Improving the Reporting of Pharmacogenetic Studies to Facilitate Evidence Synthesis: Anti-Tuberculosis Drug-Related Toxicity as an Example

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Martha Helen Chaplin

March 2021

Abstract

Title: Improving the Reporting of Pharmacogenetic Studies to Facilitate Evidence Synthesis: Anti-Tuberculosis Drug-Related Toxicity as an Example

Author: Martha Chaplin

Background

In pharmacogenetic studies, researchers explore how genetic variants impact individuals' responses to drugs. Implementation of pharmacogenetic tests in clinical practice can improve treatment efficacy and reduce toxicity. For health service providers to implement pharmacogenetic testing in clinical practice, the pharmacogenetic association of interest must be supported by strong evidence. Performing meta-analyses of pharmacogenetic studies increases sample size and power, and is therefore an indispensable tool to researchers striving to improve the strength of evidence for pharmacogenetic associations. The aim of this thesis is to identify and resolve challenges that reviewers might encounter when synthesising evidence from primary pharmacogenetic studies.

Methods

We explored methods of evidence synthesis for pharmacogenetic studies and applied them to undertake a systematic review and meta-analysis of associations between genetic variants and anti-tuberculosis drug-related toxicity. We applied both standard methods of meta-analysis, and more complex methods of meta-analysis that account for correlation between related effect sizes for each genetic variant. Conducting this systematic review and meta-analysis enabled us to identify that key information was often poorly reported in the primary pharmacogenetic studies. In order to improve the reporting of pharmacogenetic studies with a view to facilitating the evidence synthesis process, we used consensus methodology to develop a reporting guideline for pharmacogenetic studies, known as the STROPS (Strengthening The Reporting Of Pharmacogenetic Studies) guideline.

Results

Our systematic review of the association between genetic variants and anti-tuberculosis drug-related toxicity included 70 studies. Slow acetylators are more likely to experience anti-TB drug-induced hepatotoxicity than intermediate/rapid acetylators. We also observed associations between the *CYP2E1 Rsa*I and *GSTM1* null polymorphisms and hepatotoxicity. Key information, such as the ethnicity of included patients, methodological quality, and patient cohort overlap, was poorly reported. We also found that improvements in the reporting of outcome data would give systematic reviewers greater freedom in terms of their analysis approach. As part of the development of the STROPS guideline, 52 individuals from key stakeholder groups participated in two rounds of a Delphi survey. A total of eight individuals participated in a consensus meeting, before the 54-item STROPS guideline was finalised.

Conclusions

Our systematic review showed that pharmacogenetic testing may be useful in clinical practice in terms of risk stratification for hepatotoxicity during TB treatment. More studies are needed to overcome methodological limitations of the existing studies and to assess the feasibility and cost-effectiveness of a stratified medicine approach. It is currently challenging to synthesise pharmacogenetic evidence, due to poor reporting of primary studies. We encourage authors to adhere to the STROPS guideline when publishing pharmacogenetic studies. The STROPS guideline will not only improve the transparency of reporting of pharmacogenetic studies, but will also facilitate the conduct of high-quality systematic reviews and meta-analyses, and thus improve the power to detect pharmacogenetic associations.

Acknowledgements

Firstly, I would like to offer my thanks to my supervisors, Professor Andrea Jorgensen, Professor Jamie Kirkham and Dr Kerry Dwan. I am lucky to have had such an excellent supervisory team, who have been so generous with their time and expertise throughout this project. I would also like to recognise the contributions of Dr Derek Sloan and Dr Gerry Davies; this work would not have been possible without their valuable clinical input.

My studies have been supported partly by Liverpool Reviews and Implementation Group (LRiG), based on funding from the NIHR HTA programme, and partly by the Research, Evidence and Development Initiative (READ-It) project[a], which is funded by UK aid from the UK government; my thanks go to these funders. I am very grateful to Professor Rumona Dickson, Dr Angela Boland and Dr Rui Duarte at LRiG, and Professor Paul Garner at CIDG, for their continuous support of my professional development. Furthermore, my thanks go to all my colleagues at LRIG and at the Cochrane Infectious Diseases Group, who have provided me with a wonderfully supportive working environment for the last 7 years. It is a privilege to work with such friendly, funny and intelligent individuals on a daily basis.

I would like to thank Eleanor Kotas, for her assistance in developing the search strategies used in this thesis, and all individuals that contributed to the development of the STROPS reporting guideline by completing the Delphi survey and/or attending the consensus meeting.

As ever, my family are a constant source of love and humour, and I would be lost without them. In particular, I will never be able to find the words to thank my parents for everything they have done for me. With my dad's love of spreadsheets, and my mum's love of all things Maths, my choice of career is probably no great surprise to anyone.

Finally, I would like to thank Joe, who has been there for me throughout every part of this journey, and is my biggest cheerleader. He knows how to put a smile on my face after even the most stressful of days; whether that is with the most delicious cooking, a glass of wine, or a Friday night at home listening to our favourite music (or more commonly all three).

[a] READ-It (project number 300342-104) is funded by UK aid from the UK government; however, the views expressed do not necessarily reflect the UK government's official policies.

List of abbreviations

ADR	Adverse drug reaction
ALP	Alkaline phosphatise
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATD	Anti-tuberculosis drug
ATDH	Anti-tuberculosis drug-induced hepatotoxicity
ATT	Anti-tuberculosis treatment
CI	Confidence interval
CIOMS	Council for International Organizations of Medical Science
COMET	Core Outcome Measures in Effectiveness Trials
DIH	Drug-induced hepatotoxicity
DILI	Drug-induced liver injury
DOTS	Directly observed treatment, short-course
E+E	Explanation and elaboration
EMB	Ethambutol
EQUATOR	Enhancing the QUAlity and Transparency Of health Research
FDC	Fixed-dose combination
GI	Group identifier
GWAS	Genome-wide association study
Het	Heterozygous
Hom	Homozygous
HWE	Hardy-Weinberg equilibrium
IA	Intermediate acetylator
IgM	Immunoglobulin M
INH	Isoniazid
IQR	Interquartile range
MPE	Maculopapular eruption
MT	Mutant-type
N/A	Not applicable
NR	Not reported
OR	Odds ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PZA	Pyrazinamide
RCT	Randomised controlled trial
RA	Rapid acetylators
RIF	Rifampicin
SA	Slow acetylators
SM	Streptomycin
SNP	Single nucleotide polymorphism
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology
STROPS	STrengthening the Reporting Of Pharmacogenetic Studies
STREGA	STrengthening the REporting of Genetic Association studies
TB	Tuberculosis
ULN	Upper limit of normal
WHO	World Health Organization
WT	Wild-type

Initials of researchers involved in the work presented in this thesis

MC: Martha Chaplin, ALJ: Andrea L Jorgensen; JK: Jamie Kirkham; KD: Kerry Dwan

Publications

Papers

- Richardson M, Kirkham J, Dwan K, Sloan DJ, Davies G, Jorgensen AL. *NAT2* variants and toxicity related to anti-tuberculosis agents: a systematic review and meta-analysis. The *Int J Tuberc Lung Dis* 2019 Mar 1;23(3):293-305.
- Richardson M, Kirkham J, Dwan K, Sloan DJ, Davies G, Jorgensen AL. *CYP* genetic variants and toxicity related to anti-tubercular agents: a systematic review and meta-analysis. *Syst Rev* 2018;7(1):204. Doi:10.1186/s13643-018-0861-z.
- Chaplin M, Kirkham JJ, Dwan K, Sloan DJ, Davies G, Jorgensen AL. STrengthening the Reporting Of Pharmacogenetic Studies: Development of the STROPS guideline. *PLoS Med* 2020;17(9):e1003344. Doi:10.1371/journal.pmed.1003344

Protocols

- Richardson M, Kirkham J, Dwan K, Sloan D, Davies G, Jorgensen A. Influence of genetic variants on toxicity to anti-tubercular agents: a systematic review and meta-analysis (protocol). *Syst Rev* 2017;6(1):142. Doi:10.1186/s13643-017-0533-4.
- Richardson M, Kirkham JJ, Dwan KM, Sloan DJ, Davies G, Jorgensen A. Protocol for the development of the STrengthening the Reporting Of Pharmacogenetic Studies (STROPS) guideline: checklist of items for reporting pharmacogenetic studies. *BMJ Open* 2019;9(7):e030212. Doi:10.1136/bmjopen-2019-030212

Conference abstracts

- Richardson M, Kirkham J, Dwan K, Sloan D, Davies G, Jorgensen A. *NAT2* genetic variants and toxicity related to anti-tubercular agents: A systematic review and meta-analysis. In*:* ASHG 2018. *American Society of Human Genetics 68th Annual Meeting - Poster Abstracts. San Diego, October 16-20, 2018*. Available from: https://www.ashg.org/wp-content/uploads/2019/10/2018-poster-abstracts.pdf.
- Richardson M, Kirkham J, Dwan K, Sloan D, Davies G, Jorgensen A. Development of reporting guidelines for pharmacogenetic studies to facilitate evidence synthesis. *Genet Epidemiol* 2018;42(7):727.

Table of contents

Abstract i
Acknowledgements ii
List of abbreviations iii
Publications iv
Table of contents v
List of tables viii
List of figures xi
1 Introduction 1
1.1 Introduction to pharmacogenetics 1
1.2 The human genome and DNA 2
1.3 Genetic variation 2
1.4 SNPs and drug response 3
1.5 Pharmacogenetic association studies 4
1.6 Clinical implications of pharmacogenetic studies 6
1.7 How systematic reviews and meta-analysis can improve the strength of the evidence base 8
1.8 Thesis aim and structure 11
2 Methods of evidence synthesis for pharmacogenetic studies 12
2.1 Searching for studies 12
2.2 Assessment of methodological quality of pharmacogenetic studies 13
2.3 Analysis methods 22
2.4 Assessment of publication bias 25
2.5 Discussion 25
3 Influence of genetic variants on toxicity related to anti-tuberculosis drugs: a systematic review 28
3.1 Background 28
3.2 Methods 31
3.3 Results 34
3.4 Discussion 65
3.5 Conclusion 71
4 Influence of genetic variants on toxicity related to anti-tuberculosis drugs: meta-analyses 72
4.1 Methods 72

4.2 Results 75
4.3 Discussion 120
4.4 Conclusion 125
5 Further analyses of the association between genetic variants of the *NAT2* data and hepatotoxicity 126
5.1 Introduction 126
5.2 Genetic model-free and bivariate meta-analysis approaches 127
5.3 Methods 130
5.4 Results 133
5.5 Discussion 138
5.6 Conclusion 141
6 Development of the STROPS (STrengthening the Reporting Of Pharmacogenetic Studies) guideline 142
6.1 Introduction 142
6.2 Methods 143
6.3 Results 148
6.4 Discussion 161
6.5 Conclusion 162
7 Conclusions and further work 164
7.1 Introduction 164
7.2 Summary of main findings 164
7.3 Implications for practice and research 167
7.4 Future work 170
7.5 Concluding remarks 172
References 174
Appendices 188
Appendix 1. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Search strategy 188
Appendix 2. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Full quality assessment results 189
Appendix 3. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Results of the sensitivity analyses 217
Appendix 4. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Forest plots for the secondary analyses 231

Appendix 5. Further analyses of the association between *NAT2* genetic variants and hepatotoxicity: Investigating the assumption of constant λ 274
Appendix 6. Further analyses of the association between *NAT2* genetic variants and hepatotoxicity: Results of the sensitivity analyses 290
Appendix 7. Development of the STROPS guideline: Reporting items scored in the Delphi survey 291
Appendix 8. Development of the STROPS guideline: Consensus matrix 302
Appendix 9: Development of the STROPS guideline: Explanation and elaboration document 310
Glossary of terms 342

List of tables

Table 1 Study designs used for pharmacogenetic association studies 5
Table 2 Examples of different modes of inheritance 19
Table 3 Key characteristics of studies included in the systematic review 35
Table 4 Choosing which genes and SNPs to genotype: quality assessment summary 53
Table 5 Sample size: quality assessment summary 54
Table 6 Study design: quality assessment summary 54
Table 7 Reliability of genotypes: quality assessment summary 55
Table 8 Missing genotype data: quality assessment summary 56
Table 9 Population stratification: quality assessment summary 57
Table 10 Hardy-Weinberg equilibrium: quality assessment summary 57
Table 11 Mode of inheritance: quality assessment summary 58
Table 12 Choice and definition of outcomes: quality assessment summary 59
Table 13 Definitions of hepatotoxicity in the included studies 60
Table 14 Definitions of other toxicity outcomes in the included studies 64
Table 15 Treatment adherence: quality assessment summary 65
Table 16 Results of the secondary analyses: NAT2 SNPs and hepatotoxicity 94
Table 17 Results of the secondary analyses: NAT2 genetic variants and other toxicity outcomes 100
Table 18 Results of the secondary analyses: CYP genetic variants and hepatotoxicity 102
Table 19 Results of the secondary analyses: CYP genetic variants and other toxicity outcomes 106
Table 20 Results of the secondary analyses: other genetic variants and hepatotoxicity 108
Table 21 Results of the secondary analyses: association between GST and other genetic variants and toxicity outcomes (other than hepatotoxicity) 119
Table 22 Contingency table to be completed for each study 132
Table 23 Association between NAT2 acetylator status and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 134
Table 24 Association between NAT2 282C-T and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 135
Table 25 Association between NAT2 341T-C and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 136
Table 26 Association between NAT2 481C-T and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 136

Table 27 Association between NAT2 590G-A and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 137
Table 28 Association between NAT2 803A-G and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 137
Table 29 Association between NAT2 857G-A and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons 138
Table 30 Summary of decisions made at consensus meeting 152
Table 31 STROPS guideline 159
Table 32 Full quality assessment results: choosing which genes and SNPs to genotype 189
Table 33 Full quality assessment results: sample size 191
Table 34 Full quality assessment results: study design 193
Table 35 Full quality assessment results: reliability of genotypes 195
Table 36 Full quality assessment results: missing genotype data 201
Table 37 Full quality assessment results: population stratification 205
Table 38 Full quality assessment results: Hardy-Weinberg equilibrium 207
Table 39 Full quality assessment results: mode of inheritance 211
Table 40 Full quality assessment results: choice and definition of outcomes 213
Table 41 Full quality assessment results: treatment adherence 215
Table 42 GSTM1 null polymorphism and hepatotoxicity: results of pairwise comparisons 228
Table 43 Labels for each study included in Figure 83 and corresponding λ 275
Table 44 Labels for each study included in Figure 84 and corresponding limits of the bootstrapped 95% CI for λ 276
Table 45 NAT2 282C-T: λ for each study 278
Table 46 NAT2 282C-T: Limits of the bootstrapped 95% CI for λ for each study 278
Table 47 NAT2 341T-C: λ for each study 280
Table 48 NAT2 341T-C: Limits of the bootstrapped 95% CI for λ for each study 280
Table 49 NAT2 481C-T: λ for each study 281
Table 50 NAT2 481C-T: Limits of the bootstrapped 95% CI for λ for each study 282
Table 51 NAT2 590G-A: λ for each study 284
Table 52 NAT2 590G-A: Limits of the bootstrapped 95% CI for λ for each study 285
Table 53 NAT2 803A-G: λ for each study 286
Table 54 NAT2 803A-G: Limits of the bootstrapped 95% CI for λ for each study 287
Table 55 NAT2 857G-A: λ for each study 288
Table 56 NAT2 857G-A: Limits of the bootstrapped 95% CI for λ for each study 289

Table 57 Association between NAT2 282C-T and hepatotoxicity: sensitivity analysis excluding Santos 2013 290
Table 58 Association between NAT2 341T-C and hepatotoxicity: sensitivity analysis excluding Lee 2010 290
Table 59 Association between NAT2 590G-A and hepatotoxicity: sensitivity analysis excluding Xiang 2014 and Santos 2013 290
Table 60 Reporting items scored at Round 1 and Round 2 of the Delphi survey 291
Table 61 Additional reporting items suggested by Delphi participants, scored in Round 2 of the Delphi survey only 301

List of figures

Figure 1 PRISMA flowchart showing inclusion and exclusion of studies in the systematic review..... 34
Figure 2 NAT2 acetylator status and hepatotoxicity; slow/intermediate versus rapid acetylator status 76
Figure 3 NAT2 acetylator status and hepatotoxicity sensitivity analysis: slow versus rapid/intermediate acetylator status 78
Figure 4 Summary of the results obtained by applying different approaches to the analysis of NAT2 acetylator status and hepatotoxicity 79
Figure 5 Funnel plot for the analysis of NAT2 slow/intermediate versus rapid acetylator status and hepatotoxicity 80
Figure 6 CYP2E1 RsaI polymorphism and hepatotoxicity: homozygous mutant-type (TT) or heterozygous (CT) versus homozygous wild-type (CC)..... 81
Figure 7 Summary of the results obtained by applying different approaches to the analysis of CYP2E1 RsaI polymorphism and hepatotoxicity 82
Figure 8 CYP2E1 DraI polymorphism and hepatotoxicity: homozygous mutant-type (AA) or heterozygous (AT) versus homozygous wild-type (TT) 83
Figure 9 Summary of the results obtained by applying different approaches to the analysis of CYP2E1 DraI polymorphism and hepatotoxicity 84
Figure 10 CYP2E1 PstI polymorphism and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (GC) versus homozygous wild-type (GG) 85
Figure 11 Summary of the results obtained by applying different approaches to the analysis of CYP2E1 PstI polymorphism and hepatotoxicity..... 86
Figure 12 Funnel plot for the analysis of CYP2E1 RsaI polymorphism and hepatotoxicity: homozygous mutant-type (TT) or heterozygous (CT) versus homozygous wild-type (CC).... 86
Figure 13 Funnel plot for the analysis of CYP2E1 DraI polymorphism and hepatotoxicity: homozygous mutant-type (AA) or heterozygous (AT) versus homozygous wild-type (TT) ... 87
Figure 14 Funnel plot for the analysis of CYP2E1 PstI polymorphism and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (GC) versus homozygous wild-type (GG).. 87
Figure 15 GSTM1 null polymorphism and hepatotoxicity: homozygous null versus heterozygous or homozygous present 88
Figure 16 Summary of the results obtained by applying different approaches to the analysis of GSTM1 null polymorphism and hepatotoxicity 89

Figure 17 GSTT1 null polymorphism and hepatotoxicity: homozygous null versus heterozygous or homozygous present 90
Figure 18 Summary of the results obtained by applying different approaches to the analysis of GSTT1 null polymorphism and hepatotoxicity 91
Figure 19 Funnel plot for the analysis of GSTM1 null polymorphism and hepatotoxicity: homozygous null versus heterozygous or homozygous present 91
Figure 20 Funnel plot for the analysis of GSTT1 null polymorphism and hepatotoxicity: homozygous null versus heterozygous or homozygous present 92
Figure 21 Summary of the results obtained by applying different approaches to the analysis of NAT2 acetylator status and hepatotoxicity 134
Figure 22 Attrition bias: All stakeholder groups 149
Figure 23 Attrition bias: Journal editors 150
Figure 24 Attrition bias: Primary researchers 150
Figure 25 Attrition bias: Systematic reviewers 151
Figure 26 NAT2 acetylator status and hepatotoxicity: slow versus rapid acetylator status 217
Figure 27 NAT2 acetylator status and hepatotoxicity: intermediate versus rapid acetylator status 218
Figure 28 CYP2E1 RsaI polymorphism and hepatotoxicity: homozygous mutant-type (TT) or heterozygous (CT) versus homozygous wild-type (CC), sensitivity analysis excluding studies where genotypes deviated from HWE (Rana 2014, Singla 2014 and Sotsuka 2011) 219
Figure 29 CYP2E1 RsaI polymorphism and hepatotoxicity: heterozygous (CT) versus homozygous wild-type (CC) 220
Figure 30 CYP2E1 RsaI polymorphism and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 221
Figure 31 CYP2E1 DraI polymorphism and hepatotoxicity: homozygous mutant-type (AA) or heterozygous (TA) versus homozygous wild-type (TT), sensitivity analysis excluding two studies where genotypes deviated from HWE (Bose 2011 and Gupta 2013b) 222
Figure 32 CYP2E1 DraI polymorphism and hepatotoxicity: heterozygous (TA) versus homozygous wild-type (TT) 223
Figure 33 CYP2E1 DraI polymorphism and hepatotoxicity: homozygous mutant-type (AA) versus homozygous wild-type (TT) 224
Figure 34 CYP2E1 PstI polymorphism and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (GC) versus homozygous wild-type (GG), sensitivity analysis excluding one study where genotypes deviated from HWE (Chamorro 2013) 225

Figure 35 CYP2E1 PstI polymorphism and hepatotoxicity: heterozygous (GC) versus homozygous wild-type (GG) 226
Figure 36 CYP2E1 PstI polymorphism and hepatotoxicity: homozygous mutant-type (CC) versus homozygous wild-type (GG) 227
Figure 37 GSTM1 null polymorphism and hepatotoxicity: homozygous null versus heterozygous or homozygous present, sensitivity analysis excluding one study where genotypes deviated from HWE (Liu 2014) 228
Figure 38 GSTT1 null polymorphism and hepatotoxicity: heterozygous versus homozygous present 229
Figure 39 GSTT1 null polymorphism and hepatotoxicity: homozygous null versus homozygous present 230
Figure 40 NAT2 282C-T and hepatotoxicity: heterozygous (CT) versus homozygous wild-type (CC) 231
Figure 41 NAT2 282C-T and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 232
Figure 42 NAT2 341T-C and hepatotoxicity: heterozygous (TC) versus homozygous wild-type (TT) 233
Figure 43 NAT2 341T-C and hepatotoxicity: homozygous mutant-type (CC) versus homozygous wild-type (TT) 234
Figure 44 NAT2 481C-T and hepatotoxicity: heterozygous (CT) versus homozygous wild-type (CC) 235
Figure 45 NAT2 481C-T and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 236
Figure 46 NAT2 590G-A and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (GG) 237
Figure 47 NAT2 590G-A and hepatotoxicity: homozygous mutant-type (AA) versus homozygous wild-type (GG) 238
Figure 48 NAT2 803A-G and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (AA) 239
Figure 49 NAT2 803A-G and hepatotoxicity: homozygous mutant-type (GG) versus homozygous wild-type (AA) 240
Figure 50 NAT2 857G-A and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (GG) 241
Figure 51 NAT2 857G-A and hepatotoxicity: homozygous mutant-type (AA) versus homozygous wild-type (GG) 242

Figure 52 CYP2E1 96-bp SNP and hepatotoxicity: heterozygous (DI) versus homozygous wild-type (DD) 243
Figure 53 CYP2E1 96-bp SNP and hepatotoxicity: homozygous mutant-type (II) versus homozygous wild-type (DD) 244
Figure 54 CYP2C9 rs4918758 and hepatotoxicity: heterozygous (TC) versus homozygous wild-type (TT) 245
Figure 55 CYP2C9 rs4918758 and hepatotoxicity: homozygous mutant-type (CC) versus homozygous wild-type (TT) 246
Figure 56 CYP2B6 rs3745274 and hepatotoxicity: heterozygous (GT) versus homozygous wild-type (GG) 247
Figure 57 CYP2B6 rs3745274 and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (GG) 248
Figure 58 ABCB1 rs1045642 and hepatotoxicity: heterozygous (CT) versus homozygous wild-type (CC) 249
Figure 59 ABCB1 rs1045642 and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 250
Figure 60 NQO1 609C-T (rs1800566) and hepatotoxicity: heterozygous (CT) versus homozygous wild-type (CC) 251
Figure 61 NQO1 609C-T (rs1800566) and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 252
Figure 62 PXR rs3814055 and hepatotoxicity: heterozygous (CT) versus homozygous wild-type (CC) 253
Figure 63 PXR rs3814055 and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 254
Figure 64 PXR rs12488820 and hepatotoxicity: homozygous mutant-type (TT) versus homozygous wild-type (CC) 255
Figure 65 PXR rs2461823 and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (GG) 256
Figure 66 PXR rs2461823 and hepatotoxicity: homozygous mutant-type (AA) versus homozygous wild-type (GG) 257
Figure 67 PXR rs7643645 and hepatotoxicity: heterozygous (AG) versus homozygous wild-type (AA) 258
Figure 68 PXR rs7643645 and hepatotoxicity: homozygous mutant-type (GG) versus homozygous wild-type (AA) 259

Figure 69 PXR rs6785049 and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (GG) 260
Figure 70 PXR rs6785049 and hepatotoxicity: homozygous mutant-type (AA) versus homozygous wild-type (GG) 261
Figure 71 PXR rs3814057 and hepatotoxicity: heterozygous (AC) versus homozygous wild-type (AA) 262
Figure 72 PXR rs3814057 and hepatotoxicity: homozygous mutant-type (CC) versus homozygous wild-type (AA) 263
Figure 73 SLCO1B1 rs4149013 and hepatotoxicity: heterozygous (AG) versus homozygous wild-type (AA) 264
Figure 74 SLCO1B1 rs4149013 and hepatotoxicity: homozygous mutant-type (GG) versus homozygous wild-type (AA) 265
Figure 75 SLCO1B1 rs4149014 and hepatotoxicity: heterozygous (GT) versus homozygous wild-type (TT) 266
Figure 76 SLCO1B1 rs4149014 and hepatotoxicity: homozygous mutant-type (GG) versus homozygous wild-type (TT) 267
Figure 77 SLCO1B1 rs2306283 and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (GG) 268
Figure 78 SLCO1B1 rs2306283 and hepatotoxicity: homozygous mutant-type (AA) versus homozygous wild-type (GG) 269
Figure 79 SLCO1B1 rs4149056 and hepatotoxicity: heterozygous (TC) versus homozygous wild-type (TT) 270
Figure 80 SLCO1B1 rs4149056 and hepatotoxicity: homozygous mutant-type (CC) versus homozygous wild-type (TT) 271
Figure 81 SOD2 rs4880 and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (CT) versus homozygous wild-type (TT) 272
Figure 82 UGT1A1 rs4148323 and hepatotoxicity: heterozygous (GA) versus homozygous wild-type (GG) 273
Figure 83 NAT2 acetylator status: Graph showing the log OR for intermediate versus rapid acetylators against the log OR for slow versus rapid acetylators for each study 274
Figure 84 NAT2 acetylator status: Graph showing the bootstrapped 95% CIs for λ for each study 275
Figure 85 NAT2 282C-T: Graph showing the log OR for CT versus CC against the log OR for TT versus CC for each study (all studies) 277

Figure 86 NAT2 282C-T: Graph showing the log OR for CT versus CC against the log OR for TT versus CC for each study (sensitivity analysis excluding Santos 2013 due to deviation from HWE)..... 277
Figure 87 NAT2 282C-T: Graph showing the bootstrapped 95% CIs for λ for each study ... 278
Figure 88 NAT2 341T-C: Graph showing the log OR for TC versus TT against the log OR for CC versus TT for each study (all studies) 279
Figure 89 NAT2 341T-C: Graph showing the log OR for TC versus TT against the log OR for CC versus TT for each study (sensitivity analysis excluding Lee 2010 due to deviation from HWE)..... 279
Figure 90 NAT2 341T-C: Graph showing the bootstrapped 95% CIs for λ for each study ... 280
Figure 91 NAT2 481C-T: Graph showing the log OR for CT versus CC against the log OR for TT versus CC for each study 281
Figure 92 NAT2 481C-T: Graph showing the bootstrapped 95% CIs for λ for each study ... 282
Figure 93 NAT2 590G-A: Graph showing the log OR for AG versus GG against the log OR for AA versus GG for each study (all studies)..... 283
Figure 94 NAT2 590G-A: Graph showing the log OR for AG versus GG against the log OR for AA versus GG for each study (sensitivity analysis excluding Xiang 2014 and Santos 2013 (due to deviation from HWE) 283
Figure 95 NAT2 590G-A: Graph showing the bootstrapped 95% CIs for λ for each study... 284
Figure 96 NAT2 803A-G: Graph showing the log OR for GA versus AA against the log OR for GG versus AA for each study 286
Figure 97 NAT2 803A-G: Graph showing the bootstrapped 95% CIs for λ for each study... 287
Figure 98 NAT2 857G-A: Graph showing the log OR for AG versus GG against the log OR for AA versus GG acetylators for each study 288
Figure 99 NAT2 857G-A: Graph showing the bootstrapped 95% CIs for λ for each study... 289

1 Introduction

1.1 Introduction to pharmacogenetics

Pharmacogenetics has been defined as the study of how genetic differences influence the variability in patient's responses to drugs.[1] Patient response to a particular drug may be defined in terms of both benefits (therapeutic effect) and harms (*adverse effects*[b]). For example, in a study conducted by Trivedi et al.,[2] only approximately 30% of patients receiving the antidepressant drug, citalopram, for major depressive disorder achieved an adequate treatment response, i.e. full remission. However, 4% of patients from the Trivedi et al. study[2] experienced at least one serious adverse event.

