoS One. 2012; **7**(9): e46385.

287. Zhong S, Sharp DS, Grove JS, Bruce C, Yano K, Curb JD, et al. Increased coronary heart disease in Japanese-American men with mutation in the cholesteryl ester transfer protein gene despite increased HDL levels. Journal of Clinical Investigation. 1996; **97**(12): 2917.

288. Millwood I, Bennett D. CETP gene variants and risk of vascular and non-vascular diseases in 150,000 Chinese adults (Submitted). JAMA Cardiology; 2017.

289. Anderson CD, Falcone GJ, Phuah CL, Radmanesh F, Brouwers HB, Battey TWK, et al. Genetic variants in CETP increase risk of intracerebral hemorrhage. Annals of neurology. 2016; **80**(5): 730-40.

290. Schwartz CC, VandenBroek JM, Cooper PS. Lipoprotein cholesteryl ester production, transfer, and output in vivo in humans. J Lipid Res. 2004; **45**(9): 1594-607.

291. Hoekstra M, Van Berkel TJ, Van Eck M. Scavenger receptor BI: a multi-purpose player in cholesterol and steroid metabolism. World J Gastroenterol. 2010; **16**(47): 5916-24.

292. Wittrup HH, Tybjaerg-Hansen A, Nordestgaard BG. Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. Circulation. 1999; **99**(22): 2901-7.

293. Do R, Stitziel NO, Won H-H, Jorgensen AB, Duga S, Angelica Merlini P, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature. 2015; **518**(7537): 102-6.

294. The TG and HDL Working Group. Loss-of-Function Mutations in APOC3, Triglycerides, and Coronary Disease. N Engl J Med. 2014; **371**(1): 22-31.

295. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Loss-of-Function Mutations in APOC3 and Risk of Ischemic Vascular Disease. N Engl J Med. 2014; **371**(1): 32-41.

296. Dewey FE, Gusarova V, O'Dushlaine C, Gottesman O, Trejos J, Hunt C, et al. Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. N Engl J Med. 2016; **374**(12): 1123-33.

297. Dewey FE, Gusarova V, Dunbar RL, O'Dushlaine C, Schurmann C, Gottesman O, et al. Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. N Engl J Med. 2017; **377**: 211-21.

298. Graham MJ, Lee RG, Brandt TA, Tai L-J, Fu W, Peralta R, et al. Cardiovascular and Metabolic Effects of ANGPTL3 Antisense Oligonucleotides. N Engl J Med. 2017; **377**: 222-32.

299. Feigin V, Hoorn SV. How to study stroke incidence. Lancet. 2004; **363**(9425): 1920.

300. Ahmad OB, Boschi Pinto C, D. Lopez A, Jl Murray C, Lozano R, Inoue M. Age standardization of rates: A new WHO standard; 2001.

301. Wang Z, Hu S, Sang S, Luo L, Yu C. Age–Period–Cohort Analysis of Stroke Mortality in China: Data From the Global Burden of Disease Study 2013. Stroke. 2017; **48**(2): 271-5.

302. Global Burden of Disease Study 2016. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME); 2016.

303. Critchley J, Liu J, Zhao D, Wei W, Capewell S. Explaining the increase in coronary heart disease mortality in Beijing between 1984 and 1999. Circulation. 2004; **110**(10): 1236-44.

304. Emberson J, Whincup P, Morris R, Walker M, Ebrahim S. Evaluating the impact of population and high-risk strategies for the primary prevention of cardiovascular disease. Eur Heart J. 2004; **25**(6): 484-91.

305. Hankey GJ. Stroke. Lancet. 2017; **389**(10069): 641-54.

306. McQueen MJ, Hawken S, Wang X, Ounpuu S, Sniderman A, Probstfield J, et al. Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study. Lancet. 2008; **372**(9634): 224-33.

307. Houterman S, Verschuren WM, Oomen CM, Boersma-Cobbaert CM, Kromhout D. Trends in total and high density lipoprotein cholesterol and their determinants in The Netherlands between 1993 and 1997. Int J Epidemiol. 2001; **30**(5): 1063-70.

308. Kotseva K, Wood D, De Backer G, De Bacquer D, Pyörälä K, Keil U. Cardiovascular prevention guidelines in daily practice: a comparison of EUROASPIRE I, II, and III surveys in eight European countries. Lancet. 2009; **373**(9667): 929-40.

309. Tolonen H, Keil U, Ferrario M, Evans A. Prevalence, awareness and treatment of hypercholesterolaemia in 32 populations: results from the WHO MONICA Project. Int J Epidemiol. 2005; **34**(1): 181-92.

310. Ference BA. How to use Mendelian randomization to anticipate the results of randomized trials. European heart journal. 2017: ehx462-ehx.

311. Athyros VG, Papageorgiou AA, Mercouris BR, Athyrou VV, Symeonidis AN, Basayannis EO, et al. Treatment with atorvastatin to the National Cholesterol Educational Program goal versus 'usual' care in secondary coronary heart disease prevention. The GREek Atorvastatin and Coronary-heart-disease Evaluation (GREACE) study. Current medical research and opinion. 2002; **18**(4): 220-8.

312. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat Genet. 2013; **45**(11): 1345-52.

313. Helgadottir A, Gretarsdottir S, Thorleifsson G, Hjartarson E, Sigurdsson A, Magnusdottir A, et al. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. Nat Genet. 2016; **48**(6): 634-9.

314. Rosner B, Spiegelman D, Willett W. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. American Journal of Epidemiology. 1992; **136**(11): 1400-13.

315. Suh I, Jee SH, Kim HC, Nam CM, Kim IS, Appel LJ. Low serum cholesterol and haemorrhagic stroke in men: Korea Medical Insurance Corporation Study. Lancet. 2001; **357**(9260): 922-5.

316. Cholesterol Treatment Trialists' collaboration. Cholesterol Treatment Trialists' collaboration. [cited 23 September 2017]; Available from: https://www.cttcollaboration.org/

317. Stone NJ, Robinson J, Lichtenstein AH, Merz CNB, Blum CB, Eckel RH, et al. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013.

318. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren WM, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012) The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). European heart journal. 2012; **33**(17): 2126.

319. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. Blood pressure, cholesterol, and stroke in eastern Asia. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. Lancet. 1998; **352**(9143): 1801-7.

320. Lacey B, Lewington S, Clarke R. Personal communications. 2017.

321. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood

pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990; **335**(8693): 827-38.

322. Collins R, MacMahon S. Blood pressure, antihypertensive drug treatment and the risks of stroke and of coronary heart disease. British medical bulletin. 1994; **50**(2): 272-98.

323. Gueyffier F, Boutitie F, Boissel JP, Pocock S, Coope J, Cutler J, et al. Effect of antihypertensive drug treatment on cardiovascular outcomes in women and men. A meta-analysis of individual patient data from randomized, controlled trials. The INDANA Investigators. Ann Intern Med. 1997; **126**(10): 761-7.

324. Neal B, MacMahon S, Chapman N. Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. Blood Pressure Lowering Treatment Trialists' Collaboration. Lancet. 2000; **356**(9246): 1955-64.

325. The CHARGE Consortium. Cohorts for Heart and Aging Research in Genomic Epidemiology. [cited 23 September 2017]; Available from: http://www.chargeconsortium.com/

326. The International Stroke Genetics Consortium. The International Stroke Genetics Consortium. [cited 23 September 2017]; Available from: http://www.strokegenetics.org/