Variability in patient response to any given drug is likely to be highly complex and may be attributable to a number of physiological and environmental factors, such as age, sex, smoking status, and levels of physical activity. A lack of adherence to the recommended dose and interactions with concurrent medications may also impact treatment response in terms of both efficacy and toxicity. Furthermore, it is widely accepted that genetic variation can also contribute to the variation in drug response observed between different individuals.

The concept that genetic variants might account for variability in drug responses was first proposed by Garrod in 1909,[3] and the term 'pharmacogenetics' was introduced by Vogel in 1959.[4] As the study of pharmacogenetics allows researchers to explore reasons for inter-individual variability in drug response, pharmacogenetics is viewed as an important field which may lead to improvements in drug therapy and future prescribing practices.[5]

Another term, 'pharmacogenomics', is also used throughout the literature. PharmGKB (www.pharmgkb.org), a pharmacogenomics knowledge resource,[6] advises that 'pharmacogenetics usually refers to how variation in one single *gene* influences the response to a single drug', and that 'pharmacogenomics is a broader term, which studies how all of the genes (the *genome*) can influence responses to drugs'. It is also stated that these terms are often used interchangeably.[7] Throughout this thesis, the term 'pharmacogenetics' is used, since the focus of the work is on studies which aim to identify specific genetic variants that cause individuals to respond to drugs differently.

[b] Definitions are provided for terms in italics in the glossary

1.2 The human genome and DNA

The human genome contains the information required to produce all the *proteins* needed by a human being, and is inscribed in a material known as *DNA*. DNA is built from molecules called '*nucleotides*', each of which contains one of four bases: adenine (A), thymine (T), guanine (G) and cytosine (C). The order in which these bases are arranged determines which proteins are produced by specifying a sequence of *amino acids*.

In the human body, almost every cell has a nucleus, which serves as the cell's control centre and directs the cell's activities. Each nucleus contains a full copy of the individual's DNA, which is packaged into large molecules known as *chromosomes*. Human cells contain two sets of 23 chromosomes, with one set inherited from each parent. A chromosome contains many genes, each of which is a section of DNA that codes for a specific protein. Although each cell contains a full copy of the individual's DNA, cells use genes selectively depending on the proteins required for the function of that particular cell.

In addition to the sections of DNA that are genes, the genome also contains sections of DNA between genes, or '*intergenic regions*'. Intergenic regions make up a large proportion of the human genome, but little is known about the function of most intergenic DNA.[8] Furthermore, within genes, there are coding regions (exons), which code for protein, and non-coding regions, such as introns and regulatory sequences, both of which have functions relating to the regulation of *gene expression*.

1.3 Genetic variation

Humans are approximately 99.9% identical to each other in terms of their DNA sequence. Differences in the DNA sequence among individuals are known as genetic variation, and can influence many characteristics, including physical appearance, behaviour, and susceptibility to disease. The most common type of genetic variation among humans is a *single nucleotide polymorphism* (SNP). Each SNP is a *locus* i.e. a fixed location on a chromosome, where the type of nucleotide base (A, T, C, or G) present can differ between individuals. The different types of nucleotide bases that can be observed at a given locus are known as *alleles*.

SNPs are most commonly bi-allelic, i.e. all individuals have one of two alleles. The allele that is most commonly observed at a particular locus for a given population is known as the '*wild-type allele*' while the less common allele is known as the '*mutant-type allele*'. For example, at a SNP where most individuals have the C nucleotide base, but some individuals have the T nucleotide base, the wild-type allele is C and the mutant-type allele is T.

For any given locus, an individual possesses two alleles, one on each chromosome. Therefore, if the locus is a bi-allelic SNP, three distinct combinations of alleles are possible. These combinations are referred to as '*genotypes*'. For the previous example, an individual may have the CC genotype, where the C allele occurs at the locus of interest on both chromosomes, the CT genotype, where the C allele occurs on one chromosome and the T allele occurs on the other, or the TT genotype, where the T allele occurs on both chromosomes. In more general terms, an individual with two copies of the wild-type allele at the locus of interest is said to be '*homozygous* wild-type', an individual with two copies of the mutant-type allele is 'homozygous mutant-type', and an individual with one copy of each allele is '*heterozygous*'.

The focus of this work is the investigation of associations between SNPs and treatment outcomes; however, it should be noted that there are other types of genetic variation. For example, the term 'indel' refers to either the insertion or deletion of a stretch of nucleotide bases in the DNA sequence, which may range from one base to thousands of bases in length.

1.4 SNPs and drug response

SNPs may occur in both coding and non-coding regions of genes, and in intergenic regions. A SNP in a coding region is either non-synonymous or synonymous; non-synonymous SNPs alter the amino acid sequence of the protein produced by the gene, while synonymous SNPs have no impact on this sequence. A non-synonymous SNP may therefore affect the structure and/or function of the protein. Historically, synonymous SNPs were thought to be inconsequential, and were widely referred to as 'silent SNPs'. However, there is a growing body of research demonstrating a variety of mechanisms by which synonymous SNPs may alter the structure, function, and/or expression level of proteins.[9] Furthermore, recent research has identified functional SNPs in non-coding regions of genes and in intergenic regions.[10]

Proteins produced by genes have specific functions within cells, which may relate to how the body interacts with drugs. Interactions between the human body and drugs are usually described in terms of two distinct processes; '*pharmacokinetics*' and '*pharmacodynamics'*. Pharmacokinetics is the study of the relationship between drug dose and the resulting plasma and tissue drug concentrations, and is usually subdivided into the processes of absorption, distribution, *metabolism*, and *excretion*. SNPs in genes coding for proteins involved in pharmacokinetic processes can affect the concentration of the drug within the

body; individuals may experience a lack of efficacy if plasma drug levels are too low, and may experience toxicity if plasma drug levels are too high. Pharmacodynamics is the study of the relationship between drug concentration at the target and the physiological and biochemical effect of a drug. SNPs in genes which encode drug targets can alter how a drug binds to its target, and consequently the effect of the drug on the body. Finally, SNPs in genes involved in the disease pathway itself may also lead to variation in drug response. Although individuals may have the same disease, genetic factors contributing towards the disease may differ from one patient to another. If a drug acts by counteracting a specific disease pathway, then treatment response may vary from one individual to the next.

1.5 Pharmacogenetic association studies

In order to determine whether genetic variants are associated with response to a given drug, researchers may perform a pharmacogenetic association study. In such a study, participants are subject to a process called 'genotyping', whereby their genotype is determined for any number of genetic variants. Data are also collected on participants' responses to this drug, in terms of efficacy and/or toxicity. Statistical analyses can then be performed to determine whether there is a statistically significant association between genotype and drug response.

Studies designed to identify associations between genetic variants and treatment-related outcomes are usually carried out using either a case-control or cohort study design. Table 1 provides a description of how each of these designs would be applied to identify pharmacogenetic associations.

Table 1 Study designs used for pharmacogenetic association studies

Study design	Description
Case-control study	The 'case' group consists of patients who have received the treatment of interest and subsequently experienced some treatment-related outcome (i.e. a measure of either efficacy or toxicity). The 'control' group may consist of 'true controls', who have been exposed to the relevant treatment but have not developed the outcome of interest, or 'population controls', i.e. individuals who can be assumed to be controls, although we cannot ascertain whether they would have developed the outcome of interest if they had been exposed to the relevant treatment. Genotype frequencies are compared between the case and control groups.
Cohort study	A sample of patients receiving the drug treatment of interest are recruited according to the study's inclusion criteria. Following enrolment, data are collected on genotype and treatment response for these patients. Treatment responses can then be compared between genotype groups. Participants may be followed prospectively over time to determine their treatment response, or this information may be collected retrospectively from existing records. However, the STROBE guideline[11] recommends that authors refrain from simply calling a study 'prospective' or 'retrospective' as there is no clear consensus on the definition of these terms. When describing cohort studies, authors should instead report how and when data collection took place for the variables and outcomes of interest.

STROBE: STrengthening the Reporting of OBservational studies in Epidemiology

It is possible that researchers conducting studies designed for the primary purpose of comparing different treatment regimens may also collect and report data that can be used to identify pharmacogenetic associations. For example, in an RCT, or a non-randomised study designed to compare different interventions, participants may be genotyped to perform additional pharmacogenetic analyses. If data on genotype and a treatment-related outcome are obtained for a group of participants all receiving a particular treatment of interest, then this data may be used to investigate whether pharmacogenetic associations exist.

In addition to identifying pharmacogenetic associations, studies may be conducted to address other pharmacogenetic-related questions, such as: i) whether an individual's genotype modifies the treatment effect of some intervention of interest *versus* some comparator of interest; ii) whether prescribing treatments according to genotype is beneficial in terms of efficacy and/or safety outcomes. Alternative study designs to the case-control and cohort designs described in Table 1 would be required to address these questions. However, the focus of this thesis is studies that investigate whether pharmacogenetic associations exist, and studies that address other pharmacogenetic-related questions are not discussed in detail.

Pharmacogenetic association studies may be described in terms of the scale of genotyping conducted. Researchers may genotype all known SNPs in the human genome, in a '*genome-wide association study*' (GWAS), or they may choose to genotype a comparatively small set of SNPs which are likely to affect treatment response, in a '*candidate gene study*'. There are strengths and limitations to each of these approaches, and researchers may choose to use either approach for a variety of different reasons.

A pharmacogenetic association study using a candidate gene approach focuses on SNPs within genes that play a role in pharmacokinetic or pharmacodynamic processes, or within genes that are believed to be involved in the pathway of the disease being treated. As previously discussed in Section 1.4, these types of SNPs are most likely to affect drug response. As prior knowledge of the function of the selected genes is required for a candidate gene approach, this type of study is therefore not capable of identifying new genes (or gene combinations) of unknown function that have an important effect on drug response. Conducting a GWAS is more expensive, since hundreds of thousands of SNPs must be genotyped in order to provide coverage of the entire human genome. However, a GWAS is capable of identifying genes that may be important in terms of drug response, even when the function of these genes is unknown.

In addition, due to the number of SNPs genotyped, the number of statistical tests performed for a GWAS is inevitably very large in comparison to the number of tests performed for a candidate gene study. In order to control the overall probability of false positive results being reported, researchers conducting GWASs may lower the threshold at which calculated p-values would indicate statistical significance. Such an adjustment for the large number of statistical tests performed would mean that the GWAS would have lower power compared to a candidate gene study with the same sample size.

1.6 Clinical implications of pharmacogenetic studies

The key motivation for performing pharmacogenetic studies and identifying associations between SNPs and treatment outcomes is that eventually these findings might inform clinical practice. If a genetic variant has been linked to an increased risk of toxicity or poor efficacy outcome for a particular treatment in pharmacogenetic studies, a health care provider may choose to genotype patients before administering the treatment. This would allow the health care provider to make an informed decision regarding the choice of prescribed drug and dosage. In more general terms, implementing the findings of pharmacogenetic studies is a step towards a *stratified medicine* approach to health care.

The terms ‘stratified medicine’, ‘personalised medicine’ and ‘precision medicine’ are often used interchangeably in the literature; there is no consensus on how these concepts differ in practice. For simplicity, in this thesis, we use the term ‘stratified medicine’ to describe the concept of tailoring medical treatment to the individual characteristics of a patient.

In recent years there has been a dramatic rise in our knowledge of pharmacogenetic associations; for example, the Clinical Pharmacogenetics Implementation Consortium (CPIC) has published pharmacogenetic guidelines for 62 drugs as of July 2020.[12] However, few pharmacogenetic tests are routinely used in clinical practice.[13, 14] There are a variety of reasons why the findings of a pharmacogenetic study may not be implemented in clinical practice, including: logistical challenges and costs associated with introducing pharmacogenetic tests, inexperience of clinicians in interpreting and acting on results of these tests, and a lack of RCTs demonstrating the benefits of genotype-guided therapy in comparison to standard therapy.[13]

Despite these barriers, it is becoming increasingly apparent that implementation of pharmacogenetic tests in clinical practice can reduce adverse effects, improve treatment efficacy, and consequently be a cost-effective use of health care providers’ budgets. For example, Verhoef et al.[15] demonstrated that genotype-guided dosing of warfarin in patients with atrial fibrillation is cost effective in the UK. Another study[16] based in the UK investigated pre-prescription pharmacogenetic testing in HIV patients, as *HLA* B*5701 is known to be a genetic risk factor for abacavir hypersensitivity in Caucasians. The authors concluded that pharmacogenetic-guided prescribing to prevent abacavir hypersensitivity is a cost-effective strategy. Indeed, testing for *HLA* B*5701 is now widely performed in clinical practice prior to prescription of abacavir. Another example of pharmacogenetic testing in widespread use is *TPMT* testing (either genotyping or measurement of TMPT enzyme levels) prior to thiopurine prescription.[14]

In order to to encourage researchers to conduct high-quality RCTs of genotype-guided prescribing practices, it is crucial in the first instance that strong evidence exists from pharmacogenetic association studies. Once high-quality RCTs have been performed, or the evidence from association studies suggests a sufficiently large effect that RCTs are not required, health service providers may consider the implementation of pharmacogenetic testing in clinical practice.

1.7 How systematic reviews and meta-analysis can improve the strength of the evidence base

Outcomes from pharmacogenetic studies are often likely to be complex *traits*; genetic influence may be explained by several genetic variants each having only a small effect on outcome.[17] It is unlikely that a common genetic variant will have a large impact on treatment response; it is more likely that the variant will be rare and/or the impact on treatment response will be small to moderate.[18] Therefore, large sample sizes are required to provide sufficient power to identify pharmacogenetic associations. It is often not feasible for pharmacogenetic studies to be conducted on a large scale due to insufficient funding and/or resources.

Meta-analysis is a statistical technique for combining the results of a number of individual studies to produce a summary result, and can therefore be used to improve sample size and increase the power to detect pharmacogenetic associations. Researchers often perform meta-analysis as part of a systematic review, which aims to identify all published and unpublished evidence relevant to a specific research question, and to methodologically summarise this evidence base. Meta-analysis may also be performed outside the context of a systematic review; researchers may wish to consolidate data from multiple studies without performing a systematic search for all available evidence. However, the focus within this thesis is on the conduct of meta-analyses within the context of systematic reviews, as systematic reviews of all available evidence are widely considered to be the most reliable source of evidence to inform clinical practice.

An additional challenge to consider when conducting pharmacogenetic studies is the possibility of false positives occurring, where a statistically significant association is identified despite there being no association between the genetic variant and drug response outcome in reality. This is referred to as a 'type 1 error'. For each hypothesis test performed (i.e. a test of association between a single SNP and outcome), the type 1 error rate determines the probability that a statistically significant result is a false positive finding. When conducting a single hypothesis test, and assuming the usual type 1 error rate of 5%, there is a 5% chance of observing a false positive. As the number of statistical tests increases, the number of false positives likely to be observed also increases. For example, if 20 statistical tests are performed, each with a type 1 error rate of 5%, the risk of at least one false positive finding is 64%. Increasing the number of tests to 100 increases the chance of at least one false positive finding to over 99%. Therefore, due to the large number of

SNPs typically investigated, pharmacogenetic studies are at particular risk of type 1 errors. When a study reports a pharmacogenetic association to be statistically significant, researchers may therefore be interested in whether other studies also provide evidence to support the association of interest. If other studies have replicated the association (in terms of magnitude, direction and possibly statistical significance), this can improve confidence that the original finding was not a type 1 error.

However, replication of findings is particularly uncommon among genetic association studies. Ioannidis et al.[19] examined the results of 55 meta-analyses of genetic association studies, and identified that only nine meta-analyses (16%) were free from heterogeneity and bias, and identified significant and replicated genetic associations. In this study, genetic associations were determined to be sufficiently replicated when the results of the meta-analysis remained statistically significant after excluding the study or studies that first reported the significant association.

To date, there has been no study published on failure to replicate findings in pharmacogenetic studies in particular, although there have been many instances where initial reports of pharmacogenetic associations have not been replicated in subsequent studies. For example, Smeraldi et al.[20] were the first group of researchers to report that individuals carrying at least one copy of the long (l) allele at the serotonin-transporter-linked polymorphism (5-HTTLPR) showed a better response to the antidepressant drug fluvoxamine than individuals with two copies of the short (s) allele. However, many studies have subsequently investigated this association, with conflicting results. A systematic review and meta-analysis conducted by Taylor et al.[21] identified that l allele carriers were no more likely to experience remission than those with the ss genotype.

Failure to replicate a pharmacogenetic association, whether this is purely in terms of the observed magnitude and direction of effect, or whether this also requires the rejection of a null hypothesis to indicate statistical significance, does not necessarily invalidate the initial study's finding. There are various reasons why a pharmacogenetic association may not be replicated in subsequent studies. Study settings and populations may vary in terms of environmental factors that modify the pharmacogenetic effect; indeed, Keers et al.[22] reported that an interaction between stressful life events and the 5-HTTLPR polymorphism predicts response to antidepressant treatment. In scenarios where environmental factors impact the pharmacogenetic association, it would be expected that findings would differ between studies conducted in different settings and populations.

Furthermore, varying patterns of *linkage disequilibrium* and differences in allele frequencies between populations may cause replication failure of pharmacogenetic associations. Linkage disequilibrium describes the extent to which the frequency of a combination of alleles at two *loci* differs to the frequency that would be expected if the loci were independent and associated randomly. If two SNPs are in strong linkage disequilibrium, and one of these SNPs has a causal relationship with a drug response outcome, a significant association may be observed between the non-causal SNP and the outcome due to the correlation between the observed alleles at the two loci. If a particular allele is strongly associated with drug response outcome in one population, but not in another, this may be due to varying extents of linkage disequilibrium between the genotyped SNP and the causal SNP in the different populations as opposed to a true absence of association in the second population. In addition, a study may not replicate a previously identified statistically significant association if the allele of interest is observed in smaller frequencies in the replication study population than in the original study population, due to reduced power.

Finally, studies may vary considerably in terms of their methodological quality; Jorgensen and Williamson[18] identified key issues relating to the methodological quality of pharmacogenetic studies. These issues may introduce bias into studies, and cause replication failure across pharmacogenetic studies.

An important component of any well conducted meta-analysis (within a systematic review) is an exploration of heterogeneity between studies. Authors conducting meta-analyses are able to investigate how a pharmacogenetic association varies according to covariates such as study setting or population, by performing subgroup analyses or meta-regression.[23] Furthermore, authors of systematic reviews and meta-analyses ought to assess the methodological quality of included studies; such an assessment may highlight important methodological differences between studies, which could cause differences in overall findings. Authors may even choose to perform sensitivity analyses, in which studies of particularly poor methodological quality are excluded from meta-analyses.

Therefore, to researchers striving to improve the strength of evidence bases for pharmacogenetic associations, meta-analysis is an indispensable tool. Performing a systematic review and meta-analysis can not only improve power to detect pharmacogenetic associations, but also can enable investigations of heterogeneity and the impact of methodological quality on study results. Performing these investigations may

shed light on why pharmacogenetic associations have not been replicated across studies, or they may highlight the possibility that initial significant associations were spurious findings.

1.8 Thesis aim and structure

The aim of this thesis is to improve the strength of the evidence base for pharmacogenetic associations, by:

i) Identifying challenges that reviewers might encounter when synthesising evidence from pharmacogenetic studies;

ii) Resolving these challenges with the development of new methodology and guidance that will facilitate future evidence synthesis within pharmacogenetics.

In Chapter 2, we outline methodology for evidence synthesis, i.e. performing systematic reviews and meta-analyses, of pharmacogenetic studies. In Chapters 3 and 4, we demonstrate how these methods can be applied in practice, by conducting a large systematic review and meta-analysis of pharmacogenetic studies investigating associations between genetic variants and anti-tuberculosis (TB) drug-related toxicity. In Chapter 5, we explore more complex methods of meta-analysis and apply these methods to data from our systematic review. In the process of performing the systematic review and meta-analysis, we identified that improvements in the reporting of pharmacogenetic studies would enable researchers to perform high-quality meta-analyses that include all relevant evidence. In Chapter 6 therefore, we develop a reporting guideline for pharmacogenetic studies using widely accepted and robust methodology. In Chapter 7, we summarise the overall findings and implications of this thesis, outline areas for future work, and make concluding remarks.

2 Methods of evidence synthesis for pharmacogenetic studies

To date, there has been no widely disseminated guidance specifically for evidence synthesis of pharmacogenetic studies. However, guidance on evidence synthesis methods for genetic association studies is available in the HuGENet HuGE Review Handbook,[24] and the methods outlined are applicable to evidence synthesis of pharmacogenetic studies. Since the development of the HuGENet HuGE Review Handbook,[24] Jorgensen and Williamson[18] have developed a checklist of methodological issues to be considered by primary researchers when conducting a pharmacogenetic study. The authors confirm that the checklist can also be used to assess the methodological quality of studies included in a systematic review and meta-analysis of pharmacogenetic studies. In this chapter, we summarise and reflect on the key issues from the HuGENet HuGE Review Handbook[24] and Jorgensen and Williamson's checklist,[18] which ought to be considered when synthesising evidence from pharmacogenetic studies.

2.1 Searching for studies

The HuGENet HuGE Review Handbook[24] emphasises the importance of conducting a comprehensive search, as systematic reviews of genetic association studies are at particular risk of being affected by reporting biases. Reporting bias occurs when decisions about how, when or where to report results of studies are influenced by the findings of the study.[25] Considering the huge number of outcome and genetic variant combinations that are possible, it is clear that the number of genetic associations under investigation at any one time is likely to be vast. It is also therefore probable that only the most exciting results will be published in easily accessible journals.

Key types of reporting biases are publication bias, and language bias. Publication bias occurs when investigators submit manuscripts for publication, or reviewers and editors accept manuscripts for publication, based on the direction or strength of their findings.[26] Language bias occurs when choice of language for publication depends on the nature of the findings.[25] If only published studies, or only studies published in a certain language are included in the review, the estimates of effect reported in the included studies may not be reflective of the true relationship between the variables of interest. It is therefore impotant that review authors attempt to identify all relevant studies, regardless of publication status, and language.

The HuGENet HuGE Review Handbook[24] recommends including non-English language journals in the search, which may not be indexed in commonly searched databases, such as

PubMed, ISI Science Citation Index, EMBASE and BIOSIS. It is also recommended that authors attempt to identify unpublished studies by searching 'grey literature', such as conference proceedings, technical reports and books, and include any studies identified, providing there is sufficient information to perform a thorough quality assessment.[24] Checking the reference lists of existing reviews on the topic and articles identified for inclusion in the review may also identify unpublished or published studies not identified by the initial search.

2.2 Assessment of methodological quality of pharmacogenetic studies

An assessment of the internal validity of a study's results is a key component of any good systematic review. Such an assessment guides the systematic reviewer's approach to analysis, interpretation of results and conclusions.[25] Jorgensen and Williamson[18] have developed a quality assessment checklist to be referred to by researchers when conducting pharmacogenetic studies. The checklist may also be used by researchers conducting systematic reviews and meta-analyses of pharmacogenetic studies.

It is important to note here that concerns relating to methodological quality may not always have an impact on the internal validity of a study's findings; in other words, poor standards of conduct may not necessarily lead to overestimation or underestimation of the true effect. Similarly, it is possible that a study's results may be impacted by sources of bias that even a study conducted to the highest possible standard would not be able to overcome. It is for these reasons that in recent years, there has been a shift among the evidence synthesis research community, from performing assessments of methodological quality, to assessing the risk that bias may have affected particular numerical results.

To the best of our knowledge, no tool exists that has been designed to assess the risk of bias in pharmacogenetic studies. Although not specifically tailored for this purpose, Jorgensen and Williamson's checklist[18] may help systematic reviewers identify issues that would impact the internal validity of a pharmacogenetic study's results. Systematic reviewers using the tool ought to be aware that some of the listed criteria would not have implications for the internal validity of a pharmacogenetic study's results, and therefore would not need to be highlighted in the review as particular issues of concern. In addition, there are some potential sources of bias not described in the checklist; these are discussed further in Section 2.5.

In this chapter, we summarise the key issues of methodological quality identified by Jorgensen and Williamson.[18] In Chapter 3, we apply the checklist to a set of

pharmacogenetic studies, and highlight instances where criteria are of little relevance to the validity of study results included in a systematic review.

Choosing which genes and SNPs to genotype

As previously discussed in Section 1.7, pharmacogenetic studies that investigate a large number of SNPs are at particular risk of type 1 errors due to the large number of statistical tests performed. There are various methods available that can be used to overcome the issue of multiple testing in genetic association studies, each of which has its own strengths and limitations. Bouaziz et al.[27] discuss the most common methods of multiplicity adjustment used in genetic association studies, and provide advice for researchers on how to choose with method to use. Ultimately, the statistical significance of a result and conclusions of a study may vary according to the method of adjustment for multiple testing. For clarity, pharmacogenetic studies ought to describe any methods used to adjust for multiple testing.[18]

The problem of multiplicity becomes particularly prominent for GWASs, in which hundreds of thousands of SNPs are investigated. Furthermore, there may be no theoretical basis for associations identified between SNPs and drug response outcomes in a GWAS, and so particular caution ought to be taken when interpreting statistically significant results.

On the other hand, in a candidate-gene study, SNPs have been selected for investigation as there is a plausible biological explanation for association between the SNP and drug response. Therefore, multiple testing is often a less concerning, although certainly not negligible, issue for consideration. It is important that researchers conducting a candidate gene study choose the genes and SNPs to be investigated in a systematic way, using prior knowledge to guide their decisions. Jorgensen and Williamson[18] outline a procedure for the systematic selection of genes and SNPs in candidate gene studies. Such an approach ensures that the resources allocated to the study are used efficiently by prioritising genetic variants that are likely to influence drug response. Furthermore, having strong prior evidence can improve confidence that a statistically significant association is genuine, rather than a type 1 error. Clear rationale also instils confidence in the reader that all analyses performed have been reported, rather than only statistically significant results. Therefore, Jorgensen and Williamson[18] recommend that study authors report their rationale for investigating all chosen genes and SNPs.

Sample size

An important challenge for researchers conducting pharmacogenetic association studies is achieving sufficient statistical power. As previously outlined in Section 1.7, large sample sizes are required to provide sufficient power to identify pharmacogenetic associations. Indeed, Jorgensen and Williamson[18] identified that in the majority of hypothetical scenarios considered, thousands of patients would be required to detect a pharmacogenetic association, assuming 80% power.

Software programmes are available that can be used to perform power calculations for pharmacogenetic studies with binary and continuous outcomes,[28-30] and time-to-event outcomes,[31] although authors may of course use other methods. Jorgensen and Williamson[18] recommend that authors of pharmacogenetic studies provide details of the sample size calculation and the *a priori* power to detect effect sizes of varying degrees, so that the reader is aware of the likelihood of false negative results (i.e. type 2 errors) occurring.

Study design

Pharmacogenetic studies can be conducted using either a case-control, cohort or RCT design (see Section 1.5, Table 1,). It is important for study design to be clearly described in pharmacogenetic study reports, to allow the reader to take this into consideration when interpreting the findings and comparing results with other studies. For example, case-control studies and cohort studies where covariate data were collected retrospectively may be at risk of recall bias. Recall bias is a systematic error caused by differences in the accuracy or completeness of recollected data between those who did and did not experience the outcome of interest.[32] It is only environmental factors that may be recalled differentially between participants; however, the overestimation or underestimation of environmental effects may subsequently lead to an unreliable estimate of the contribution of genetic factors. In cohort studies where outcome status is unknown at the time of recruitment and covariate data collection (i.e. participants are followed up prospectively), recall bias is unlikely to be an issue of concern.

However, it is often impractical to perform a cohort study and follow patients prospectively. In particular, if the outcome is rare, the study would require a large number of participants to ensure that an adequate number of patients experience the outcome of interest. Case-control designs are often chosen as participants are recruited based on their

case-control status, and researchers can ensure that a sufficient number of cases are included.

Jorgensen and Williamson[18] make specific recommendations for case-control studies. Firstly, the case and control groups ought to be clearly defined so that researchers attempting to replicate the study's findings would know which individuals to recruit. Secondly, it is recommended that cases and controls should be genotyped in mixed batches, rather than in separate batches, to ensure that genotyping quality is comparable between groups.

Reliability of genotypes

A variety of genotyping methods are available to researchers conducting pharmacogenetic studies, each of which has its own strengths and weaknesses.[33] Genotyping error rates vary between these methods, and are often non-negligible; error rates per locus of between 0.2% and 15% have previously been reported.[34] Genotyping errors can lead to loss of power[35] and an increase in type 1 error rates[36] in genetic association studies. It is therefore essential that researchers quality assess their genotyping results by re-genotyping all or a random sample of participants.[18] Comparing the observed genotype frequencies to those reported in previous studies can also provide an indication of the reliability of genotyping results.[18]

For studies in which human inference is part of the genotyping procedure, Jorgensen and Williamson[18] recommend that genotyping is undertaken by two individuals separately, who are blinded to outcome status, with any discrepancies in genotyping results resolved by a third individual. Using such methodology ought to reduce the likelihood of bias due to incorrect genotype allocation. It is also recommended that researchers describe genotyping methods and quality control methods, and report the findings of quality checks.[18]

Missing genotype data

If an individual's genotype cannot be determined using the chosen genotyping method, then the genotype will be recorded as missing. Researchers may decide to re-genotype certain SNPs or individuals with large proportions of missing data. However, it is likely that some genotype data will still be missing following this re-genotyping.

Jorgensen and Williamson[18] recommend that authors of pharmacogenetic studies report the extent of missing genotype data, and summarise reasons for this missing data. If the extent of missing data is minimal, it may be acceptable to simply exclude individuals with missing genotype data from the analysis. If the extent of missing data is not minimal,

genotypes can be imputed to avoid loss of power by excluding individuals.[37-39] Regardless of whether missing data are imputed or not, authors should report how many individuals were included in each analysis as an indication of the available power.

It is important to consider whether data are missing at random,[18] i.e. the fact that the genotype is missing for an individual is unrelated to the individual's true genotype. If study authors exclude individuals with missing genotype data from the analyses, the results may be biased if the missing data are not missing at random. Missing genotype data are unlikely to be missing at random, as *heterozygotes* are notoriously more difficult to call than *homozygotes.* In practice, even carefully designed studies, such as the HapMap project,[40] have missing genotype data that are not missing at random.[41] If authors decide to impute missing genotype data, it is particularly important to consider whether data are likely to be missing at random so that an appropriate method of imputation can be chosen. Jorgensen and Williamson provide guidance on testing whether data are missing at random.[18]

Population stratification

Population stratification exists when the studied population contains subpopulations that differ in terms of allele frequencies due to ethnic diversity. Population stratification can be problematic for pharmacogenetic association studies when the prevalence of the relevant outcome also differs between the different subpopulations. In this case, subpopulation therefore influences both genotype and outcome, and can confound the relationship between the two. If not properly accounted for in analyses, population stratification can cause false positive findings to occur in pharmacogenetic studies.

Researchers conducting pharmacogenetic studies may choose a study design that reduces the likelihood of confounding due to population stratification. For example, the eligibility criteria for a study might state that each participant's four grandparents must all be from a specific ethnic population. Although such a method would minimise ethnic diversity within the study population, subtle differences may still exist and the effects of population stratification may not be completely eradicated.[42] These subtle differences between subpopulations, that might not be instantly recognisable to the researchers, are sometimes referred to as 'cryptic population stratification'.

Jorgensen and Williamson recommend performing tests to detect cryptic population stratification, and if necessary, adjusting for cryptic population stratification in the analyses.[18] In a recent review, Hellwege et al.[43] summarise well-established and more novel

methods for detecting the presence of, and accounting for, cryptic population stratification, and outline the capabilities and limitations of each method.

Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium (HWE) principle states that when mating is random in a large population, and there are no disturbing factors, genotype and allele frequencies for a bi-allelic SNP will remain constant from one generation to the next. If a population is in HWE, and a bi-allelic SNP has wild-type and mutant-type allele frequencies, p and q, respectively, then then the frequency of wild-type homozygotes, mutant-type homozygotes, and heterozygotes, will be p^2, q^2, and $2pq$, respectively.

Human populations rarely meet the assumptions underlying HWE exactly; disturbing factors such as *natural selection*, mutations and migration all act on human populations (and most other large populations). Therefore, allele frequencies are likely to vary from generation to generation. However, as a null hypothesis, it is reasonable to expect that HWE holds in a population of healthy controls.[24]

A variety of methods are available for testing and/or measuring deviation of genotype frequencies from HWE.[44] Although deviations from HWE may be due to disturbing factors which the researcher has no control over, it is also possible that deviation may be caused by genotyping errors,[45] by the existence of population stratification or by biased selection of controls.[46] It is therefore crucial that researchers conducting pharmacogenetic studies should assess deviation from HWE for each investigated SNP, and explore possible reasons for deviation from HWE where applicable. Jorgensen and Williamson[18] recommend that study authors outline details of the method used to test deviation from HWE and highlight any SNPs that were found to deviate.

Mode of inheritance

Mode of inheritance refers to the way in which a genetic trait is passed from one generation to the next. In humans, there are two alleles present at almost all loci, one from each parent; the mode of inheritance determines what the observed *phenotype* will be for each combination of alleles. An exception to this rule occurs on the sex chromosomes; males have one X chromosome and one Y chromosome (XY), while females have two X chromosomes (XX). The Y chromosome is smaller than the X chromosome, and contains a number of genes which are involved in the initiation and maintenance of maleness, but lacks copies of most of the genes that are found on the X chromosome. Therefore, males will only have one allele present at some loci. Modes of inheritance for genetic variants on

the sex chromosomes differ to modes of inheritance for genetic variants on the autosomes (i.e. the non-sex chromosomes), and are not discussed further here.

For genetic variants on the autosomes, there are various modes of inheritance, some of which are described in Table 2. For simplicity, it is assumed that the hypothetical genetic variant described for each mode of inheritance is solely responsible for an observable trait, whereas in reality, observable traits are likely to be controlled by multiple genetic variants.

Table 2 Examples of different modes of inheritance

Mode of inheritance	Description	Example
Dominant	The phenotype associated with a particular allele is observed when an individual has one copy of the allele (heterozygous) or two copies of the allele (homozygous).	Seed shape in peas: allele R is dominant and associated with round seed shape, allele r is associated with wrinkled seed shape. RR and Rr genotypes will result in round seed shape, rr genotype will result in wrinkled seed shape.
Recessive	The phenotype associated with a particular allele is only observed when an individual has two copies of the allele (homozygous).	Sickle cell disease: only individuals who have two copies of the allele which is associated with sickle cell disease will have the condition.
Co-dominant	Homozygous individuals display the phenotype associated with the allele that they have two copies of. Both alleles are simultaneously expressed in heterozygous individuals.	Blood type: A and B alleles code for proteins that exist on the surface of red blood cells. Heterozygous individuals, with one copy of the A allele and one copy of the B allele, form a new blood group, AB, and both A and B proteins are expressed on the surface of red blood cells equally.

When investigating association between a SNP and a treatment response outcome, researchers may assume a specific mode of inheritance. The analysis plan can then be tailored to the chosen mode of inheritance, improving the power to detect pharmacogenetic associations. For example, if the mode of inheritance for a particular allele is assumed to be dominant, there are only 2 observable phenotypes; one will be shown in individuals who are homozygous for the dominant allele or heterozygous, the other will be shown in individuals who are homozygous for the non-dominant (recessive) allele. It is therefore only necessary to make comparisons between two genotype groups, rather than all three genotypes. Jorgensen and Williamson demonstrate how regression models might be structured if other modes of inheritance are assumed, to ensure that the analysis is as efficient as possible.[18]

Authors of pharmacogenetic studies may use statistical methods to guide their choice of genetic model selection,[47] or consider data from functional studies. However, strong evidence to support a particular mode of inheritance is often lacking, and if this is the case, authors should carefully consider their options.[18]

Authors may choose to assume a specific mode of inheritance, taking into consideration the possible benefits and risks of such an approach; statistical power may be improved if the assumed mode is true, but may be reduced if the assumed mode is incorrect. It is also possible to perform analyses making no assumptions about the underlying mode of inheritance (i.e. two pairwise comparisons of the following genotypes: heterozygous *versus* homozygous wild-type, and homozygous mutant-type *versus* homozygous wild-type). In this case, Jorgensen and Williamson[18] recommend that authors explain the limitations of such an approach; namely, that statistical power will be reduced in comparison to the power that would be achieved by using an approach tailored to the true mode of inheritance. Finally, authors may undertake multiple analyses, each making a different assumption about the underlying mode of inheritance, or making no assumption about the mode of inheritance. In this case, adjustments for multiple comparisons ought to be made.[18]

As there are various approaches to the analyses of pharmacogenetic data, readers of pharmacogenetic study reports may have concerns regarding selective reporting of results. Selective reporting would have occurred if authors performed multiple analyses assuming various modes of inheritance, but only reported the most interesting, or most statistically significant results. To reassure readers that analyses were not selectively reported, study authors ought to report all analyses that were undertaken, and provide justification for performing each of these analyses.[18]

Choice and definition of outcomes

Pharmacogenetic studies aim to identify how treatment response is influenced by genetic variants. In some cases, researchers may be interested in a single measure of treatment response, such as score on a self-report depression scale. However, treatment response is often multifaceted; there may be various ways to measure treatment efficacy, and also adverse events to be considered.

Due to the fact that a variety of relevant outcomes often exist for pharmacogenetic studies, selective reporting of outcomes is an issue of concern. In particular, authors may only report outcomes for which statistically significant results were observed. Omitting non-statistically significant findings from the study publication distorts the results, and

conclusions ought to be interpreted with caution. It is therefore essential that study authors clearly list all investigated outcomes, and provide justification for their choice of outcomes.[18]

Outcome definitions across pharmacogenetic studies have previously been reported to be heterogenous.[48, 49] Jorgensen and Williamson[18] recommend that study authors provide clear definitions for all investigated outcomes, so that readers are able to interpret the results correctly and compare results across studies.

Treatment adherence

In Jorgensen and Williamson's[18] checklist, the term 'compliance' is used. Here, we use the term 'adherence', as in recent years, bodies such as the World Health Organization (WHO)[50] and in the UK, the National Co-ordinating Centre for NHS Service Delivery and Organisation (NCCSDO)[51] have demonstrated a preference for this term.

Treatment adherence can be defined as the extent to which a patient's drug-taking history coincides with the prescribed drug regimen. In general, treatment adherence is not an issue of primary concern for non-pharmacogenetic studies of drug efficacy. This is because the aim of these studies is to estimate how effective the drug will be when used in a real-world setting; in reality, patients are likely to occasionally be non-adherent with the prescribed regimen, and so primary trial results will be reflective of clinical practice. Effects of adhering to the drug can also of course be explored, by adopting a per-protocol approach to analysis, but these are usually secondary concerns. In contrast, the primary aim of a pharmacogenetic study is typically to estimate how much impact genetic variants have on the drug's biologic pathway through the body. Taking too much or too little of the prescribed drug undoubtedly may also impact the body's response to the drug. Therefore, unless adherence is adjusted for, the estimates of association between genetic factors and the treatment-related outcome of interest may be underestimated or overestimated. Treatment adherence should therefore be of primary concern when undertaking statistical analyses of pharmacogenetic studies, and consequently extent of this adherence should be measured as accurately as possible.

Generally, adjusting for a non-confounding covariate (such as treatment adherence) can explain variability in the outcome, consequently reducing noise and increasing power to detect pharmacogenetic associations.[52] However, it is important to note that when the drug response outcome is binary, and individuals are recruited according to case or control status, adjusting for the covariate can actually reduce power.[53] Nevertheless, methods have

been developed to overcome this issue,[54-56] which account for non-confounding covariates while increasing power to detect genetic associations in case-control studies. It is therefore advisable to account for treatment adherence when investigating pharmacogenetic associations, even in the analyses of case-control studies (providing careful consideration is given to the choice of analysis method). Indeed, Jorgensen and Williamson[18] recommend both measuring and adjusting for treatment adherence.

2.3 Analysis methods

Methods of meta-analysis are well-established for RCTs. There is a wealth of information available for researchers who wish to meta-analyse data from RCTs, including the Cochrane Handbook,[25] the Centre for Reviews and Dissemination's (CRD) guidance for undertaking reviews in health care,[57] and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement.[58] The application of meta-analysis methods to observational studies has been controversial, due to the greater potential for biased results, and the extreme diversity in study designs that is likely to be observed.[59] However, it has been argued that genetic association studies are more closely-related to RCTs than other types of observational study, due to the concept of 'Mendelian randomisation'.[60] Mendelian randomisation refers to the random assortment of alleles at the time of *gamete* formation, and the use of this random assortment to make inferences about environmentally modifiable exposures. In some circumstances, genetic association study designs can therefore be thought of as akin to randomised comparisons.

The HuGENet HuGE Review Handbook[24] strongly encourages quantitative synthesis of the results of genetic association studies, with a specific focus on the investigation of heterogeneity; exploring possible reasons for variability between the included studies can help to identify important sources of bias.

Data synthesis

For a meta-analysis of genetic association studies, authors may be comparing groups determined by the presence of a particular allele i.e. carrier *versus* non-carrier, or they may be comparing groups determined by genotype. Generally, authors should perform analyses by comparing genotype or allele groups without any adjustment for confounding.[24] However, it may be informative to also extract confounder-adjusted measures of association from the original study reports for comparative purposes.

Standard methods of meta-analysis (i.e. those applied in meta-analysis of RCTs) can be used when there are two groups to compare. Clearly, these methods can easily be applied to

comparisons based on carrier or non-carrier status of a given allele. For comparisons of different genotypes, there are usually three genotype groups to be compared, since most SNPs are bi-allelic. Therefore, to apply standard methods of meta-analysis, authors must reduce these three groups to two. For a bi-allelic SNP with two alleles, A and a, where A is thought to be associated with a particular outcome (i.e. treatment response or toxicity in a meta-analysis of pharmacogenetic studies), the following analysis approaches are possible:

i. Assuming a dominant mode of inheritance, thus comparing 'AA+Aa' to 'aa'
ii. Assuming a recessive mode of inheritance, thus comparing 'AA' to 'Aa + aa'
iii. Assuming a co-dominant mode of inheritance, where the effect of the Aa genotype is mid-way between the effect of AA and aa
iv. Performing two separate pairwise comparisons (i.e. 'Aa' *versus* 'AA' and 'aa' *versus* 'AA')
v. Using a 'genetic model-free' approach[61]

As previously discussed in Section 2.2, analyses that assume a specific mode of inheritance (i.e. approaches i-iii) may improve statistical power if the assumed mode is true, but may reduce power if the assumed mode is incorrect. The HuGENet HuGE Review Handbook[24] recommends presenting separate effect estimates for heterozygotes and homozygotes where possible; this allows the reader to make their own judgements regarding the most likely mode of inheritance. Approaches iv and v can both be used to obtain these separate effect estimates, and make no assumption about the underlying mode of inheritance. Approach v developed by Minelli et al.[61] is more complex than the other approaches (all of which use standard methods of meta-analysis) and will be discussed further in Chapter 5.

If a particular mode of inheritance was proposed at the time of the first reporting of an association, and it seems unlikely that selective reporting of this particular analysis occurred, meta-analysts ought to also perform analyses assuming this mode of inheritance.[24] If analyses assuming other modes of inheritance are to be undertaken, these ought to be pre-specified in the analysis plan, and reported in full in the publication of the review.[24]

Investigation of heterogeneity

Methods that would be used for an investigation of heterogeneity in a systematic review of RCTs can be applied to a systematic review of genetic association studies. Statistical tests can be used to detect the presence of heterogeneity, and statistics such as the I^2 statistic can be used to quantify heterogeneity.[25] Subgroup analyses, meta-regression and

sensitivity analyses are all tools that can be used to explore potential sources of heterogeneity.

Specifically, the HuGENet HuGE Review Handbook[24] recommends performing sensitivity analyses to investigate how results from meta-analyses are impacted by HWE deviation; Trikalinos et al.[62] reported that excluding HWE-violating studies or making adjustments for deviations from HWE can impact the statistical significance of meta-analysis results and the extent of between study-heterogeneity observed. Since the HuGENet HuGE Review Handbook was published,[24] Minelli et al.[63] have performed an empirical evaluation of how studies that deviate from HWE are handled in meta-analyses, and concluded that HWE studies ought not to be automatically excluded. The authors argue that departure from HWE does not necessarily indicate that a study's results are biased, and that the exclusion of large studies with small departures from HWE may be detrimental rather than beneficial. The authors also advise that when sensitivity analyses are performed to compare meta-analysis results when studies that deviate from HWE are excluded and included, reviewers should tend toward accepting the analysis that includes all studies.

An examination of how the results of studies change over time may also be informative. The 'Proteus phenomenon' has been observed among meta-analyses of genetic association studies, whereby the earliest study included in a meta-analysis demonstrates the strongest effect, and the next earliest study demonstrates the strongest effect in the opposite direction.[64] It is hypothesised that the phenomenon may be prominent in the field of genetic research, due to the enormous number of outcome, genetic variant and mode of inheritance combinations that researchers may choose to investigate.[65] Consequently, it is inevitable that some spurious findings will be identified and reported, and subsequent studies will fail to replicate the results. If the 'proteus phenomenon' has occurred, the earliest studies will report the most extreme results which are not representative of the true association. The HuGENet HuGE Review Handbook[24] advises routinely checking the results of the earliest study/studies against subsequent research as a way of investigating whether this phenomenon has occurred.

Due to the possibility of confounding by population stratification (see Section 2.2), it is recommended that meta-analyses are always stratified by ethnicity. Pooling of results should only be performed if effect estimates for different ethnic groups appear sufficiently similar.[24]

Furthermore, as previously discussed in Section 1.7, a significant association between a SNP and a treatment response outcome does not necessarily indicate a causal relationship; it is possible that the association only exists due to the SNP of interest being in strong linkage disequilibrium with the causal SNP. Patterns of linkage disequilibrium vary from one population to the next,[66] and therefore heterogeneity may be observed across the results of studies conducted in different populations. This provides further rationale for stratifying all meta-analyses by ethnicity.

2.4 Assessment of publication bias

As outlined in Section 2.1, genetic association studies are at particular risk of being affected by publication bias. The HuGENet HuGE Review Handbook[24] advises that there is no way to definitively confirm the existence or non-existence of publication bias within a systematic review. Funnel plots may be used to visually explore the possibility that publication bias has occurred, with or without accompanying statistical tests to detect asymmetry in the funnel plot.[67-70] However, it has been reported that funnel plots can be misleading, and only a limited number of researchers are able to correctly interpret these graphs.[71] Furthermore, the tests for detecting funnel plot asymmetry also have their limitations, as outlined by Sterne et al.[72] It is particularly important to note that visually examining funnel plots and statistical tests can only highlight that asymmetry exists in the funnel plot, suggesting that small studies give different results to larger studies. However, publication bias is just one of several possible explanations of these 'small study effects'. Other possible sources of asymmetry in funnel plots, or 'small study effects', include differences in methodological quality between smaller and larger studies, and true heterogeneity, i.e. the true effect size varies according to study size.[67]

As it is difficult to determine whether publication bias exists in a systematic review, it is important to take measures to minimise this bias when searching for studies to include in the review. Such measures are outlined in Section 2.1.

2.5 Discussion

Due to the rapid growth of the field of pharmacogenetics in recent years, the potential for synthesising evidence from multiple pharmacogenetic studies has increased. Researchers wishing to conduct systematic reviews and/or meta-analyses of pharmacogenetic data ought to follow guidance outlined in the HuGENet HuGE Review Handbook,[24] and assess the quality of included studies using the checklist developed by Jorgensen and Williamson.[18] Using such methodology will ensure that findings from the review are valid and interpreted

appropriately, and that readers are aware of any important limitations of the evidence base.

Review authors ought to be aware that there may be sources of bias in pharmacogenetic studies not covered by Jorgensen and Williamson's checklist.[18] In particular, there are issues relating to the recruitment of participants that may impact the validity of a study's reported results. Firstly, if patients are recruited to studies after treatment has commenced, and patients are less likely to be recruited to the study if they experience early adverse reactions to the drug treatment of interest, this will introduce selection bias. However, some studies that recruit patients after treatment initiation would inherently provide data on events occurring prior to recruitment (case-control studies would commonly be conducted in this way), and post-treatment enrolment in these circumstances is unlikely to be an issue of concern. Secondly, if participants are less likely to be recruited to the study due to results of previously conducted genotyping (for example, if investigators believe that the study treatment will lead to toxicity in patients with a particular genotype), then this would also introduce selection bias to the study.

In addition, although there is a criterion in Jorgensen and Williamson's checklist[18] that considers whether case and control groups are well-defined, it may also be informative for reviewers to assess whether cases and controls are from the same underlying patient population. There should be no important differences between the populations in terms of factors that may influence treatment outcome, such as the distribution of genotypes within the population, or history of treatment prior to recruitment.

There are also two issues relating to measurement of outcomes that it would be beneficial for reviewers to consider when assessing the internal validity of pharmacogenetic study results. Although Jorgensen and Williamson's checklist[18] has a criterion which assesses whether genotyping personnel were blinded to outcome status, the tool does not consider whether those assessing outcome status were blinded to patients' genotypes; a lack of blinding of outcome assessors could introduce detection bias to the study. Finally, there is no explicit consideration of whether evidence exists to suggest that researchers assessed outcomes according to multiple definitions, and selectively reported only the most interesting or statistically significant results. This issue is of particular importance when composite definitions of outcomes are used, as the potential for selective reporting increases as the number of possible outcome definitions increases. Seeking out study

protocols or analysis plans may provide evidence that reported outcome measures differed to those specified prior to study initiation.

In this chapter, we focus on methods for aggregate data meta-analysis, whereby summary data obtained from study reports or communication with study authors are synthesised. We do not discuss methods for individual participant data (IPD) meta-analysis, which involves the collection of detailed data for each participant from each trial. Although there are advantages to conducting IPD meta-analysis, such as the possibility of standardising definitions of outcomes and covariates, and more flexibility in terms of the analysis strategy, IPD meta-analysis is also significantly more resource-intensive than aggregate data meta-analysis. It is often the case for research teams that it would not be feasible to conduct an IPD meta-analysis, due to a lack of time and/or funds to commit to the project. Furthermore, authors of pharmacogenetic studies may be unwilling or unable to share data for the purposes of IPD meta-analysis. There may be data security issues or data sharing restrictions; these are likely to differ between countries and will also depend on the initial consent given for use of the genetic data. For these reasons, throughout this chapter, and the thesis as a whole, our focus is on aggregate data meta-analysis of pharmacogenetic studies.

Finally, it is important to note that here we have outlined guidance on methods of evidence synthesis for genetic association studies,[24] which can be applied to evidence synthesis of pharmacogenetic studies. However, this guidance was made available in 2006, and since this time, new methods for meta-analysis of genetic association studies have been proposed. Namely, Salanti and Higgins[73] have developed methodology to enable the synthesis of data from studies that report data for each genotype group separately and studies that report data for combined genotype groups. Verzilli et al.[74] have introduced methodology which enables the synthesis of data from studies investigating different, but highly correlated, SNPs; this method maximises power to detect true genetic associations. Finally, Thompson et al.[75] have developed the work of Minelli et al.,[61] demonstrating how combining information from related meta-analyses of genetic association studies can lead to increases in the precision of estimated genetic effects. We do not discuss these methods in further detail in this thesis; it is beyond the scope of our work to explore all methodology that has been developed in this area. Instead, we focus on methods outlined in the available guidance on evidence synthesis for genetic association studies, as these methods are most likely to be used by future researchers who wish to undertake systematic reviews of pharmacogenetic studies.

3 Influence of genetic variants on toxicity related to anti-tuberculosis drugs: a systematic review

We have previously discussed how systematic reviews and meta-analyses can improve the strength of the evidence base for pharmacogenetic associations (Section 1.7), and outlined methodology for evidence synthesis of pharmacogenetic studies (Chapter 2). In this chapter, and the following chapter (Chapter 4) we demonstrate how these methods can be applied in practice, by conducting a systematic review and meta-analysis of associations between genetic variants and anti-TB drug-related toxicity. Many studies have investigated the pharmacogenetics of anti-TB drugs, making this topic area a prime candidate for the application of evidence synthesis methods. In the process of conducting the systematic review and meta-analysis, we hoped to identify whether obstacles exist which hinder the potential for reviewers to conduct high-quality and well-powered meta-analyses of pharmacogenetic studies.

In order to identify the total accumulated evidence base on the pharmacogenetics of anti-TB drug-related toxicity, we conducted a systematic review in accordance with the methods described in Chapter 2. This included a thorough quality assessment of each included study using the checklist developed by Jorgensen and Williamson.[18] In this chapter, we summarise the studies identified for inclusion in the systematic review, and the results of the quality assessment. Where appropriate, data from studies included in the systematic review were synthesised in meta-analyses for each combination of genetic variant and toxicity outcome; methods and results of the meta-analyses are provided in Chapter 4.

3.1 Background

TB is an infectious disease caused by *Mycobacterium tuberculosis* bacteria and is the leading cause of death from a single infectious agent (ranking above HIV/AIDS).[76] For individuals with drug-susceptible TB, the WHO currently recommends combination treatment with four first-line drugs: isoniazid, rifampicin, ethambutol and pyrazinamide.[76]

Treatment with anti-TB drugs may cause patients to experience serious adverse effects, such as anti-TB drug-induced hepatotoxicity (ATDH). Incidence rates of ATDH for patients treated with the standard combination treatment have been reported to vary from 2% to 28%, depending on the treatment regimen, patient characteristics (e.g. age, race and sex), and definition of ATDH.[77] ATDH may be fatal, with reported mortality rates of 6–12% if treatment is not stopped promptly.[78] ATDH and other anti-TB drug-related toxicity

outcomes may also lead to poor patient adherence, which in turn may result in treatment failure, relapse and the emergence of drug resistance.[77]

Proposed genetic risk factors for ATDH include polymorphisms of the N-acetyltransferase 2 (*NAT2)*, cytochrome P450 2E1 (*CYP2E1*), glutathione S-transferase mu 1 (*GSTM1*) and glutathione S-transferase theta 1 (*GSTT1*) genes. These genes encode drug-metabolizing *enzymes,*[79] and therefore polymorphisms of these genes may affect enzyme activity, altering the metabolic pathway of anti-TB drugs in the liver. Consequently, hepatic adverse reactions may occur. Toxic metabolites may also cause other adverse reactions, such as peripheral neuropathy and maculopapular eruption, although hepatotoxicity is the most widely studied outcome in pharmacogenetic studies of anti-TB drugs.

Isoniazid is the anti-TB drug for which mechanisms of the genetic contribution to ATDH have been most widely studied. Specifically, it is thought that *NAT2* acetylator status may be associated with increased risk of isoniazid-related hepatic adverse reactions, as NAT2 is the main enzyme involved in the metabolism of isoniazid in the liver. There are three phenotypes of acetylator status. Individuals who are slow *NAT2* acetylators acetylate isoniazid slowly, resulting in high plasma drug levels.[80] This may be beneficial for treatment efficacy, but slow acetylators may also experience an accumulation of toxic metabolites; there is uncertainty surrounding which specific metabolites are responsible for causing this toxicity. Fast acetylators have lower plasma drug levels, and so, treatment may be not only less effective but also less toxic, and intermediate acetylators fall between these two extremes.

Acetylator status is governed by polymorphisms in a number of alleles on the *NAT2* gene, making the genetic definition of acetylator status difficult to standardise. Classification of acetylator status according to genetic information varies between studies. For example, Azuma et al.[81] classified individuals with two copies of the *NAT2*4* allele as rapid acetylators, individuals with one copy to be intermediate acetylators, and those with no copies to be slow acetylators. Alternatively, Ho et al.[82] classified individuals with two copies of any of the following *NAT2* alleles: *NAT2*4*, *NAT2*11A*, *NAT2*12A*, *NAT2*12B*, *NAT2*12C*, *NAT2*13*; to be rapid acetylators, individuals with one of these alleles and one allele associated with slow acetylation activity (*NAT2*5*, *NAT2*6*, *NAT2*7*, and *NAT2*14*) to be intermediate acetylators, and those with two slow acetylation alleles to be slow acetylators.

Another enzyme that may be involved in the metabolism of isoniazid is the CYP2E1 enzyme, although its precise role in the pathway remains uncertain. The *Rsa*I (1053C-T; rs2031920) *Pst*I (1293G-C; rs3813867) and *Dra*I (7632T-A; rs6413432) polymorphisms are the most widely studied polymorphisms of the *CYP2E1* gene in relation to hepatotoxicity. The *Rsa*I and *Pst*I polymorphisms have been shown to be in complete linkage disequilibrium in East Asian populations[83] and to change the transcriptional activity of *CYP2E1 in vitro.*[84] It is uncertain whether the *Dra*I polymorphism itself modifies *CYP2E1* expression; the *Dra*I polymorphism has been reported to be in linkage disequilibrium with the *Rsa*I polymorphism,[85] although the *Dra*I polymorphic allele may be observed at higher frequencies than the *Rsa*I polymorphic allele.[85, 86]

Finally, it has been proposed that the GST enzyme may detoxify toxic metabolites from earlier in the metabolic pathway of isoniazid, although these specific metabolites have not been identified.[87] Two functionally important polymorphisms of *GST* genes (*GSTM1* and *GSTT1* null polymorphisms) have been widely studied in relation to isoniazid-induced hepatotoxicity, although a biological basis for an association between these polymorphisms and hepatotoxicity is uncertain. Due to the proposed impact of genetic variants of *NAT2*, *CYP2E1*, and *GST* genes on the metabolism of isoniazid in the liver, studies investigating the genetic contribution to anti-TB drug-related toxicity have mostly focused on variants of these genes.

Considering the other drugs that are administered as part of the WHO recommended regimen for drug-susceptible TB, rifampicin and pyrazinamide have also been reported to cause hepatotoxicity.[88] However, the biological mechanisms for rifampicin- and pyrazinamide-induced hepatotoxicity remain unknown.[89] The *OATP1B1*15* haplotype has been reported to be an important risk factor for rifampicin-induced liver injury,[90] while no research into genetic risk factors for pyrazinamide-induced hepatotoxicity has been reported.[91] Ethambutol has not previously been reported to cause hepatotoxicity.[88]

The aim of this systematic review and meta-analysis was to evaluate the current evidence for associations between genetic variants and anti-TB drug-related toxicity. Meta-analyses investigating the association between *NAT2, CYP2E1, GSTM1* and *GSTT1* genetic variants and hepatotoxicity had previously been published[92-102] when we started the review. However, despite the informative nature of the previously conducted meta-analyses, we identified some important methodological limitations of these reviews:

- Cai et al.,[92] Deng et al.,[94] Du et al.,[95] Li et al.,[96] Sheng et al.,[97] Shi et al.,[98] Tang et al.,[100] Wang et al.[101] and Wang et al.[102] all excluded studies if data required for meta-analysis were not included in the study report.
- Cai et al.[92] excluded three studies that were non-RCTs, and Cai et al.,[93] Deng et al.,[94] Li et al.,[96] Sheng et al.,[97] Shi et al.,[98] Sun et al.,[99] Tang et al.[100] and Wang et al.[101] all included only case-control studies. Important evidence may have been omitted from these reviews, as pharmacogenetic data may be reported in RCTs, case-control studies, or cohort studies.
- Cai et al.,[92] Cai et al.,[93] Du et al.[95] and Li et al.[96] did not assess the methodological quality of included studies. The other previously conducted meta-analyses used a checklist developed by Little et al.[103] to assess study quality.
- None of the previously conducted meta-analyses aimed to identify and synthesise data for genetic variants other than *NAT2*, *CYP2E1*, *GSTM1* and *GSTT1* genetic variants, or for outcomes other than hepatotoxicity; such exclusions may limit evidence-based recommendations.

We planned to overcome these limitations in our systematic review by: contacting study authors to obtain data required for meta-analysis when it was not included in the study report; including relevant studies regardless of their design; and performing a rigorous quality assessment of included studies. In addition, we did not exclude studies that did not report hepatotoxicity, and we aimed to identify and synthesise data for all genetic variants. Therefore, the scope of our review is wider than the previously conducted meta-analyses.

3.2 Methods

The methods described in the following section were pre-specified in a published protocol (PROSPERO registration number: CRD42017068448).[104]

Selection criteria

Types of studies

Eligible study designs were cohort studies, case-control studies and RCTs.

Types of participants

We included studies that recruited TB patients who were either already established on anti-TB treatment or were commencing treatment (at least one of isoniazid, rifampicin, pyrazinamide, or ethambutol), and who were genotyped, in order to investigate the association between genetic variants and anti-TB drug-related toxicity outcomes. Specifically, we only included studies where over 50% of included patients were TB patients

receiving anti-TB treatment; we contacted study authors to obtain data specifically for the subgroup of TB patients if these data were not available in the study publication.

Types of outcomes

Studies that measured any anti-TB drug-related toxicity outcomes were eligible for inclusion.

Search strategy

We designed the search strategy with assistance from an information specialist. MEDLINE, PubMed, EMBASE, BIOSIS and Web of Science were searched for relevant studies. Details of the number of records identified in each of these databases are provided in Appendix 1, together with the full search strategy for EMBASE. We hand-searched the reference lists of relevant studies, and contacted clinical experts to identify further eligible studies. Only studies published in English were included, but we did not restrict by year of publication or by publication status.

Study selection

We imported the results of the search into Covidence.[105] One researcher (MC) removed duplicates and scanned the study abstracts to remove obviously irrelevant studies. A sample of 10% of studies were independently screened by one of three researchers (ALJ, JK or KD). One reviewer (MC) obtained the full text for each potentially relevant study and assessed eligibility based on the eligibility criteria. A sample of 10% of studies were independently screened by one of three researchers (ALJ, JK or KD). Any disagreements between the two reviewers at both the abstract and full-text screening stages were resolved through discussion, or by consulting a third researcher if necessary.

After full-text screening, we decided to exclude studies that only reported data for associations between *HLA* genes and anti-TB drug-related toxicity outcomes. We did not include these studies as *HLA* genes have many possible variations, with some having hundreds of alleles.[106] The standard methods of meta-analysis described in Chapter 2 therefore cannot be used to synthesise data for the association between *HLA* genetic variants and treatment response outcomes. Finally, we excluded articles that included patient cohorts for whom data were also reported in other articles, if clear data for additional genetic variants were not reported.[107-110] If multiple articles reported clear data for the same set of genetic variants, we included the more recent article.

Outcomes

The primary outcome of our review was hepatotoxicity by any definition used by the original investigators. The secondary outcomes were all other toxicity outcomes reported in the included studies.

Data extraction

We pre-piloted a data extraction form, which was designed to enable collection of data on study design, participant characteristics, treatment regimen, genotype groups and outcomes. One researcher (MC) extracted data, following methods described in the HuGENet HuGE Review Handbook.[24] A second researcher (ALJ, JK or KD) independently extracted all outcome data. Any disagreements between the two reviewers were resolved through discussion, or by consulting a third researcher if necessary. We contacted study authors if relevant outcome data were not reported in the required format for meta-analysis.

We examined author lists, locations, dates of recruitment and other study characteristics to identify cases of multiple articles reporting data for overlapping or identical patient cohorts. If we suspected that this may be the case, we contacted authors to clarify whether patient cohorts were distinct. If an author clarified that multiple articles reported outcomes for the same patient cohort or overlapping cohorts, or if we suspected this based on reported study characteristics, we assigned a group identifier (GI) to these articles. Assigning this GI ensured that data for each patient cohort were only included once in any meta-analysis.

We extracted data for each article rather than each patient cohort, as key details such as sample size and genetic variants investigated often varied between multiple articles reporting data for overlapping or identical patient cohorts.

Quality assessment

One researcher (MC) used the criteria developed by Jorgensen and Williamson[18] specifically for pharmacogenetic studies, to assess the methodological quality of each included study. A second researcher (ALJ) independently assessed the quality of a sample of 10% of studies. Any disagreements between the two reviewers were resolved through discussion. We summarised the number of studies meeting each criterion in the text.

We quality assessed each article rather than each patient cohort, as different methods may have been used across multiple articles reporting data for overlapping or identical patient cohorts.

3.3 Results

Included and excluded studies

A PRISMA flowchart, showing selection and elimination of studies during the literature search, is provided in Figure 1. We identified 70 relevant articles (51 distinct cohorts of patients). We contacted 15 study authors to obtain clarification on whether patient cohorts were distinct, overlapping or identical to other patient cohorts, or for clarification on reported outcome data. We obtained responses from 11 study authors.

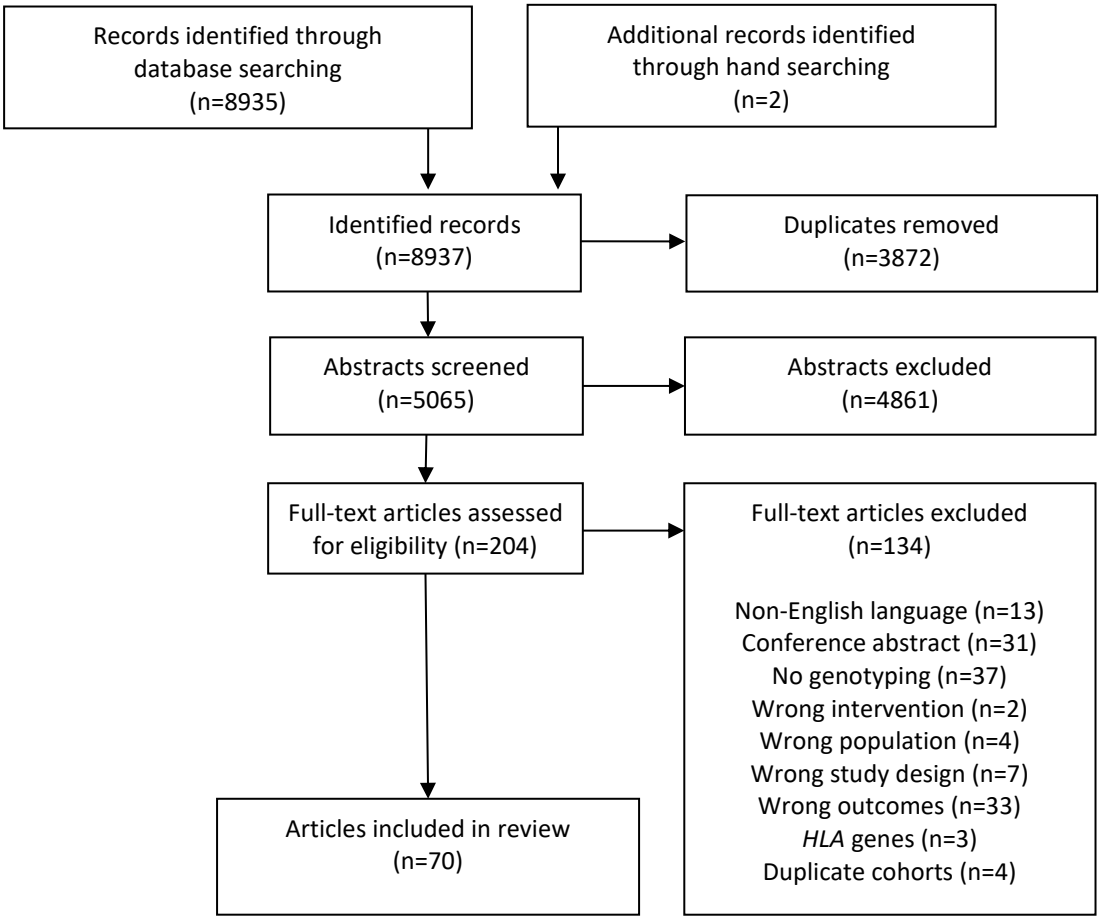

Figure 1 PRISMA flowchart showing inclusion and exclusion of studies in the systematic review

In this review, we include data from 70 articles[81, 82, 86, 90, 111-176] (51 distinct patient cohorts). Key characteristics of studies included in our systematic review are provided in Table 3. Thirty-six articles described cohort studies, although as recommended by the STROBE (STrengthening the Reporting of Observational Studies in Epidemiology) guideline,[11] we do not attempt to categorise these studies as 'retrospective cohort' or 'prospective cohort' here, as there is no clear consensus on the definitions of these terms. Thirty-three articles described case-control studies, and one article described a RCT of pharmacogenetic-guided therapy *versus* standard therapy.

Table 3 Key characteristics of studies included in the systematic review

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
ADACS: Chen 2015 [117]	Case-control	China	NR	All patients took INH 600 mg, RIF 600 mg (or 450 mg if body weight was <50 kg), PZA 2,000 mg and EMB 1,250 mg every other day for the first 2 months (re-treatment patients were injected with SM 750 mg each time simultaneously). Then with the same regimen, PZA and EMB were discontinued for primary patients, whereas PZA and SM were discontinued for re-treatment patients, the rest of the drugs were continued for another 4 months.	445	Hepatotoxicity	*SLC10A1* rs4646285 *SLCO1B1* rs4149013 *SLCO1B1* rs4149014 *SLCO1B1* rs2306283 *SLCO1B1* rs4149056 *SLCO1B1* rs2291075
ADACS: Lv 2012 [145]	Case-control	China	6-9 months	All patients took INH (600 mg), RIF (600 mg, or 450 mg if body weight was <50 kg), PZA (2,000 mg) and EMB (1,250 mg) every other day in the first 2 months; INH and RIF were continued for another 4/6 months. Re-treatment patients took SM (750 mg) every other day in the first 2 months and continued EMB for another 6 months.	445	Hepatotoxicity	*NAT2* acetylator status *NAT2* 481C-T *NAT2* 590G-A *NAT2* 857G-A
ADACS: Tang 2012 [160]	Case-control	China	6–9 months according to treatment episode	All patients took INH 600 mg, RIF 600 mg (or 450 mg if body weight was <50 kg), PZA 2,000 mg and EMB 1,250 mg every other day for the first 2 months (re-treatment patients were injected with SM 750 mg each time simultaneously). PZA and EMB were then discontinued for primary patients, whereas INH and RIF were continued for another 4 months. PZA and SM were discontinued for re-treatment patients, whereas INH, RIF and EMB continued.	445	Hepatotoxicity	*CYP2E1 Dra*I *GSTM1* null *GSTT1* null
ADACS: Tang 2013a [161]	Case-control	China	6–9 months according to treatment episode	Same as Tang 2012 [GI: ADACS]	445	Hepatotoxicity	*CYP2E1 Rsa*I *CYP2E1* rs2080672 *CYP2E1* rs915908 *CYP2E1* rs8192775 *CYP2E1* rs2515641 *CYP2E1* rs2515644

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
ADACS: Tang 2013b [162]	Case-control	China	NR	Same as Tang 2012 [GI: ADACS]	445	Hepatotoxicity	*CYP2C9* rs4918758 *CYP2C9* rs9332098 *CYP3A4* rs12333983 *CYP2C19* rs11568732 *CYP2C19* rs4986894
ADACS: Wang 2015a [167]	Case-control	China	NR	All primary/re-treatment patients with pulmonary TB were treated with a combination regimen including INH (600 mg), RIF (600 mg, or 450 mg if body weight was <50 kg), PZA (2,000 mg) and EMB (1,250 mg) for the first 2 months (re-treatment patients were injected with SM 750 mg each time simultaneously) and then with the same regimen, without PZA and EMB, for another 4 months for primary patients and with the same regimen, without PZA and SM, for another 6 months for re-treatment patients.	445	Hepatotoxicity	*HSPA1L* rs2227956 *IL6* rs2066992 *IL6* rs2069837 *IL6* rs1524107 *STAT3* rs1053004 *STAT3* rs1053023 *STAT3* rs1053005
ADACS: Wang 2015c [169]	Case-control	China	NR	Patients were given INH (600 mg), RIF (600 or 450 mg if body weight was <50 kg), PZA (2,000 mg) and EMB (1,250 mg) for the first 2 months (re-treatment patients were injected with SM 750 mg each time) and then the same regimen, without PZA and EMB, for another 4 months for primary patients and the same regimen, without PZA and SM, for another 6 months for re-treatment patients.	445	Hepatotoxicity	*IL4* rs2243289 *IL4* rs2243250 *IL4* rs2070874 *IL10* rs1800896 *IL10* rs1800871 *IL10* rs1800872

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
An 2012 [111]	Case-control	China	6 months	Daily treatment with INH, RIF, PZA and EMB for 2 months, followed by 4 months treatment with INH and RIF, with dosages calculated according to body weight: Body weight <45 kg: RIF 300 mg, INH 200 mg, PZA 1,000 mg Body weight of 45–55 kg: RIF 450 mg, INH 300 mg, PZA 1,500 mg Body weight >55 kg: RIF 600 mg, INH 400 mg, PZA 2,000 mg	208	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *NAT2* 190C-T *NAT2* 282C-T *NAT2* 341T-C *NAT2* 481C-T *NAT2* 499G-A *NAT2* 590G-A *NAT2* 803A-G *NAT2* 857G-A
Azuma 2013 [81]	RCT	Japan	6 months	6-month regimen comprising INH, RIF, PZA, and EMB/SM for the first 2 months followed by RIF and INH for 4 months. All patients started taking the standard oral dose (approx. 5 mg/kg body weight, once daily). For pharmacogenetic-guided treatment patients, dosages were adjusted based on individual *NAT2* status within 3 days. Modified daily INH doses were approximately 7.5, 5 and 2.5 mg/kg for rapid, intermediate and slow acetylators, respectively. Regarding the other drugs for the standard regimen, standard daily doses of RIF (10 mg/kg, max 600 mg/body), PZA (25 mg/kg, 1,500 mg/body), EMB (15 (20) mg/kg, 750 (1,000) mg/body), and SM (15 mg/kg, 750 mg/body) were recommended with the following dose ranges allowed at the discretion of the physician: RIF, 8–12 mg/kg; PZA, 20–30 mg/kg; EMB, 15–20 mg/kg; SM, 12–18 mg/kg.	172	Hepatotoxicity; peripheral neuropathy	*NAT2* acetylator status

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Bose 2011 [112]	Cohort	India	8 weeks	All patients received ATT (RIF, INH, EMB, and PZA) according to their body weight. All 4 drugs were given for 2 months. PZA and EMB were discontinued, while INH and RIF were continued for another 4 months. RIF, body weight (kg): mg/day: ≤35: 300; 36–50: 450; >50: 600 INH, body weight (kg): mg/day: ≤35: 200; >35: 400 PZA, body weight (kg): g/day: ≤50: 1.0; >50: 1.5 EMB: 20 mg/kg/day	218	Hepatotoxicity; adverse DIH outcome	*NAT2* acetylator status *CYP2E1 Rsa*I [a] *CYP2E1 Dra*I
Çetintaş 2008 [113]	Cohort	Turkey	NR	Patients received INH 5 mg/kg (max 300 mg/day), RIF 10 mg/kg (max 600 mg/day), PZA 25 mg/kg (max 2,000 mg/day), and EMB 15-25 mg/kg (max 1,500 mg/day).	100	Hepatotoxicity	*NAT2* acetylator status *NAT2* 191G-A *NAT2* 481C-T *NAT2* 590G-A *NAT2* 857G-A
Chamorro 2013 [114]	Cohort	Argentina	NR	The patients began a standard TB-treatment protocol for the first 2 months (INH: 5 mg/kg/day, max 300 mg/day; RIF: 10 mg/kg/day, max 600 mg/day; PZA: 20 mg/kg/day; EMB: 20 mg/kg/day), followed by INH and RIF for 4 months or more, depending on the disease severity or the presence of extrapulmonary foci.	175	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Pst*I
Chang 2012 [115]	Cohort	Taiwan	NR	First-line anti-TB medications	98	Hepatotoxicity	*NAT2* acetylator status *UGT1A1* 211G-A (rs4148323) *UGT1A1* 686C-A *UGT1A1* TA6→TA7 at the *promoter region* *UGT1A1* 1091C-T

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Chatterjee 2010 [116]	Case-control	India	NR	ATD regimens comprising of INH, RIF and PZA as per the Revised National Tuberculosis Control Program of India – DOTS	151	Hepatotoxicity	*GSTM1* null *GSTT1* null
Cho 2007 [118]	Cohort	Korea	End of treatment	All patients received oral INH (300 mg), RIF (600 mg), PZA (20 mg/kg body weight), and EMB (800 mg) daily for the first 2 months. PZA was then discontinued, while INH, RIF and EMB were continued for another 4 months.	132	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I
Costa 2012 [119]	Cohort	Brazil	NR	All patients were treated with INH (300 mg/kg/day), RIF (300 mg/kg/day), and PZA (1,500 mg/kg/day) for the first 2 months, and then INH and RIF for a further 4 months.	129	ADRs	*NAT2* acetylator status *CYP2E1 Dra*I *CYP2E1 Pst*I *GSTM1* null *GSTT1* null
Dhoro 2013 [120]	Case-control	Zimbabwe	NR	NR	Unclear - 189 on INH	Peripheral neuropathy	*NAT2* 191G-A *NAT2* 341T-C
Feng 2014 [121]	Case-control	China	6 months	Treatment with ATD regimens at the usual dosage, including 300 mg/day INH, 450 mg/day RIF, and 1,500 mg/day PZA	346	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *CYP2E1 Pst*I *GSTM1* null *GSTT1* null
Fredj 2016 [123]	Cohort	Tunisia	End of treatment	INH (5 mg/kg per day), RIF (10 mg/kg per day), PZA (25 mg/kg per day) and EMB (15 mg/kg per day) for the first 2 months, followed by INH and RIF for 4 to 7 additional months, depending on TB clinical presentation	71	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *CYP2E1 Dra*I *CYP2E1 Pst*I
Gogtay 2016 [124]	Case-control	India	NR	Anti-tuberculous medications - no further details provided	214	Hepatotoxicity	*CYP2E1 Rsa*I
GUPTA: Gupta 2013a [125]	Cohort	India	6-9 months	INH: 5 mg/kg (max 300 mg/day); RIF: 10 mg/kg (max 600 mg/day); PZA: 25 mg/kg (max 2,000 mg/day); EMB: 15-25 mg/kg (max 1,500 mg/day)	296	Hepatotoxicity	*GSTM1* null *GSTT1* null

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
GUPTA: Gupta 2013b [126]	Cohort	India	End of treatment	Combination regimen including INH 5 mg/kg (max 300 mg daily), RIF 10 mg/kg (max 600 mg daily), PZA 25 mg/kg (max 1,500 mg daily), and EMB 15–25 mg/kg (max 2,000 mg daily) for a period of 2 months and then for an additional 4 months with INH and RIF	215	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *CYP2E1 Dra*I *NAT2* 481C-T *NAT2* 590G-A
He 2015 [127]	Case-control	China	6 months	Daily 2S(E)HRZ4HR: S, streptomycin; E, ethambutol; H, isoniazid, R, rifampicin; Z, pyrazinamide; dose increased for 2 months and then consolidated for 4 months	254	Hepatotoxicity	*CYP1A1 Msp*I *GSTP1* Ile105Val
HIGHUCI: Higuchi 2007 [128]	Cohort	Japan	NR	Treated with an INH (400 mg/day) and RIF (450 mg/day) containing regimen for 6 or 9 months	100	Hepatotoxicity; skin rash; eosinophilia	*NAT2* acetylator status

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
HIGUCHI: Nanashima 2012 [148]	Cohort	Japan	NR	Treatment including INH (400 mg/day) and RIF (450 mg/day) for 6-9 months	100	Hepatotoxicity	*BACH1 (*rs2300301, rs1153285, rs2070401) *HMOX1* (rs2071746, rs2071749, rs5755720) *KEAP1 (*rs1048290, rs11545829) *MAFF (*rs2413508, rs2267373*,* rs2235264, rs4821765) *MAFK* (rs4720833, rs3808337) *NFE2L2* (rs2886161*,* rs4243387*,* rs6726395, rs2001350) *NOS2A* (rs10459953, rs3794764, rs12944039, rs11080344, rs2314810, rs3729966, rs944722, rs2255929, rs3794756) *NQO1 (*609C-T/rs1800566, rs689452, rs2917669, rs10517) *XPO1 (*rs7606167, rs11125883, rs1050567)

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Ho 2013 [82]	Cohort	Taiwan	180 days	Oral INH 300 mg, RIF 600 mg (or 450 mg if body weight was <50 kg), PZA 25 mg/kg body weight (max daily dose: 2,000 mg), and EMB 15 mg/kg body weight daily (max daily dose: 1,600 mg) for the first 2 months. PZA was then discontinued; INH, RIF, and EMB were continued for another 4 months.	348	Hepatotoxicity	*NAT2* acetylator status *NAT2* rs1495741
HUANG: Huang 2003 [129]	Cohort	Taiwan	End of treatment	The standard daily anti-TB regimen for the first 2 months included INH (300 mg), RIF (600 mg or 450 mg if body weight was <50 kg), PZA (20 mg/kg body weight), and EMB (25 mg/kg body weight). PZA was then discontinued; INH, RIF, and EMB (15 mg/kg body weight) were continued for another 4 months.	318	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I
HUANG: Huang 2007 [130]	Case-control	Taiwan	NR	NR	230 (126 on ATDs)	Hepatotoxicity	*GSTM1* null *GSTT1* null *NQO1* 609C-T (rs1800566) *SOD2* (*MnSOD*) 47T-C (rs4880)
Jung 2015 [131]	Cohort	Korea	4 weeks	INH (5 mg/kg, usually 300 mg), RIF (450 mg for <50 kg or 600 mg for ≥50 kg body weight), EMB (15 mg/kg), and PZA (20–30 mg/kg), given daily for 2 months and followed by INH and RIF with or without EMB for 4 months	206	Hepatotoxicity	*NAT2* acetylator status
Khalili 2011 [132]	Case-control	Iran	2 months	Treated daily with INH (300 mg), RIF (600 mg), PZA (20 mg/kg), EMB (15 mg/kg) for the first 2 months followed by INH and RIF daily for 4 additional months	100	Hepatotoxicity	*NAT2* acetylator status

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
KIM: Kim 2009 [133]	Case-control	Korea	NR	All patients with pulmonary TB were treated daily with a combination regimen including INH (300-400 mg daily), RIF (450-600 mg daily), EMB (600-800 mg daily) and PZA (1,000-1,500 mg daily) for 2 months and then without PZA for 4 or more following months. Doses of each drug were adjusted based on body weight.	226	Hepatotoxicity	*CYP2E1 Rsa*I *NAT2* 590G-A *NAT2* 857G-A *NAT2* rs4646244 *NAT2* rs4646267 *CYP2E1* rs2070672 *CYP2E1* rs2070673 *CYP2C9* rs4918758 *CYP2C9* rs9332096 *CYP2C9* rs1057910 *CYP2C19* rs17878465 *CYP2C19* rs4986893 *CYP2D6* rs1080983 *CYP2D6* rs1080989 *UGT1A1* 211G-A (rs4148323) *UGT1A1* rs3755319 *UGT1A1* rs2003569 *UGT1A3* rs2008584 *UGT1A3* rs6431625
KIM: Kim 2010 [134]	Case-control	Korea	NR	Same as Kim 2009 (GI: KIM)	341	Hepatotoxicity; ATD-induced cutaneous reactions	*GSTM1* null *GSTT1* null

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
KIM: Kim 2011 [135]	Case-control	Korea	NR	Same as Kim 2009 (GI: KIM)	221	ATD-induced MPE	*NAT2* 590G-A *NAT2* 857G-A *NAT2* rs4646244 *NAT2* rs4646267 *CYP2E1 Rsa*I *CYP2E1* rs2070672 *CYP2E1* rs2070673 *CYP2C9* rs4918758 *CYP2C9* rs9332096 *CYP2C9* rs1057910 *CYP2C19* rs4986893 *CYP2C19* -1418C-T
KIM: Kim 2012a [136]	Case-control	Korea	NR	Same as Kim 2009 (GI: KIM)	221	ATD-induced MPE	*ABCB1* rs1045642 *ABCB1* rs10261685 *ABCC2* 1774G-del *ABCC2* rs1885301 *ABCC2* rs717620 *ABCC2* rs2804400 *ABCC2* rs2273697 *ABCC2* rs3740070 *ABCC2* rs3740066

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
KIM: Kim 2012b [137]	Case-control	Korea	NR	Same as Kim 2009 (GI: KIM)	226	Hepatotoxicity	*ABCB1* rs1045642 *ABCB1* rs10261685 *ABCC2* 1774G-del *ABCC2* rs1885301 *ABCC2* rs717620 *ABCC2* rs2804400 *ABCC2* rs2273697 *ABCC2* rs3740070 *ABCC2* rs3740066 *SLCO1B1* rs4149013 *SLCO1B1* rs4149014 *SLCO1B1* rs2306283 *SLCO1B1* rs4149056
KIM: Kim 2012c [138]	Case-control	Korea	NR	Same as Kim 2009 (GI: KIM)	306	Hepatotoxicity	*TNF-alpha* -308G-A
KIM: Kim 2015 [139]	Case-control	Korea	NR	Same as Kim 2009 (GI: KIM)	321	Hepatotoxicity	*SOD1* rs2070424 *SOD2* (*MnSOD*) 47T-C (rs4880) *SOD3* rs1799895 *SOD3* rs2536512
Kwon 2012 [140]	Case-control	Korea	14 days	ATDs	238 (not all on TB drugs)[b]	Hepatotoxicity	*TXNRD1* rs10735393 *TXNRD1* rs4964287 *TXNRD1* rs4595619 *TXNRD1* rs10861201 *TXNRD1* rs11111997 *TXNRD1* rs4246270 *TXNRD1* rs4246271

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Lee 2010 [141]	Cohort	Taiwan	NR	All patients received oral INH 300 mg, RIF 600 mg (or 450 mg if body weight was <50 kg), PZA 200 mg/kg body weight and EMB 800 mg daily for the first 2 months. PZA was then discontinued, while INH, RIF and EMB were continued for another 4 months.	140	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Pst*I *NAT2* 341T-C *NAT2* 590G-A *NAT2* 857G-A
LEIRO: Leiro 2008 [142]	Case-control	Spain	NR	Regimens that included at least INH, RIF and PZA at the usual drug dosages (INH 5 mg/kg/day – max 300 mg/day, RIF 10 mg/kg/day – max 600 mg/day and PZA 25–30 mg/kg/day – max 2,500 mg/day)	95	Hepatotoxicity	*GSTM1* null *GSTT1* null
LEIRO: Leiro-Fernandez 2011 [143]	Case-control	Spain	End of treatment	Treatment with regimens that included INH, RIF and PZA at the usual dosages (INH 5 mg/kg/day to max 300 mg/day, RIF 10 mg/kg/day to max 600 mg/day and PZA 25–30 mg/kg/day to max 2,500 mg/day)	117	Hepatotoxicity	*NAT2* acetylator status
Li 2012 [90]	Case-control	China	NR	RIF treatment	273	Hepatotoxicity	*SLCO1B1* rs2306283 *SLCO1B1* rs4149056
Liu 2014 [144]	Case-control	China	NR	INH 10–20 mg/kg/day (max 300 mg/day), RIF 10–20 mg/kg/day (max 450 mg/day), PZA 20–30 mg/kg/day (max 1,500 mg/day), EMB 15–25 mg/kg/day, and SM 20–30 mg/kg/day (max 750 mg/day)	163	Hepatotoxicity	*GSTM1* null *GSTT1* null
Mahmoud 2012 [146]	Cohort	Tunisia	3 months	INH and RIF containing regimen	66	Hepatotoxicity	*NAT2* acetylator status *NAT2* 481C-T *NAT2* 590G-A *NAT2* 857G-A
Monteiro 2012 [147]	Cohort	Brazil	NR	Ongoing TB treatment	177	Hepatotoxicity	*GSTM1* null *GSTT1* null

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Ng 2014 [149]	Case-control	Cases recruited in the UK. Unknown for controls	NR	All patients were prescribed INH with all but 1 patient also taking additional ATDs. Patients were prescribed 300 mg INH/day.	127	Hepatotoxicity	*NAT2* acetylator status
NTUH: Wang 2011 [166]	Cohort	Taiwan	End of treatment	All participants received a standard ATT of daily INH, RIF, EMB, and PZA in the first 2 months, and daily INH and RIF for the succeeding 4 months. The regimen was modified if necessary by the primary care physician.	360	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I
NTUH: Wang 2015b [168]	Cohort	Taiwan	6 months	Daily INH, RIF, EMB, and PZA in the first 2 months, and daily INH and RIF for the next 4 months. The daily dosage of each drug was calculated by weight.	Derivation cohort: 355; validation cohort: 182	Hepatotoxicity	*NAT2* 191G-A *NAT2* 282C-T *NAT2* 341T-C *NAT2* 481C-T *NAT2* 590G-A *NAT2* 803A-G *NAT2* 857G-A *PXR* rs3814055 *PXR* rs12488820 *PXR* rs2461823 *PXR* rs7643645 *PXR* rs6785049 *PXR* rs3814057
Ohno 2000 [150]	Cohort	Japan	3 months	Initial chemotherapy always included INH (400 mg/day) and RIF (450 mg/day); the third drug used was EMB or SM.	77	Hepatotoxicity	*NAT2* acetylator status

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
POSSUELO: Brito 2014 [86]	Cohort	Brazil	NR	Daily treatment with INH, RIF, and PZA for the first 2 months, followed by INH and RIF for an additional 4 months	245	Hepatotoxicity	*CYP2E1 Rsa*I *CYP2E1 Dra*I *CYP2E1 Pst*I *GSTM1* null *GSTT1* null
POSSUELO: Possuelo 2008 [151]	Cohort	Brazil	NR	Treatment daily with INH, RIF, and PZA for the first 2 months followed by INH and RIF daily for 4 additional months. Drug dosages used were calculated according to body weight (weight <45 kg: RIF 300 mg, INH 200 mg, PZA 1,000 mg; 45–55 kg: RIF 450 mg, INH 300 mg, PZA 1,500 mg; >55 kg: RIF 600 mg, INH 400 mg, PZA 2,000 mg).	254	Hepatotoxicity; gastrointestinal ADRs	*NAT2* acetylator status
Rana 2014 [152]	Cohort	India	End of treatment	Daily ATT for the first 2 months included INH (300 mg), RIF (600 or 450 mg for body weight <50 kg), PZA (20 mg/kg body weight) and EMB (25 mg/kg body weight). After 2 months, EMB and PZA were discontinued, whereas INH and RIF were continued for an additional 4 months.	300	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *GSTM1* null *GSTT1* null
Roy 2001 [153]	Case-control	India	NR	Cases and controls were treated daily with the same drug regimen for the first 2 months: INH (300 mg), RIF (450 mg), PZA (1,500 mg) and EMB (800 mg). Subsequently, INH and RIF were continued for a further 4 months.	66	Hepatotoxicity	*GSTM1* null *GSTT1* null
Roy 2006 [154]	Cohort	India	NR	Most of the patients were treated with INH (5 mg/kg body weight per day); RIF (10 mg/kg body weight per day) and PZA (20–35 mg/kg body weight per day). A few patients with TB meningitis also received EMB (20 mg/kg body weight per day). After 2 months, PZA and EMB were discontinued and INH and RIF were continued for the next 4 months.	109	Hepatotoxicity	*CYP2E1 Rsa*I *CYP2E1 Dra*I *CYP2E1 Pst*I *CYP2E1* 96-bp

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
SANTOS: Fernandes 2015 [122]	Cohort	Brazil	NR	INH, RIF, and PZA for the first 2 months, followed by INH and RIF daily for 4 months	220	Hepatotoxicity	*CYP2B6* rs3745274 *CYP3A5* rs776746
SANTOS: Santos 2013 [155]	Cohort	Brazil	NR	Treatment with INH, RIF and PZA for the first 2 months, followed by INH and RIF daily for 4 months	270	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *CYP2E1 Dra*I *CYP2E1 Pst*I *NAT2* 282C-T *NAT2* 341T-C *NAT2* 481C-T *NAT2* 590G-A *NAT2* 803A-G *NAT2* 857G-A *CYP2E1* 96-bp
Sharma 2014 [156]	Case-control	India	NR	INH, RIF, PZA, EMB; dosages administered to patients according to body weight: RIF, body weight (kg): mg/day: ≤35: 300; 36–50: 450; >50: 600 INH, body weight (kg): mg/day: ≤35: 200; >35: 300 PZA, body weight (kg): g/day: ≤50: 1.0; >50: 1.5 EMB, 15 mg/kg/day	314	Hepatotoxicity	*CYP2E1 Rsa*I *GSTM1* null
Shimizu 2006 [157]	Cohort	Japan	3 months	Treatment with INH and RIF. The dose of INH ranged from 300 to 400 mg, and that of RIF ranged from 300 to 450 mg.	42	Hepatotoxicity	*NAT2* acetylator status
Singla 2014 [158]	Cohort	India	NR	NR	408	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *GSTM1* null *GSTT1* null

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Sotsuka 2011 [159]	Cohort	Japan	3 months	INH, RIF and PZA, plus EMB or SM during the first 2 months, followed by administration of INH and RIF plus EMB or SM during the final 4 months	144	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *CYP2E1 Dra*I *GSTM1* null *GSTT1* null
Teixeira 2011 [163]	Case-control	Brazil	NR	ATD regimens that include INH at the usual dosage (400 mg/day)	167	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *GSTM1* null *GSTT1* null
Vuilleumier 2006 [164]	Cohort	Switzerland	NR	INH 300 mg daily and vitamin B6 40 mg per day for a period of 6 months	89	Hepatotoxicity	*NAT2* acetylator status
Wang 2010 [165]	Case-control	China	NR	INH, RIF, PZA, and EMB for 2 months followed by INH and RIF for 4 months	215	Hepatotoxicity	*CYP2E1 Rsa*I *GSTM1* null
Xiang 2014 [170]	Cohort	China	2 months	All patients were prescribed INH (600 mg), RIF (600 mg, or 450 mg if the body weight was <50 kg), PZA (2,000 mg), and EMB (1,250 mg) every other day in the first 2 months. After 2 months, INH and RIF were continued for a further 4 to 6 months. Re-treatment patients in addition received SM (750 mg) every other day in the first 2 months and continued receiving EMB for another 6 months.	2244	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *GSTM1* null *GSTT1* null *NAT2* 481C-T *NAT2* 590G-A *NAT2* 857G-A
YAMADA: Yamada 2009 [171]	Cohort	Canada	9 months	INH 300 mg	170	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I
YAMADA: Yamada 2010 [172]	Cohort	Canada	9 months	INH 300 mg	170	Hepatotoxicity	*CES1* – 28 variants[c]

Study	Design	Country	Follow-up time	Drugs and dosage	Sample size	Outcomes	Genetic variants
Yimer 2011 [173]	Cohort	Ethiopia	56 weeks	All study participants received RIF based short-course chemotherapy for TB following the national TB treatment guideline.	353	Hepatotoxicity	*NAT2* acetylator status *CYP2B6* rs3745274 *CYP3A5*1:* no. of copies *ABCB1* rs1045642 *SLCO1B1* rs2306283 *SLCO1B1* rs4149056 *UGT2B7* rs7662029
Yuliwulandari 2016 [174]	Case-control	Indonesia	NR	ATT	241	Hepatotoxicity	*NAT2* acetylator status *NAT2* 282C-T *NAT2* 341T-C *NAT2* 481C-T *NAT2* 590G-A *NAT2* 803A-G *NAT2* 857G-A
Zaverucha-do-Valle 2014 [175]	Cohort	Brazil	End of treatment	600 mg/day of RIF, 400 mg/day of INH and 2 g/day of PZA for all patients with corporal weight >45 kg or adjusted for corporal weight <45 kg. After 2 months of therapy, PZA was discontinued.	131	Hepatotoxicity	*NAT2* acetylator status *CYP2E1 Rsa*I *CYP2E1 Pst*I *CYP3A4* -392A-G
Zazuli 2015 [176]	Cohort	Indonesia	1 month after last ATD therapy	All patients received FDC-ATD category I intensive phase (RIF 150 mg, INH 75 mg, PZA 400 mg and EMB 275 mg per tablet) and FDC-ATD category I continuation phase (RIF 150 mg and INH 150 mg) in 6 months of therapy. The dosage of ATDs was selected according to the patient's weight.	106	Hepatotoxicity	*PXR* rs3814055

Group identifier is provided before the name of the study where applicable.

[a] For *CYP2E1 Rsa*I polymorphism, results were only reported for hepatotoxicity, and not for adverse DIH outcome.

[b] We were unable to obtain data (from the report or through contact with trial authors) specifically for the subgroup of TB patients, so no outcome data from this study are presented in this review.

[c] The outcome data presented in this study were unclear, and we were unable to obtain clarification from trial authors; no outcome data from this study are presented in this review.

ADR: adverse drug reaction; ATD: anti-tuberculosis drug; ATT: anti-tuberculosis treatment; DIH: drug-induced hepatotoxicity; DOTS: directly observed treatment, short-course; EMB: ethambutol; FDC: fixed-dose combination; INH: isoniazid; MPE: maculopapular eruption; NR: not reported; PZA: pyrazinamide; RCT: randomised controlled trial; RIF: rifampicin; SM: streptomycin; TB: tuberculosis; WHO: World Health Organization

The ethnicity of included patients was poorly reported across the included studies; only 17 studies (24%) provided this information.[119, 138, 139, 142, 143, 147, 149, 151, 157, 164, 167, 169-172, 174, 175] Therefore, in Table 3, we have instead provided the country in which each study was conducted. The majority of studies were conducted in Asia (50/70, 71%). The remaining studies were conducted in South America (9 studies), Europe (4 studies), Africa (4 studies), North America (Canada, 2 studies) and Turkey (a transcontinental country, 1 study). The median sample size across the included studies was 218 (interquartile range 131–314).

Quality assessment

Here, summary results are presented for each criterion of Jorgensen and Williamson's quality assessment checklist,[18] grouped by the particular issue of methodological quality they relate to. Full quality assessment results for each study are provided in Appendix 2.

Choosing which genes and SNPs to genotype

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 4. The criteria relating to methods of adjustment for multiple testing and the reporting of precise *p*-values were not assessed, as we extracted raw data (i.e. the number of patients in each genotype group and outcome data for each genotype group) for inclusion in our meta-analyses, rather than the results of statistical analyses. Hence, the reporting of precise *p*-values and adjustment for multiple testing would have no impact on the validity of data included in our meta-analyses.

Table 4 Choosing which genes and SNPs to genotype: quality assessment summary

Criteria	Results
Was a literature review undertaken and the findings summarized?	Yes: 70 (100%) No: 0 (0%)
Are reasons given for choosing the genes and SNPs genotyped?[a]	Yes: 49 (70%) No: 21 (30%)
If reasons include previous association studies are key details from these provided?	Yes: 0 (0%) No: 67 (96%) N/A: 3 (4%)
If reasons include functional studies are supporting data provided?	Yes: 0 (0%) No: 45 (64%) N/A: 25 (36%)
Is method to adjust for multiple testing described?	Not assessed
Are precise *p*-values provided for all associations?	Not assessed

[a] We marked this as 'yes' if reasons were given for all genes and SNPs genotyped.
N/A: not applicable; SNP: single nucleotide polymorphism

All studies appeared to have reviewed the literature and summarised their findings, and most (49/70, 70%) provided justification for investigating the chosen genes and SNPs. However, when the justification included previous association studies or functional studies, no studies reported key details from these studies, such as odds ratios (ORs), or *p*-values.

Sample size

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 5.

Table 5 Sample size: quality assessment summary

Criteria	Results
What is the sample size?	Median: 218 (IQR: 131-314)[a]
Are details given of how the sample size was calculated?	Yes: 5 (7%) No: 65 (93%)
Are details given of the *a priori* power to detect effect sizes of varying degrees, or a justified specific effect size?[b]	Yes: 3 (4%) No: 67 (96%)

[a] Sample size for one study was unclear so we did not include this study in our calculation of median and IQR.
[b] We modified this criterion from Jorgensen and Williamson's checklist[18] to include 'or a justified specific effect size', as we deemed it satisfactory for a study to have calculated *a priori* power for a justified specific effect size, or for effect sizes of varying degrees.
IQR: interquartile range

Very few studies provided details of how the sample size was determined (5/70, 7%) or of *the a priori* power detect effect sizes of varying degrees, or a justified specific effect size (3/70, 4%).

Study design

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 6.

Table 6 Study design: quality assessment summary

Criteria	Results
What is the study design?	Cohort: 36 (51%) Case-control: 33 (47%) RCT: 1 (1%)
If study is case-control, are the two groups clearly defined?	Yes: 31 (94%) No: 2 (6%)
If study is case-control, were they genotyped in mixed batches?	Yes: 1 (3%) Unclear: 32 (97%)

RCT: randomised controlled trial

Amongst the 33 case-control studies, most (31/33, 94%) clearly defined the case and control groups. However, only one of the case-control studies reported that the two groups were genotyped in mixed batches;[127] for the remaining studies, it was unclear whether the case and control groups had been genotyped in mixed or separate batches.

Reliability of genotypes

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 7.

Table 7 Reliability of genotypes: quality assessment summary

Criteria	Results
Is the genotyping procedure described?	Yes: 69 (99%) No: 1 (1%)
Are the *primers* described?	Yes: 47 (67%)[a] No: 22 (31%)[b] Available upon request: 1 (1%)
Were quality control methods used and described?	Yes: 14 (20%)[c] Not mentioned: 56 (80%)
Were findings from quality control methods, if used, described?	Yes: 8 (11%) No: 6 (9%) N/A: 56 (80%)
Are any genotype frequencies previously reported quoted?	Yes: 38 (54%)[d] No: 32 (46%)
Were genotyping personnel blinded to outcome status?	Yes: 47 (67%) Not mentioned: 23 (33%)
If human inference required, was this independently undertaken by at least two people?	Yes: 6 (9%) Not mentioned: 5 (7%) Not known if applicable: 58 (83%) N/A: 1 (1%)

[a] For five of these studies, primers were reported for a subset of the investigated genes, and for three of these five studies, primers may be inherent to the *assays* used for genes for which primers were not reported.
[b] For 12 of these studies, primers may be inherent to the assays used.
[c] For two studies, quality control methods were described for a subset of investigated genetic variants only.
[d] For 20 studies, previously reported genotype frequencies were quoted for a subset of investigated genetic variants only, and for one study, previously reported haplotype frequencies were reported.

Almost all studies (69/70, 99%) described the genotyping procedure, but fewer studies (47/70, 67%) provided details of the primers used. Only 14 studies used and described genotype quality control checks, and only eight of these studies reported the findings of these checks. Around half of the studies (38/70, 54%) quoted previously reported genotype frequencies, although only 17 studies did this for all investigated genetic variants.

For the criterion 'Were genotyping personnel blinded to outcome status', if genotyping personnel were blinded to outcome status, or outcome status was unknown at the time of genotyping, or genotyping calls were made wholly using technology rather than personnel, we marked this criterion as 'yes' (47/70, 67%). It was not reported for 23 studies (23/70, 33%) whether genotyping personnel were blinded to outcome status.

For the criterion, 'If human inference required, was this independently undertaken by at least two people?', it was unknown for most studies (58/70, 83%) if human inference was required for genotyping, and consequently whether the criterion was applicable. For one study,[138] it was clear that the criteria did not apply, as genotyping calls were made wholly by technology rather than personnel. For the remaining 10 studies, only six reported that genotyping was independently undertaken by at least two people.

Missing genotype data

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 8.

Table 8 Missing genotype data: quality assessment summary

Criteria	Results
Is extent of missing data summarised?	Yes: 54 (77%)[a] No: 16 (23%)
If yes, are reasons for missing data given?	Yes: 4 (6%) No: 4 (6%) N/A: 62 (89%)
If yes, are checks for missingness at random reported?	No: 8 (11%) N/A: 62 (89%)
Are missing genotype data imputed?	No: 24 (34%) N/A: 46 (66%)[b]
Does paper quote number of patients contributing to each analysis?	Yes: 66 (94%)[c] No: 4 (6%)
If paper does quote number of patients contributing to analyses, does this agree to sample size?	Yes: 46 (66%) No: 20 (29%) N/A: 4 (6%)

[a] For one study, the extent of missing data was summarised for a subset of investigated genetic variants only.
[b] This criterion was not applicable to studies were there were no missing genotype data.
[c] For one study, we spotted errors in the reported numbers of patients contributing to each analysis, and we emailed authors to obtain corrected data.
N/A: not applicable

Most studies quoted the number of patients contributing to each analysis (66/70, 94%). It was therefore possible to compare the number of participants included in the analyses with the study sample size, and determine that there were no missing genotype data for 46 studies (66%). This approach assumes that no studies reported the sample size as the number of patients with genotype data as opposed to the number recruited. For the remaining 24 studies, only eight studies summarised the extent of missing data, and only four provided reasons for missing data. No studies described checking whether missing data were randomly distributed, or reported that imputations were performed for missing data.

Population stratification

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 9. The criterion relating to whether cryptic population stratification was adjusted for in the analyses was not assessed, as adjustments for cryptic population stratification would have no impact on the reliability of the raw data included in our meta-analyses.

Table 9 Population stratification: quality assessment summary

Criteria	Results
Are tests undertaken for cryptic population stratification?	Yes: 2 (3%) No: 68 (97%)
If so, are results quoted?	Yes: 2 (3%) N/A: 68 (97%)
Is cryptic population stratification adjusted for in the analyses?	Not assessed

N/A: not applicable

Two studies[122, 155] mentioned undertaking tests for population stratification; no population stratification was identified. One study[158] used a study design that ensured that the included patients were from a non-diverse ethnic group, however, cryptic population stratification may still exist, and no testing was performed. No other studies reported testing for cryptic population stratification.

Hardy-Weinberg equilibrium

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 10.

Table 10 Hardy-Weinberg equilibrium: quality assessment summary

Criteria	Results
What test is undertaken to check for HWE?	Chi-square: 27 (39%)[a] Other: 5 (7%) Test not described: 10 (14%)[b] Testing for HWE not mentioned: 28 (40%)
Where test undertaken, is *p*-value threshold applied to determine deviation from HWE quoted?	Yes: 31 (44%) No: 11 (16%) N/A: 28 (41%)
Where test undertaken, are SNPs deviating from HWE highlighted?	Yes: 26 (37%)[c] No: 16 (23%) N/A: 28 (40%)
Where test undertaken, and some SNPs found to deviate, are steps taken to explore deviation from HWE reported?	No: 6 (9%) N/A: 64 (91%)
Where test undertaken, and some SNPs found to deviate, are deviating SNPs excluded from further analysis?	Yes: 2 (3%) No: 4 (6%) N/A: 64 (91%)

[a] Testing of HWE was only mentioned for a subset of investigated SNPs in one study.
[b] Testing of HWE was only mentioned for a subset of investigated SNPs in two studies.
[c] For 20 studies, no SNPs deviated.
HWE: Hardy-Weinberg equilibrium; N/A: not applicable; SNP: single nucleotide polymorphism

Thirty-nine studies reported testing for HWE for all investigated SNPs, and a further three studies reported testing for HWE for a subset of SNPs. The remaining 28 studies did not report testing for HWE.

Among the 42 studies that mentioned testing for HWE, 31 studies quoted the *p*-value threshold applied to determine deviation from HWE. In six studies, SNPs deviating from

HWE were highlighted, and in a further 20 studies, it was clear that no SNPs deviated. Among the six studies where deviating SNPs were highlighted, no studies reported that steps were taken to explore deviation from HWE, and only two studies[119, 172] excluded deviating SNPs from further analysis.

It is important to note that a systematic reviewer would be able to perform their own tests for deviation from HWE if the trial authors have reported (or provided on request) the number of individuals in each genotype group. However, if problems with the data were flagged by HWE testing, it is rare that a systematic reviewer would have access to the necessary data to explore potential reasons for deviation from HWE, such as population stratification or genotyping errors.

Mode of inheritance

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 11. Two criteria were not assessed ('If no mode of inheritance is assumed does the paper explain limitations of this?' and 'If several analyses are undertaken under different assumptions, are they adjusted for multiple testing?') as satisfying these criteria would have no impact on the validity of the raw data included in our meta-analyses.

Table 11 Mode of inheritance: quality assessment summary

Criteria	Results
Is a specific mode of inheritance assumed? If so which?	Additive: 1 (1%) Dominant: 20 (29%) Different modes for different genetic variants: 14 (20%) Multiple modes assumed: 16 (23%) N/A: 2 (3%)[a] None: 9 (13%) Unclear: 8 (11%)
Is justification provided for assumptions made regarding mode of inheritance?	Yes: 5 (7%) No: 55 (78%) N/A: 10 (14%)[b]
If no mode of inheritance is assumed does the paper explain limitations of this?	Not assessed
If several analyses are undertaken under different assumptions, are they adjusted for multiple testing?	Not assessed

[a] This criterion was not applicable to two studies where no homozygous mutant-type individuals were identified.
[b] This criterion was not applicable to eight studies for which the assumptions made regarding mode of inheritance were unclear, and two studies where no homozygous mutant-type individuals were identified.
N/A: not applicable

Thirty-five studies made a specific assumption (i.e. assumed an additive mode, a dominant mode, or different modes for different SNPs) regarding the underlying mode of inheritance.

Only five studies provided justification for the assumptions made regarding mode of inheritance (i.e. no mode, or specific mode/modes).

Choice and definition of outcomes

The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 12.

Table 12 Choice and definition of outcomes: quality assessment summary

Criteria	Results
Does the paper clearly define all outcomes investigated?	Yes: 62 (89%) No: 8 (11%)
Is justification provided for the choice of outcomes?	Yes: 58 (83%) No: 12 (17%)
Are results shown for all outcomes mentioned for all variants investigated in the paper?[a]	Yes: 56 (80%) No, but the authors provided reasons to measure some outcomes only for certain variants: 4 (6%) No: 10 (14%)

[a] We modified this criterion from Jorgensen and Williamson's checklist[18] to include 'for all variants investigated in the paper', as we thought it was important to highlight studies that only reported some outcomes for certain variants.

Most studies clearly defined the outcomes investigated (62/70, 89%), and provided justification for the choice of outcomes (58/70, 83%). For the studies that did not provide justification for the choice of outcomes, we judged the outcomes to be in line with the main aims of the study. Most studies presented results for all pre-specified outcomes for all variants investigated in the study (56/70, 80%), or provided rationale for having measured some outcomes for certain variants only (4/70, 6%).

There was large variation in definition of hepatotoxicity across the included studies (Table 13). Of the 66 studies reporting hepatotoxicity data, two did not provide a definition, three provided vague definitions, and the remaining 61 studies provided 43 different definitions. Definitions of other toxicity outcomes were generally not sufficiently detailed (Table 14).

Table 13 Definitions of hepatotoxicity in the included studies

Study	Definition
An 2012	An increase of > twice the ULN range in ALT or conjugated bilirubin levels or a concurrent increase in AST levels, according to the criteria developed at an international consensus meeting[177]
Azuma 2013	INH-induced hepatotoxicity was assessed according to the diagnostic criteria of the Manual for Serious Side-Effects of Drug-induced Liver Injury from the Ministry of Health, Labor and Welfare of Japan.[178, 179] In brief, hepatocellular injury was defined as a >2-fold increase in the ULN concentration of ALT alone or a serum ALT ratio/ALP ratio >5, where the ALT ratio=ALT value/ULN of ALT, and ALP ratio=ALP value/ULN of ALP. Cholestatic injury was defined as an increase of > twice the ULN range of ALP or a serum ALT ratio/ALP ratio <2. Mixed injury was defined as a serum ALT ratio/ALP ratio of between 2 and 5. Causality assessments showed a relationship to the INH administration if the total score was >3, i.e. 'possible'.
Bose 2011, Yimer 2011	Defined according to the international consensus criteria.[177] Liver biochemical parameters > twice the ULN value was considered as hepatotoxicity.
Brito 2014, Possuelo 2008 (both GI: POSSUELO)	Criteria for the diagnosis of hepatotoxicity was an elevation in liver function tests, AST and/or ALT of >3-fold the ULN (reference values: 40 and 65 U/L, respectively) and/or total bilirubin up to >2.0 mg/dL in the presence of gastrointestinal symptoms such as anorexia, nausea, vomiting and/or jaundice, with serum ALT level normalisation after anti-TB drug discontinuation.
Çetintaş 2008	Hepatotoxicity criteria were defined as follows: • An increase in AST and ALT levels of >3-fold above normal or >5-fold above starting level or, • a greater than normal increase in ALT and AST levels together with hepatitis symptoms or, • a high bilirubin level.
Chamorro 2013	Hepatotoxicity was defined as when serum transaminase concentrations were ≥3 times the ULN (normal values: AST between 0–32 IU/L and ALT between 0–31 IU/L) with report of jaundice (bilirubin normal values: 0–1 mg/dL) and/or hepatitis symptoms (nausea, vomiting, abdominal pain), or >5 times the ULN with or without symptoms.
Chang 2012	'Defined according to the classification of the CIOMS'[177] - no further information provided
Chatterjee 2010	Hepatotoxicity was defined as: a rise of serum (i) ALT level ≥3 times the ULN or (ii) total bilirubin >1.0 mg⁄dL or, (iii) ALT <3 times the ULN but associated with severe anorexia, nausea, vomiting.
Chen 2015, Tang 2012, Tang 2013a, Tang 2013b, Wang 2015a, Wang 2015c (all GI: ADACS)	Hepatotoxicity was defined as: (i) an increase to > twice the ULN in ALT or a combined increase in AST and total bilirubin, provided one of them was > twice the ULN; (ii) causality was assessed as certain, probable or possible based on the WHO Uppsala Monitoring Center criteria.[180]
Cho 2007, Jung 2015, Lee 2010	An increase in serum ALT level to > twice the ULN value after anti-TB treatment, according to the criteria of drug-induced liver injuries developed by the international consensus meeting[177]
Feng 2014, Teixeira 2011,	An increase in serum transaminase values to >3 times the ULN value (40 IU/L ALT in Feng 2014) and symptoms compatible with hepatitis
Fernandes 2015, Santos 2013 (both GI: SANTOS)	An increase in serum ALT level >3 times the ULN after treatment
Fredj 2016	The causality of drug-induced hepatotoxicity was determined according to the report of an international consensus meeting.[177] These criteria include (i) an

Study	Definition
	increase of liver transaminases levels of > twice the normal value (<40 UI per litre) for AST and ALT (ii) an improvement of this pattern after the drug withdrawal, and (iii) the absence of alternative causes of this disorder.
Gogtay 2016	Patients with symptoms/signs such as anorexia, nausea, vomiting, malaise, icterus, and raised serum aminotransferase levels > twice the ULN or >5 times the ULN without clinical symptoms were considered to have hepatotoxicity.
Gupta 2013a (GI: GUPTA), Gupta 2013b (GI: GUPTA), Nanashima 2012 (GI: HIGUCHI)	An increase in ALT > twice the ULN or a combined increase in AST and bilirubin levels, provided one of them is > twice the ULN, was defined as hepatotoxicity according to the international consensus meeting.[177] ULN ranges - ALT: normal ≤42 IU/L; AST: normal ≤33 IU/L; total bilirubin: normal ≤1.5 mg/dL.
He 2015	Defined according to the Danan criteria promulgated in 1990.[177, 181] No further information was provided.
Higuchi 2007 (GI: HIGUCHI)	Defined according to the criteria of the international consensus meeting,[177] i.e. development of an increase in serum ALT level ≥ 2 times the ULN (≤42 IU/L), or a combined increase of >2*N* in serum AST (*N* ≤33 IU/L) and total bilirubin (*N* ≤1.5 mg/dL)
Ho 2013	An elevation in liver-function tests, AST and/or ALT of >5 times the ULN; or AST and/or ALT of >3 times the ULN in the presence of symptoms such as nausea, vomiting, poor appetite, abdominal pain, or jaundice; or AST and/or ALT of >3 times the ULN in the presence of total bilirubin of >2 times the ULN.
Huang 2003 (GI: HUANG)	Hepatotoxicity was diagnosed as: (i) an increase in serum ALT level > twice the ULN during treatment, according to the criteria established by the international consensus meeting;[177] (ii) negative serum hepatitis B virus surface antigen, IgM antibody to hepatitis A virus, and antibody to hepatitis C virus when ALT or AST is elevated; (iii) without any other major hepatic or systemic diseases that may induce elevation of liver biochemical tests, such as alcoholic liver disease, autoimmune hepatitis, congestive heart failure, hypoxia, and bacteraemia; and (iv) a causality assessment score >5 (classified as 'probable' or 'highly probable' drug-induced hepatitis), as derived from the international consensus meeting.[177]
Huang 2007 (GI: HUANG)	The inclusion criteria of hepatotoxicity patients were based on the suggestion of the Drug-induced Liver Injury Network[182] as follows: (i) an increase in serum ALT or AST level >5 times the ULN, or an elevation in ALP > twice the ULN, confirmed on at least two consecutive blood draws; (ii) if baseline ALT, AST or ALP are known and elevated, then ALT or AST >5 times the baseline value, or ALP > twice the baseline level on at least 2 consecutive blood draws; (iii) any elevation of serum ALT, AST, or ALP, associated with an increased serum total bilirubin (>2.5 mg/dL), in the absence of prior diagnosis of liver disease, Gilbert's syndrome, or evidence of haemolysis; (iv) a causality assessment score[178] >5.
Khalili 2011	Hepatotoxicity was defined as: (i) increased levels of liver transaminases >3 times the normal value (<40UL-1 for AST and ALT) with any other clinical signs and symptoms; or (ii) elevation of transaminases >5 times the ULN, if patients had no symptoms. For evaluation of causality, The Roussel Uclaf Causality Assessment Method scoring system was used.[178]
Kim 2009, Kim 2012b (both GI: KIM)	An elevation in the serum levels of ALT > twice the ULN range (≤40 U/mL) during treatment and normalisation of these values after cessation of medication according to the criteria from the international consensus meeting[177]
Kim 2010 (GI: KIM), Kim 2012c (GI: KIM), Leiro 2008 (GI: LEIRO)	An elevation in the serum level of ALT or AST >3 times the ULN range (≤40 U/L) during treatment, according to the American Thoracic Society guidelines[183]

Study	Definition
Kim 2015 (GI: KIM)	Elevation in the serum levels of AST or ALT of >3 times the ULN during treatment and normalisation of these values after cessation of treatment
Kwon 2012	AST level ≥ULN and an (AST/ULN)/(ALP/ULN) ratio ≥5
Leiro-Fernandez 2011 (GI: LEIRO)	Increase in serum transaminase (either AST or ALT) to values >3 times the ULN (i.e. >120 IU/L) at any time during the treatment period
Li 2012	Serum ALT levels ≥3 times the ULN, and/or serum bilirubin levels ≥2 times the ULN
Liu 2014	The diagnostic criteria of hepatotoxicity were based on international consensus:[177, 178, 184] (i) serum ALT > twice the ULN (40 IU/L); or (ii) serum direct bilirubin > twice the ULN (6.8 µmol/L); or (iii) increases of serum AST (40 IU/L), ALP (220 IU/L), and total bilirubin (19.0 µmol/L); moreover, one of them > twice the ULN; or (iv) any index mentioned above > the ULN and associated with liver damage symptoms, such as skin or sclera yellow dye, severe anorexia, nausea, vomiting, fever, rash, itching.
Lv 2012 (GI: ADACS)	Hepatotoxicity was designated as: • An increase of > twice the ULN value in ALT or a combined increase in AST and total bilirubin provided one of them is > twice the ULN. In this study, the ULN of ALT, AST and total bilirubin were 40 U/L, 40 U/L and 19 µmol/L, respectively. • Causality assessment result was highly probable, probable or possible based on the CIOMS scale.[185]
Mahmoud 2012	Hepatotoxicity was diagnosed as: • an increase in serum ALT level > twice the ULN during the treatment, according to the criteria established by the international consensus meeting;[177] • negative serum hepatitis B virus surface antigen, IgM antibody to hepatitis A virus, and antibody to hepatitis C virus when ALT or AST was elevated; • without any other major hepatic or systemic diseases that may induce elevation of liver biochemical tests, such as alcoholic liver disease, autoimmune hepatitis, congestive heart failure, hypoxia, and bacteremia; • when the French imputability score[186] was classified as ‘‘probable’’ or ‘‘likely’’ or ‘‘certain’’.
Monteiro 2012, Zaverucha-do-Valle 2014	Increase in serum ALT levels beyond twice the ULN (ALT ≥42 IU/L), or at least a 2-fold increase in ALT initial levels for those patients with a baseline ALT of >84 IU/L, during the treatment period
Ng 2014	All cases met at least one of the following biochemical criteria for enrolment into this study: (a) ALT >5 times the ULN, (b) ALP > twice the ULN, or (c) ALT >3 times the ULN and bilirubin > twice the ULN.
Ohno 2000	AST and/or ALT >1.5 times the ULN and 2 times the level observed before treatment administration
Rana 2014	Hepatotoxicity was defined according to international consensus criteria.[177] Patients with a rise in serum AST or ALT levels ≥5 times the ULN, irrespective of symptoms and serum bilirubin levels, or patients with rise in serum AST or ALT levels ≥ twice the ULN with hyperbilirubinaemia and an absence of serological evidence of infection with hepatitis viruses (A, B, C and E) were considered as having hepatotoxicity.
Roy 2001	Defined according to the international consensus criteria[177] with regard to chronology and causation for drug-induced liver diseases. However, only icteric hepatitis cases (serum bilirubin >3.0 mg/dL), among those fulfilling the above criteria, were included in the study. No further information was provided.

Study	Definition
Roy 2006	NR
Sharma 2014	Hepatotoxicity was diagnosed if any one of criteria (i), (ii) or (iii) were present along with criteria (iv) and (v). The criteria were: (i) an increase of 5 times the ULN (50 IU/L) of serum AST and/or ALT levels on one occasion or >3 times (>150 IU/L) on 3 consecutive occasions; (ii) serum total bilirubin level >1.5 mg/dL; (iii) any increase in serum AST and/or ALT above pre-treatment values, together with anorexia, nausea, vomiting and jaundice; (iv) absence of serological evidence of infection with hepatitis viruses A, B, C or E; and (v) improvement in liver function (serum bilirubin <1 mg/dL, AST and ALT <100 IU/L) after the withdrawal of anti-TB drugs.
Shimizu 2006	ALT and/or AST level > twice the institutional ULN, according to the modified criteria of the international consensus meeting for drug-induced liver disorders.[177] The ULN for AST was 33 IU/L and that for ALT was 42 IU/L.
Singla 2014	International consensus criteria[177] define hepatotoxicity as development of > twice the ULN value of ALT and AST. The ULN values used in this study were 35 U/L ALT and 40 U/L AST.
Sotsuka 2011	The severity of hepatotoxicity (hepatotoxicity A-D) was judged by the increase in either AST or ALT levels from the ULN range (AST, 33 U/L; ALT, 42 U/L): hepatotoxicity A, above the upper limit and <2-fold increase; hepatotoxicity B, 2- to 3-fold increase; hepatotoxicity C, 3- to 4-fold increase; hepatotoxicity D, >4-fold increase. Results for grades B–D of hepatotoxicity were used in this review as clinical opinion was that the hepatotoxicity A patients would not have met the criteria for hepatotoxicity in many of the other studies included in this review.
Vuilleumier 2006	Criteria for the diagnosis of INH-H consisted of a 4-fold elevation in AST and/or ALT levels above the upper reference limit (168 IU/L) with or without symptoms. Clinical diagnostic scales were used to assess the likelihood of drug involvement when INH-H was suspected.[187] Based on the clinical diagnostic scales, causality assessment of INH-H was then categorised as definite (score >17), probable (14–17), possible (10–13), unlikely (6–9) or excluded (<6). INH-H with possible to probable scores were considered for statistical analysis; unlikely scores were still considered when no other factor was identifiable.
Wang 2010	The selection criteria for hepatotoxicity were as follows: (i) ALT ≥ twice the ULN; (ii) increased AST/ALT/serum proteins (i.e. liver damage based on an increase in ALT or bilirubin ≥ twice the ULN, or an increase in AST, ALP and total bilirubin with at least one of these being ≥ twice the ULN); (iii) negative for hepatitis A antibody, hepatitis B surface antigen and hepatitis C marker; (iv) no other factors influencing the levels of AST/ALT/serum proteins, such as alcohol-induced liver disease, hypoxia, auto-immune disease, congestive heart failure and bacteraemia; and (v) causality assessment score >5.
Wang 2011 (GI: NTUH)	Among patients with normal baseline liver function (including AST, ALT, and total bilirubin), hepatotoxicity was defined as increased serum AST and/or ALT >3 times the ULN in symptomatic, or >5 times the ULN in asymptomatic patients. Among those with increased baseline AST and/or ALT, hepatotoxicity was defined as increased serum AST and/or ALT >1.5 times the baseline level.
Wang 2015b (GI: NTUH)	Hepatotoxicity was defined as increased serum AST and/or ALT >3 times the ULN in symptomatic patients, or >5 times the ULN in asymptomatic patients. The diagnosis of INH- or RIF-induced hepatitis required a positive re-challenge test (at least doubling of serum AST or ALT levels and recurrence of clinical symptoms of hepatitis after re-challenge), whereas PZA-induced hepatitis was diagnosed either by a positive re-challenge test or by exclusion. Results are presented for overall drug-induced hepatotoxicity and INH-induced hepatotoxicity separately. In this review, we used the results for overall drug-induced hepatotoxicity as our review focusses on hepatotoxicity induced by any anti-TB drug.
Xiang 2014	ALT, AST or bilirubin value > twice the ULN value. The ULN used in the study was 40 U/L for ALT, 40 U/L for AST, and 19 mmol/L for total bilirubin.

Study	Definition
Yamada 2009, Yamada 2010 (both GI: YAMADA)	Increase in serum AST level > twice the ULN during 9 months of treatment with INH according to the criteria of the international consensus meeting in Paris;[177] normalisation of serum AST level after discontinuation of isoniazid; and a causality assessment score[178] >8, corresponding to the category of highly probable hepatotoxicity
Yuliwulandari 2016	NR
Zazuli 2015	ALT and/or AST levels above the normal threshold on the 2nd, 4th and 6th months of monitoring during TB treatment

ALP: alkaline phosphatise; ALT: alanine aminotransferase; AST: aspartate aminotransferase; CIOMS: Council for International Organizations of Medical Science; IgM: immunoglobulin M; INH: isoniazid; NR: not reported; PZA: pyrazinamide; RIF: rifampicin; TB: tuberculosis; ULN: upper limit of normal; WHO: World Health Organization

Table 14 Definitions of other toxicity outcomes in the included studies

Outcome	Study	Outcome definition
Peripheral neuropathy	Azuma 2013	NR
	Dhoro 2013	NR
Adverse DIH outcome	Bose 2011	'16 [patients] showed an adverse outcome of anti-TB treatment hepatotoxicity with icterus, severe nausea, and vomiting'. No further details reported
ADRs	Costa 2012	The presence of at least one of the following symptoms during the follow-up period: gastric, joint, neuromuscular, or skin reactions; and hepatotoxicity (in accordance with the criteria of drug-induced liver injuries developed by the international consensus meeting)[183]
Skin rash	Higuchi 2007 (GI: HIGUCHI)	NR
Eosinophilia	Higuchi 2007 (GI: HIGUCHI)	The presence of >450 eosinophils/mL
ATD-induced cutaneous reactions	Kim 2010 (GI: KIM)	The development of any cutaneous symptom or skin lesion after receiving ATD medication
ATD-induced MPE	Kim 2011, Kim 2012a (both GI: KIM)	The development of MPE after receiving first-line ATD and the disappearance of MPE after discontinuing ATD
Gastrointestinal ADRs	Possuelo (2008) (GI: POSSUELO)	Anorexia, nausea, vomiting, and/or abdominal pain

ATD: anti-tuberculosis drug; ADR: adverse drug reaction; DIH: drug-induced hepatotoxicity; GI: group identifier; MPE: maculopapular eruption; NR: not reported; TB: tuberculosis

Treatment adherence

As previously outlined in Section 2.2, we use the term 'adherence' as opposed to 'compliance', which is used in Jorgensen and Williamson's[18] checklist. The assessment criteria for this issue of methodological quality, and the results of our assessment are provided in Table 15.

Table 15 Treatment adherence: quality assessment summary

Criteria	Results
Is adherence to treatment measured?	Yes: 11 (16%) No: 56 (80%) N/A: 3 (4%)[a]
If adherence is measured, are adjustments for non-adherence made in the analyses?	Yes: 7 (10%) No: 2 (3%) N/A: 61 (87%)

[a] This criteria was not applicable to studies that reported that treatment was administered by directly observed therapy, short-course (DOTS).
N/A: not applicable

A small proportion of studies (11/70, 16%) mentioned assessing treatment adherence. Three studies[116, 158, 176] reported that treatment was administered by directly observed therapy, short-course (DOTS), so it was unnecessary to measure adherence. Of the 11 studies that reported assessing adherence, two studies[151, 172] did not report adjusting the analyses for adherence. It was not necessary to adjust the analyses of two studies that measured adherence,[118, 119] as patients were reported to have good treatment adherence.

3.4 Discussion

Characteristics of included studies

The included studies varied greatly in terms of how hepatotoxicity was defined (43 different definitions across 66 studies). Jorgensen et al.[48] and Contopoulos-Ioannidis et al.[49] have previously made similar observations about the variability of outcome definitions across pharmacogenetics studies. It would be beneficial for consensus to be reached between experts in this clinical area on the definitions of outcomes that are commonly reported in studies of anti-TB drug-related toxicity. We observed that several studies provided a reference to an international consensus meeting for drug-induced liver disorders[177] when defining hepatotoxicity, but the definitions were still not consistent between this subset of included studies.

Since this commonly referred to international consensus meeting[177] was conducted, it has been argued that the recommended threshold for drug-induced liver injury (DILI) as an elevation in the serum concentration of alanine aminotransferase (ALT), conjugated bilirubin, or alkaline phosphatase exceeding two times the upper limit of normal (ULN), may be too low.[188] Such a low threshold may lead to unnecessary investigations of temporary and modest ALT elevations that would revert to baseline even if drug therapy were to be continued, and potentially withdrawal of efficacious drug treatment. Non-alcoholic fatty liver disease may also lead to elevations in liver parameters, and withdrawal of drugs suspected to be hepatotoxic in these instances would also be unnecessary. For these

reasons, an international DILI expert working group recommended raising the cut-off level of ALT elevation to five times the ULN in order to exclude clinically unimportant DILI, and elevations in liver parameters unrelated to treatment with potentially hepatotoxic drugs.[188] The working group developed an algorithm to firstly identify cases of DILI, and then to assess the pattern, causality, severity and chronicity of the identified DILI. These recommendations ought to ensure consistency when classifying the wide spectrum of conditions that constitute DILI.

The COMET (Core Outcome Measures in Effectiveness Trials) Initiative[189] encourages the development of standardised sets of outcomes, known as 'core outcome sets' for specific conditions. If a core outcome set exists for a particular condition, all trials of this condition ought to measure and report data for these outcomes. If a core outcome set were to be developed for studies of anti-TB drug-related toxicity, outcomes would be comparable between studies and synthesis of data from these studies would become more clinically meaningful.

We also observed that evidence from studies conducted in Africa, where TB is endemic, was very limited. Only four studies included in this review were conducted in Africa; two were conducted in Tunisia,[123, 146] one in Ethiopia,[173] and one in Zimbabwe.[120] Therefore, most of the evidence included in this review is not representative of the global population most affected by TB. Furthermore, genotype frequencies vary greatly across the African continent,[190] and it is therefore often not appropriate to extrapolate results from one African ancestral group to another. To better understand how the relationship between genetic variants and anti-TB drug-related toxicity outcomes varies across different African populations, more pharmacogenetic studies are required from this setting. However, it is important to note that investigating anti-TB drug-related toxicity in African countries may be complicated by the fact that a considerable proportion of TB patients in these settings are likely to be co-infected with HIV. It can therefore be difficult to determine whether adverse events such as hepatotoxicity have been caused by anti-HIV drugs, anti-TB drugs, or both.

Finally, the lack of reporting of ethnicity of participants in the included studies was particularly problematic. As previously discussed in Section 2.3, the HuGENet HuGE Review Handbook[24] recommends that meta-analyses of genetic association studies are stratified by ethnicity, and that pooling of results should only be performed if effect estimates for different ethnic groups appear sufficiently similar. Based on the small number of studies

that reported the required information, we concluded that it would not be possible to stratify our meta-analyses by ethnicity.

Quality assessment of included studies

A key feature of our systematic review was the quality assessment using the criteria developed by Jorgensen and Williamson[18] specifically for pharmacogenetic studies. The primary purpose of this assessment was to highlight potential issues that may impact the internal validity of the results of the studies included in our review. This would allow us to determine the appropriate level of caution to use when drawing conclusions based on the results of the included studies, and syntheses of these studies.

It is important to note here that Jorgensen and Williamson's tool[18] is not explicitly designed to comprehensively assess risk of bias (see Section 2.2 for discussion of the differences between assessing risk of bias and assessing methodological quality). However, it is feasible that many of the issues raised by Jorgensen and Williamson[18] would impact the internal validity of a pharmacogenetic study's results.

We did not assess some criteria that clearly would have no impact on the reliability of the data included in our systematic review, such as whether results were adjusted for multiple testing, or cryptic population stratification, as we used only raw data in our meta-analyses. However, we did assess some criteria that would be unlikely to have an impact on the reliability of the data included in our systematic review: for example, we assessed the criteria relating to sample size, as this highlighted the benefits of performing meta-analysis for this set of studies.

Systematic reviewers using the tool ought to be aware that some of the listed criteria would not have implications for the internal validity of a pharmacogenetic study's results, and therefore would not necessarily need to be highlighted in the review as particular issues of concern. Importantly, there are also some potential sources of bias not addressed by the checklist; these are discussed further under the "Limitations" heading of this section. Overall, future work developing Jorgensen and Williamson's checklist[18] further, to create a tool that could be used to comprehensively address risk of bias in systematic reviews of pharmacogenetic studies would undoubtedly be beneficial.

The quality of included studies was variable, with many areas of concern. An important issue relating to the conduct of pharmacogenetic studies is the possibility that authors may selectively report results based on their statistical significance or perceived importance. Selective reporting may occur in relation to genetic variants, assumed mode of inheritance,

and/or outcomes. Some studies did not provide rationale for investigating all genes and SNPs investigated; a clear rationale would provide reassurance that results were not selectively reported. Furthermore, very few studies justified their assumed mode of inheritance, or the lack of assumed mode. A reader may therefore have concerns that multiple analyses might have been performed and the results selectively reported. We identified that justification for the choice of outcomes was more widely reported, and most studies presented results for all listed outcomes for all variants investigated in the study, or provided rationale for having measured some outcomes for certain variants only.

Most studies were significantly smaller than typically required to provide sufficient power.[18] The typically small sample sizes of the studies that form the evidence base suggest that performing meta-analyses to increase power to detect genetic associations is essential in this area of research.

All but one of the studies employed case-control or cohort designs; the remaining study was a RCT[81] where patients were randomised to receive genotype-guided therapy or standard therapy. For the case-control studies, only one study reported that cases and controls were genotyped in mixed batches; for all other case-control studies, we do not know whether the genotyping quality was comparable between cases and controls.[18] Furthermore, few studies provided information regarding genotyping quality procedures, so it is difficult to assess the likelihood of bias due to incorrect genotype allocation in the included studies.

The fact that no studies described checking that missing data were missing at random is a concern; missing genotype data are unlikely to be missing at random, as heterozygotes are notoriously more difficult to call than homozygotes.[18] Analyses including only individuals with non-missing genotype data may be biased if the missing data are not missing at random.

Very few studies reported on testing for population stratification, and so the results reported in most studies are at risk of confounding by population stratification. Testing for deviation from HWE can highlight that there may be probems with the data, such as cryptic population stratification or genotyping errors;[18] however, a considerable number of studies did not report any testing of HWE, and no studies took steps to explore reasons for SNPs deviating from HWE.

Finally, most studies did not report that treatment adherence had been measured, so it is not possible to consider the impact that adherence may have had on outcomes in these studies.[18]

It is important to note here, we could only judge a study to have met a particular criterion if sufficient relevant information was provided in the study publication. Lack of information for a particular criterion suggests that the study may be of poor methodological quality, as we do not have enough detail to judge the study as high quality. However, lack of information does not necessarily indicate that a study is poor quality, or at high risk of bias. For example, in 30% of studies included in our review, authors did not provide reasons for choosing the genes and SNPs investigated in their study. However, it is entirely possible that these authors did have valid reasons for choosing the genes and SNPs investigated, but these reasons were not explicitly stated in the study publication. For many criteria in the checklist, it is feasible that if more information on methodology was available in the study publications, our overall confidence in the findings of these studies would improve.

Limitations

A limitation of this systematic review is that it was not possible to adhere completely to the HuGENet HuGE Review Handbook[24] guidance on searching for studies. Due to a lack of expertise and funding, we did not include studies published in languages other than English, nor did we search grey literature for relevant studies. Therefore, there may be relevant evidence that was not included in our systematic review from studies published in non-English languages and unpublished studies. We were however able to check the reference lists of existing reviews in our topic area, which can sometimes identify unpublished studies. In this case, no unpublished studies were identified.

Furthermore, during the study selection process, it was necessary to make some additions to the pre-specified eligibility criteria, as outlined in Section 3.2. When we designed the search strategy, we did not anticipate that we would identify multiple articles reporting data for the same genetic variants and same patient cohort. We also did not anticipate that we would identify studies investigating *HLA* alleles, that we would be unable to apply the standard methods of evidence synthesis to. Although the decisions to exclude these studies were post-hoc, we made these decisions by considering whether it would be sensible to include these studies in our review, rather than based on the results of these studies. We are therefore confident that these exclusions would not introduce bias to the results of our systematic review and meta-analysis.

Due to the number of references identified by the search strategy and the number of studies included in this review, dual abstract screening, full text assessments and quality assessments were only performed for a sample of the included studies. At the abstract screening stage, if there was any uncertainty about the relevance of an abstract, the abstract would be included. At the full text eligibility assessment and quality assessment stages, agreement was good and all discrepancies were minor. All outcome data were extracted independently by two reviewers. Therefore, we believe that any errors during study selection, quality assessment and data extraction are likely to be minimal and unlikely to influence the results of our review.

As outlined in Section 2.5, there are some additional sources of bias not listed in Jorgensen and Williamson's checklist[18] that could impact the validity of the results reported by studies included in our systematic review. It was outside the scope of this work to develop the tool further and formally assess each of these additional issues. However, considering studies that recruited patients already established on TB treatment, there was no reason to believe that any of these studies selectively recruited patients who had not already experienced early adverse drug reactions. Studies that recruited patients already established on TB treatment employed case-control designs, and therefore inherently provided data on relevant events occurring prior to recruitment. Considering whether results of genotyping tests performed prior to recruitment may have introduced selection bias to the included studies, it was often difficult to determine precisely when genotyping was performed in the included studies (i.e. pre- or post-recruitment), and so we consider it may be difficult for future review authors to assess the risk of bias in relation to this issue.

For the case-control studies included in this systematic review, cases and controls were almost always recruited from the same underlying patient population, and only differed in terms of outcome status. One exception to this was a case-control study[149] that used population controls; there was limited information about the underlying population that these controls were sampled from. Furthermore, two studies[90, 120] did not provide clear definitions of case and control groups. For all other studies, we did not have concerns about differences between case and control populations in terms of factors that may influence treatment outcome.

Regarding the issues relating to measurement of outcomes, limited information was available in the included study reports on blinding of outcome assessors. It is possible that detection bias may impact the results of the studies included in this review. Finally, it is

important to note that without access to the trial protocols or pre-specified analysis plans for each study (which were unavailable for the majority of included studies), it is very difficult to rule out the possibility of selective reporting of outcomes. Study authors may have measured more outcomes than are listed in the published paper. Indeed, the wide variety of possible definitions of hepatotoxicity (as indicated by Table 13), including complex composite definitions, raises concerns that outcome definitions in the included studies may have been influenced by the magnitude, statistical significance or perceived importance of observed effect sizes.

3.5 Conclusion

There is a substantial evidence base for the association between genetic variants and anti-TB drug-related toxicity outcomes, as previously identified and as our systematic review confirmed. However, by considering the characteristics of included studies and performing a rigorous quality assessment, we established that performing robust synthesis of the evidence base would be challenging. Studies varied substantially in terms of definition of outcomes, and in terms of methodological quality. Reporting of key details, such as the ethnicity of included patients and aspects of methodological quality, was poor across the included studies.

4 Influence of genetic variants on toxicity related to anti-tuberculosis drugs: meta-analyses

The aim of this chapter was to perform meta-analyses of the studies identified in the systematic review presented in Chapter 3. It was hoped that synthesising the identified studies would improve the power to detect associations between genetic variants and anti-TB drug related toxicity outcomes. Each meta-analysis was performed for a specific combination of genetic variant and anti-TB drug related toxicity outcome. Here, we outline the methods used to perform the meta-analyses, the results generated, and our interpretation of these results.

4.1 Methods

Primary analyses

To inform our decisions regarding which analyses would be the primary analyses, we examined the existing literature and consulted clinical experts. The primary analyses focused on hepatotoxicity, as this is the most widely studied anti-TB drug-related toxicity outcome, and on key genetic variants that researchers have hypothesised may play a role in the biological pathway between anti-TB treatment and hepatotoxicity (*NAT2* acetylator status*; CYP2E1 Rsa*I*, Pst*I and *Dra*I polymorphisms*;* and *GSTM1* and *GSTT1* null polymorphisms; see Section 3.1).

Clinical advice was that a comparison of slow and intermediate *NAT2* acetylator groups combined *versus* the rapid acetylator group would be clinically useful. For the *CYP2E1*, *GSTM1* and *GSTT1* polymorphisms, combinations of particular genotype groups have been consistently used throughout the literature. For the key *CYP2E1* polymorphisms, homozygous mutant-type and heterozygous genotype groups are commonly combined, whereas for the *GSTM1* and *GSTT1* null polymorphisms, heterozygous and homozygous present (or 'non-null') genotype groups are commonly combined. The evidence base therefore suggested that assuming a dominant mode of inheritance for the *CYP2E1* polymorphisms, and a recessive mode of inheritance for the *GSTM1* and *GSTT1* null polymorphisms would be appropriate. Consequently, the primary analyses compared the likelihood of experiencing hepatotoxicity for:

1. *NAT2* slow/intermediate acetylators *versus* rapid acetylators

Results were combined from studies that reported data for each acetylator group separately together with data from studies that combined slow and intermediate acetylator

groups. Data from studies that combined intermediate and rapid acetylator groups were included in sensitivity analyses (see below).

2. Homozygous mutant-type or heterozygous individuals *versus* homozygous wild-type individuals for the *Rsa*I, *Pst*I and *Dra*I polymorphisms of *CYP2E1*

Results were combined from studies that reported data for each genotype group separately with data from studies that combined homozygous mutant-type and heterozygous genotype groups. No studies combined homozygous wild-type and heterozygous genotype groups.

3. Individuals with homozygous null genotype *versus* those with heterozygous or homozygous present genotype for *GSTM1* and *GSTT1*

Results were combined from studies that reported data for each genotype group separately and studies that combined homozygous present and heterozygous genotype groups. No studies combined homozygous null and heterozygous genotype groups.

The following sensitivity analyses were conducted to investigate the robustness of the primary analyses:

- *NAT2* acetylator status: Pairwise comparisons of slow *versus* rapid acetylators, and intermediate *versus* rapid acetylators, using data from studies that reported data for each acetylator group separately
- *NAT2* acetylator status: Comparison of slow *versus* rapid/intermediate acetylators, using data from studies that combined data for intermediate and rapid acetylator groups, and from studies that reported data for each acetylator group separately
- *CYP2E1* polymorphisms: Pairwise comparisons of heterozygous *versus* homozygous wild-type genotype, and homozygous mutant-type *versus* homozygous wild-type genotype for the *CYP2E1 Rsa*I, *Pst*I and *Dra*I polymorphisms, using data from studies that reported data for each genotype group separately
- *GSTM1/GSTT1* null polymorphisms: Pairwise comparisons of heterozygous *versus* homozygous present genotype, and homozygous null *versus* homozygous present genotype for both *GSTM1* and *GSTT1*, using data from studies that reported data for each genotype group separately.

Prior to undertaking meta-analyses, we performed an exact test for HWE within each study. A significance level of $p<0.05$ was used to indicate deviation from HWE. We did not test HWE for *NAT2* acetylator status, as acetylator status is a trait defined by several *NAT2* SNPs. Furthermore, if we did not have the number of individuals in each genotype group (i.e. homozygous wild-type, heterozygous, homozygous mutant-type) for a particular study, it was not possible to test for HWE. Where genotypes for a study were found to deviate from HWE ($p<0.05$), a sensitivity analysis was conducted excluding that study.

We produced funnel plots for each of the primary analyses to investigate the possibility of small study effects (such as publication bias). We would not have produced a funnel plot for a primary analysis if less than 10 studies had been included, as a sizeable number of studies are required to visually assess a funnel plot for small study effects.[191]

Secondary analyses

The secondary analyses investigated all other associations between genetic variants and anti-TB drug-related toxicity outcomes. We performed meta-analyses according to the following strategy:

- For SNPs where all studies presented data for each genotype group separately, we performed two pairwise comparisons; heterozygous *versus* homozygous wild-type genotype, and homozygous mutant-type *versus* homozygous wild-type genotype.
- For SNPs where all studies presented data for the same combined genotype groups, we performed one comparison of the combined genotype groups.
- For SNPs where the approach varied between studies, we performed both pairwise comparisons (using data from studies that reported on each genotype group separately), and a comparison of the combined genotype groups (using data from all studies).

Data synthesis

All meta-analyses were performed using the metan package in Stata 14;[192] ORs with 95% CIs were the chosen measure of effect. A random-effects model was employed because we anticipated both methodological heterogeneity (due to differences in study design and quality of methods) and clinical heterogeneity (due to differences in participants' ethnic backgrounds and outcome definitions). The DerSimonian and Laird random-effects model, [193] which makes use of a 'moment-based' estimate of the heterogeneity variance, was implemented; Cochran's Q was used as the homogeneity statistic. We assessed statistical heterogeneity by visually examining the forest plots, and by referring to the I^2 statistic.

We considered performing meta-analysis for all associations between a genetic variant and a toxicity outcome that were investigated by at least two studies; however, we would not have pooled data if we judged that clinical, methodological or statistical heterogeneity would render the overall effect estimate to be meaningless.

If no events occurred in one of the genotype groups for a particular study, a continuity correction of 0.5 was applied.[25] If there were no patients in one of the genotype groups for a particular study, data from this study were excluded from the meta-analysis.

The HuGENet HuGE Review Handbook recommends that meta-analyses of genetic association studies are stratified by ethnicity, and that pooling of results should only be performed if effect estimates for different ethnic groups appear sufficiently similar.[24] Information on participants' ethnicity was not commonly reported; however, in an attempt to follow this recommendation, we stratified our analyses by the countries in which studies were performed.

For SNPs investigated by one study only, ORs comparing genotype groups and corresponding 95% confidence intervals (CIs) were calculated and reported in tables.

4.2 Results

Primary analyses: *NAT2* acetylator status and hepatotoxicity

A total of 22 studies reported data for each *NAT2* acetylator group separately, or for combined slow and intermediate acetylator groups *versus* rapid acetylators, and were therefore included in the primary analysis for *NAT2* acetylator status and hepatotoxicity. A forest plot displaying the results of this primary analysis is provided in Figure 2.

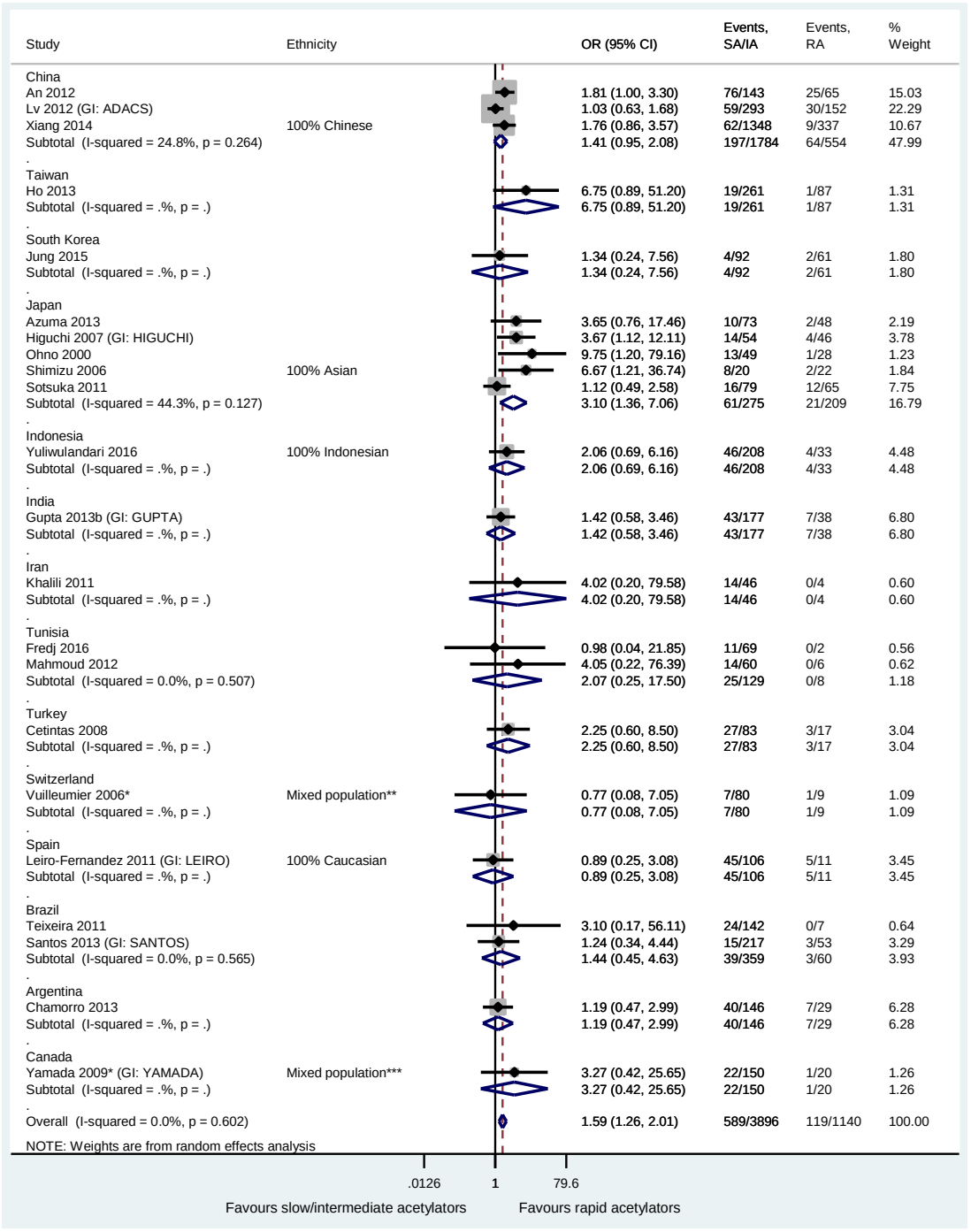

Figure 2 NAT2 *acetylator status and hepatotoxicity; slow/intermediate* versus *rapid acetylator status*

*Vuilleumier 2006[164] and Yamada 2009[171] were both conducted in the latent TB population.
**Caucasian: 38 (43%), Hispanic: 8 (9%), African: 22 (25%), South American: 15 (17%), Asian: 5 (6%), Middle Eastern: 1 (1%)
*** Asian: 72 (42%), Caucasian: 49 (29%), South Asian: 22 (13%), Hispanic: 7 (4%), Middle Eastern: 8 (5%), First nations: 5 (3%), other/mixed/unknown: 7 (4%)
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; IA: intermediate acetylators; OR: odds ratio; RA: rapid acetylators; SA: slow acetylators; TB: tuberculosis

Slow/intermediate acetylators were significantly more likely to experience hepatotoxicity than rapid acetylators (OR=1.59, 95% CI: 1.26 to 2.01). No heterogeneity was detected in this analysis (I^2=0.0%).

Interestingly, the result from a sensitivity analysis comparing the likelihood of hepatotoxicity for slow *versus* rapid/intermediate acetylators (Figure 3) resulted in a stronger estimate of association (OR=3.12, 95% CI: 2.45 to 3.97, I^2=59.0%) than the primary analysis (of slow/intermediate *versus* rapid acetylator status). A total of 34 studies reported data for each *NAT2* acetylator group separately, or for combined rapid and intermediate acetylator groups *versus* slow acetylators, and were therefore included in this sensitivity analysis.

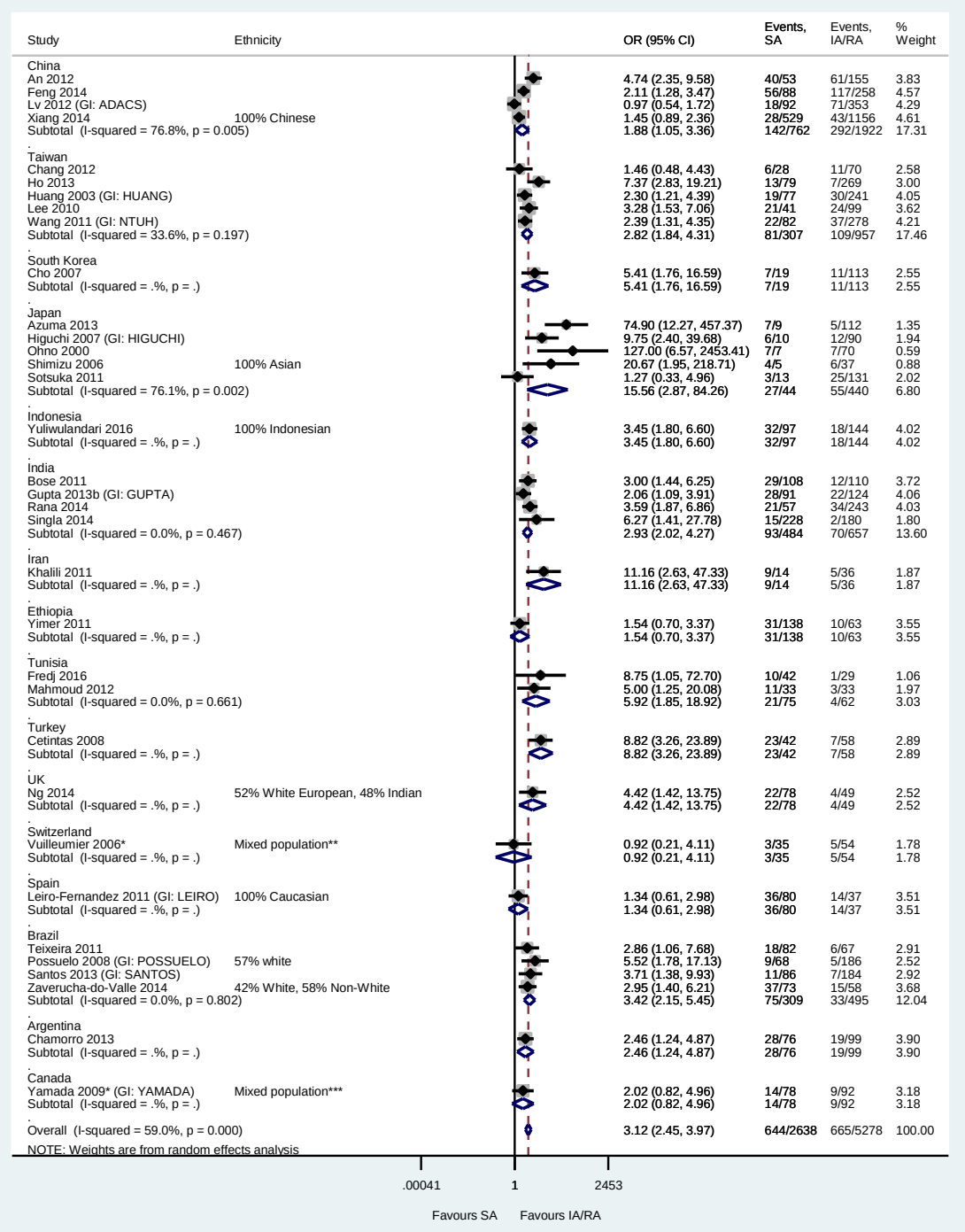

Figure 3 NAT2 *acetylator status and hepatotoxicity sensitivity analysis: slow* versus *rapid/intermediate acetylator status*

* Vuilleumier 2006[164] and Yamada 2009[171] were both conducted in the latent TB population.
**Caucasian: 38 (43%), Hispanic: 8 (9%), African: 22 (25%), South American: 15 (17%), Asian: 5 (6%), Middle Eastern: 1 (1%)
*** Asian: 72 (42%), Caucasian: 49 (29%), South Asian: 22 (13%), Hispanic: 7 (4%), Middle Eastern: 8 (5%), First nations: 5 (3%), other/mixed/unknown: 7 (4%)
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; IA: intermediate acetylators; OR: odds ratio; RA: rapid acetylators; SA: slow acetylators; TB: tuberculosis

Results of the pairwise comparison sensitivity analyses for *NAT2* acetylator status and hepatotoxicity are provided in Appendix 3. Results from the pairwise comparisons suggested that slow acetylators were significantly more likely to experience hepatotoxicity than rapid acetylators (OR=3.68, 95% CI: 2.23 to 6.09, I^2=60.0%, 21 studies), but that there were no significant differences between intermediate and rapid acetylators (OR=1.12, 95% CI: 0.87 to 1.45, I^2=0.0%, 21 studies). A comparison of the results obtained by applying different approaches to the analysis of *NAT2* acetylator status and hepatotoxicity, i.e. assuming a dominant mode of inheritance (slow/intermediate *versus* rapid acetylators), assuming a recessive mode of inheritance (slow *versus* intermediate/rapid acetylators), and performing pairwise comparisons, is provided in Figure 4.

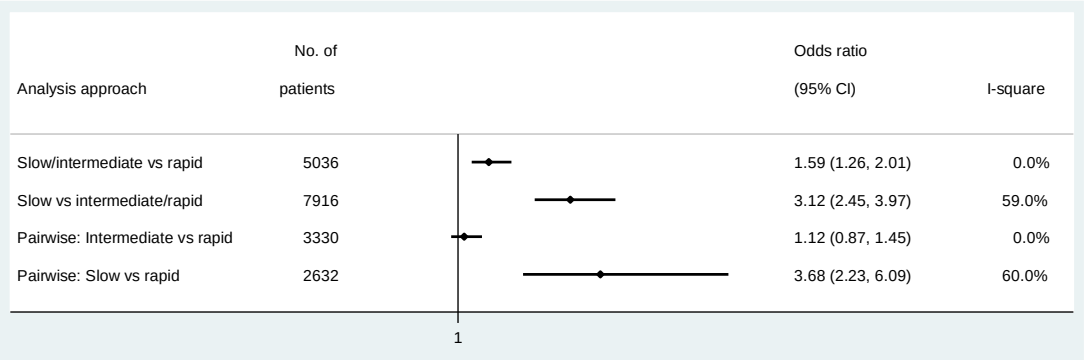

Figure 4 Summary of the results obtained by applying different approaches to the analysis of NAT2 *acetylator status and hepatotoxicity*

CI: confidence interval

The funnel plot for the analysis of slow/intermediate acetylators *versus* rapid acetylators for the outcome of hepatotoxicity (Figure 5) provided no evidence of small study effects.

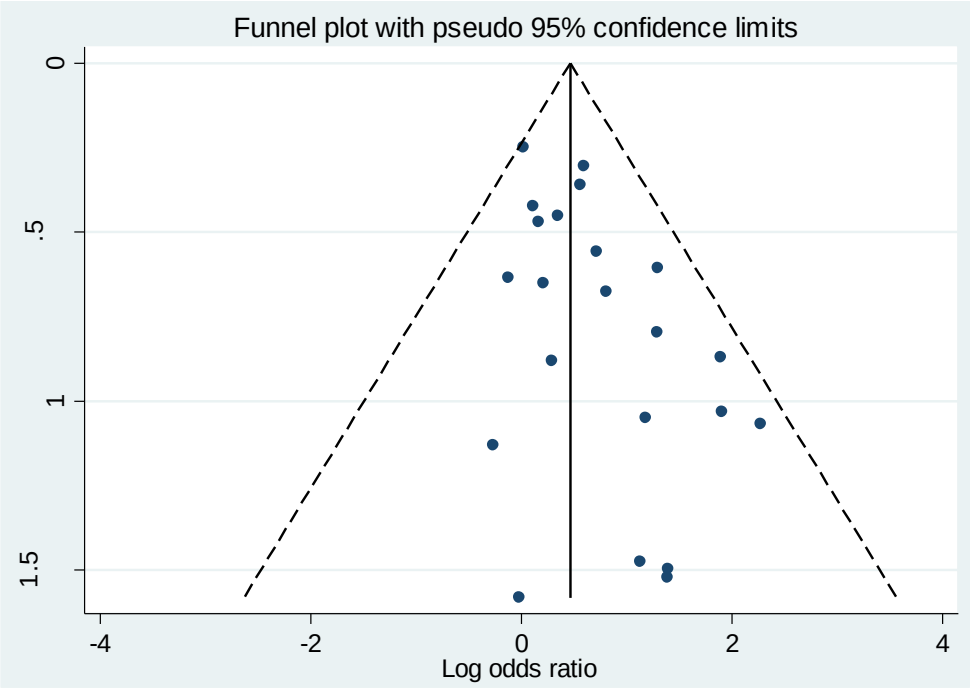

Figure 5 Funnel plot for the analysis of NAT2 *slow/intermediate* versus *rapid acetylator status and hepatotoxicity*

Primary analyses: Key *CYP2E1* SNPs and hepatotoxicity

Forest plots displaying the results of the primary analyses for the *CYP2E1* key SNPs are provided in Figure 6, Figure 8 and Figure 10.

A total of 23 studies reported data for each *CYP2E1 Rsa*I genotype group separately, or for combined homozygous mutant-type (TT) and heterozygous (CT) groups *versus* the homozygous wild-type group (CC), and were therefore included in the primary analysis for the *CYP2E1 Rsa*I polymorphism and hepatotoxicity. Patients with TT or CT genotype at the *CYP2E1 Rsa*I polymorphism were significantly less likely to experience hepatotoxicity than patients with CC genotype (OR=0.75, 95% CI: 0.56 to 1.00) (Figure 6). Moderate heterogeneity was observed in this analysis (I^2=58.2%).

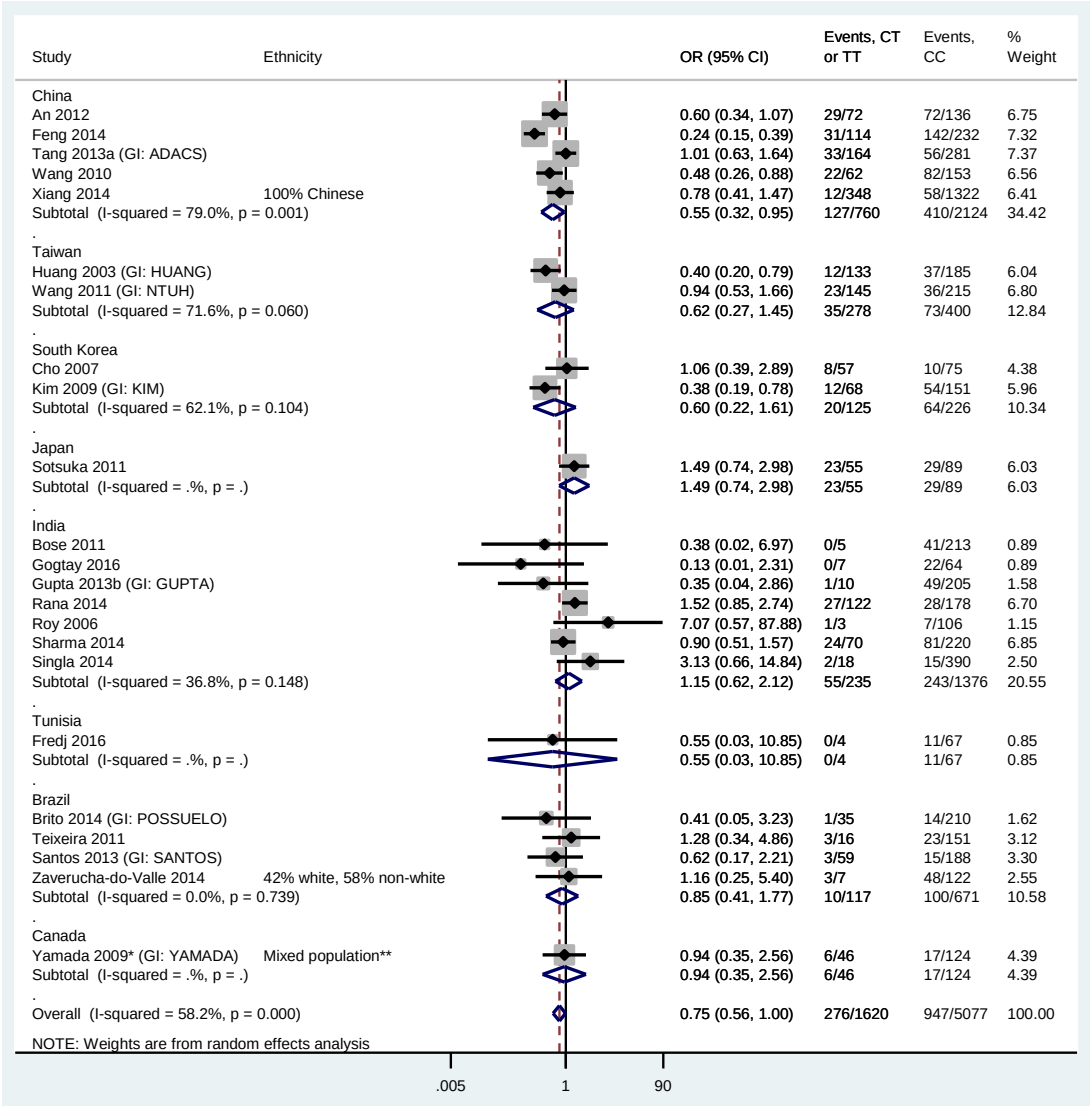

Figure 6 CYP2E1 Rsa*I polymorphism and hepatotoxicity: homozygous mutant-type (TT) or heterozygous (CT)* versus *homozygous wild-type (CC)*

* Yamada 2009[171] was conducted in the latent TB population.
**Asian: 72 (42%), Caucasian: 49 (29%), South Asian: 22 (13%), Hispanic: 7 (4%), Middle Eastern: 8 (5%), First Nations: 5 (3%), other/mixed/unknown: 7 (4%)
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio; TB: tuberculosis

The results of the sensitivity analyses for key *CYP2E1* SNPs and hepatotoxicity are provided in Appendix 3. For the *CYP2E1 Rsa*I polymorphism, when three studies[152, 158, 159] were excluded due to deviation from HWE, the association remained statistically significant, and heterogeneity was slightly reduced (OR=0.64, 95% CI: 0.48 to 0.84, I^2=47.2%, 20 studies). No significant differences were observed for either pairwise comparison (CT *versus* CC: OR=0.80, 95% CI: 0.58 to 1.10, I^2=48.4%, 18 studies; TT *versus* CC: OR=1.03, 95% CI: 0.68 to 1.55, I^2=2.7%, 18 studies). A comparison of the results obtained by applying different approaches to the analysis of *CYP2E1 Rsa*I polymorphism and hepatotoxicity, i.e. assuming a dominant mode of inheritance (TT/CT *versus* CC), excluding studies due to deviation from

HWE from this dominant analysis approach, and performing pairwise comparisons, is provided in Figure 7.

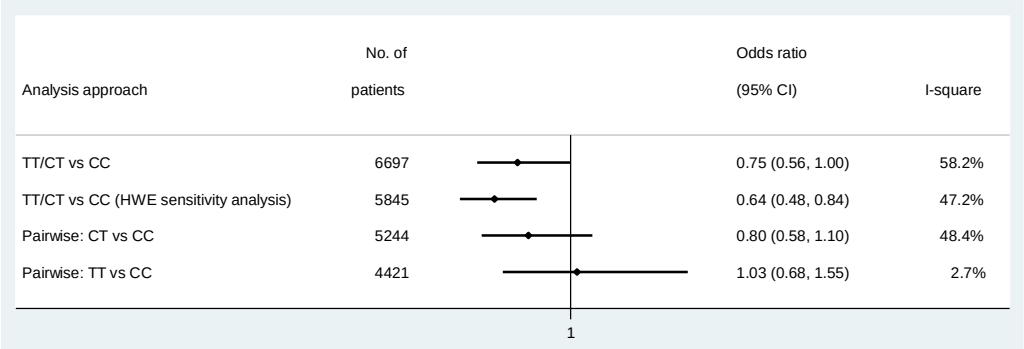

Figure 7 Summary of the results obtained by applying different approaches to the analysis of CYP2E1 Rsa*I polymorphism and hepatotoxicity*

CI: confidence interval; HWE: Hardy-Weinberg equilibrium

Eight studies reported data for each *CYP2E1 Dra*I genotype group separately, or for combined homozygous mutant-type (AA) and heterozygous (TA) groups *versus* the homozygous wild-type (TT) group, and were therefore included in the primary analysis for the *CYP2E1 Dra*I polymorphism and hepatotoxicity. There was no significant difference in the likelihood of hepatotoxicity between patients with AA or TA genotype at the *CYP2E1 Dra*I polymorphism and patients with TT genotype (OR=1.23, 95% CI: 0.92 to 1.66) (Figure 8). Minimal heterogeneity was observed in this analysis (I^2=3.1%).

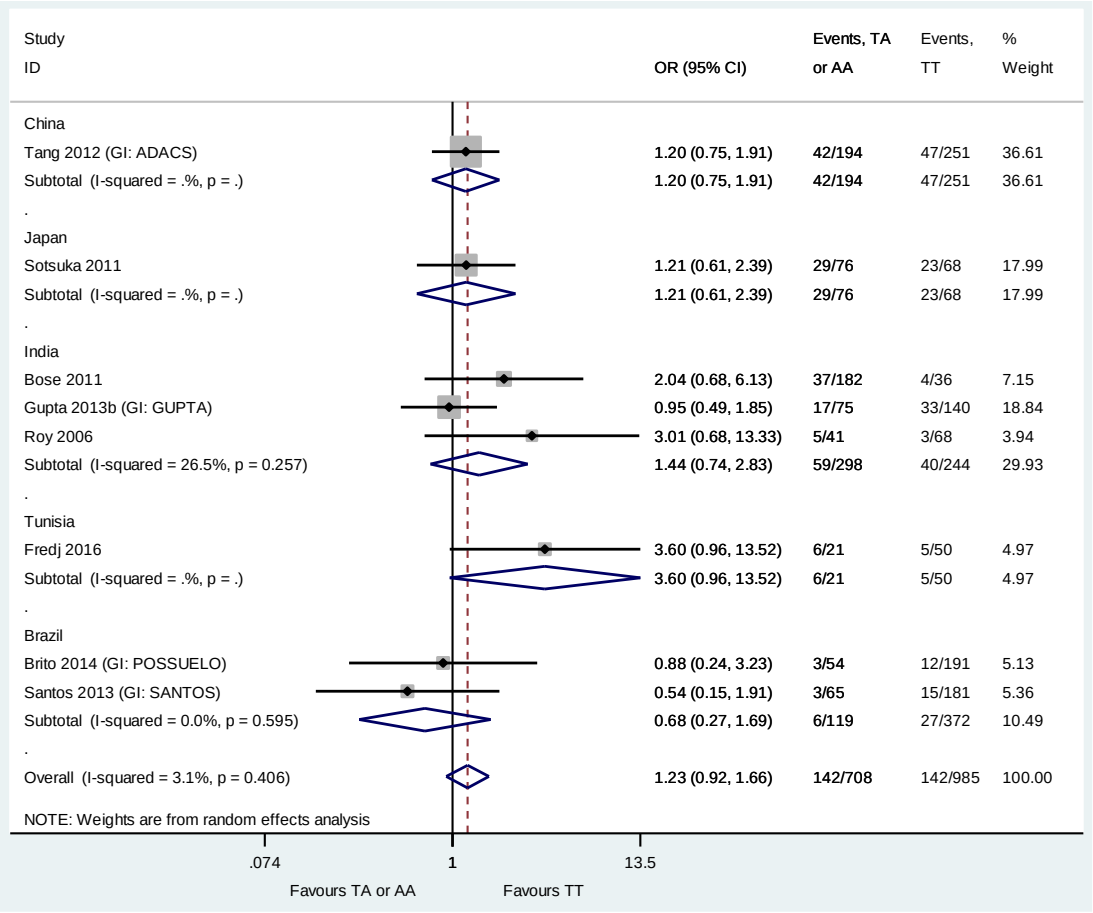

Figure 8 CYP2E1 DraI polymorphism and hepatotoxicity: homozygous mutant-type (AA) or heterozygous (AT) versus homozygous wild-type (TT)

None of the included studies reported ethnicity so this information is not provided on the forest plot.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

The sensitivity analysis excluding two studies[112, 126] where genotypes were found to deviate from HWE also showed no significant association (OR=1.27, 95% CI: 0.86 to 1.87, I^2=14.2%, 6 studies). Furthermore, the two pairwise comparisons showed no significant differences between genotype groups (TA *versus* TT: OR=1.28, 95% CI: 0.93 to 1.77, I^2=6.4%, 7 studies; AA *versus* TT: OR=1.34, 95% CI: 0.57 to 3.16, I^2=29.4%, 7 studies) (Appendix 3). A comparison of the results obtained by applying different approaches to the analysis of *CYP2E1 Dra*I polymorphism and hepatotoxicity, i.e. assuming a dominant mode of inheritance (AA/TA *versus* TT), excluding studies due to deviation from HWE from this dominant analysis approach, and performing pairwise comparisons, is provided in Figure 9.

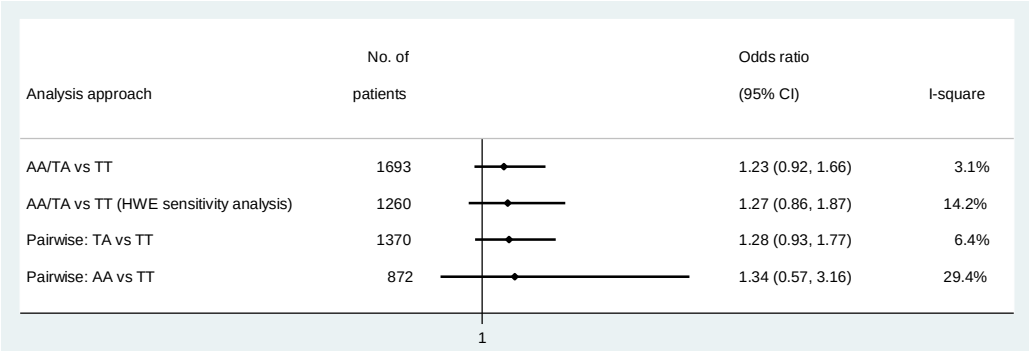

Figure 9 Summary of the results obtained by applying different approaches to the analysis of CYP2E1 Dra*I polymorphism and hepatotoxicity*

CI: confidence interval; HWE: Hardy-Weinberg equilibrium

Eight studies reported data for each *CYP2E1 Pst*I genotype group separately, or for combined homozygous mutant-type (CC) and heterozygous (GC) groups *versus* the homozygous wild-type (GG) group, and were therefore included in the primary analysis for the *CYP2E1 Pst*I polymorphism and hepatotoxicity. There was no significant difference in the likelihood of experiencing hepatotoxicity between patients with CC or GC genotype at the *CYP2E1 Pst*I polymorphism and patients with GG genotype (OR=0.78, 95% CI: 0.46 to 1.34) (Figure 10). Moderate heterogeneity was observed in this analysis (I^2=50.3%).

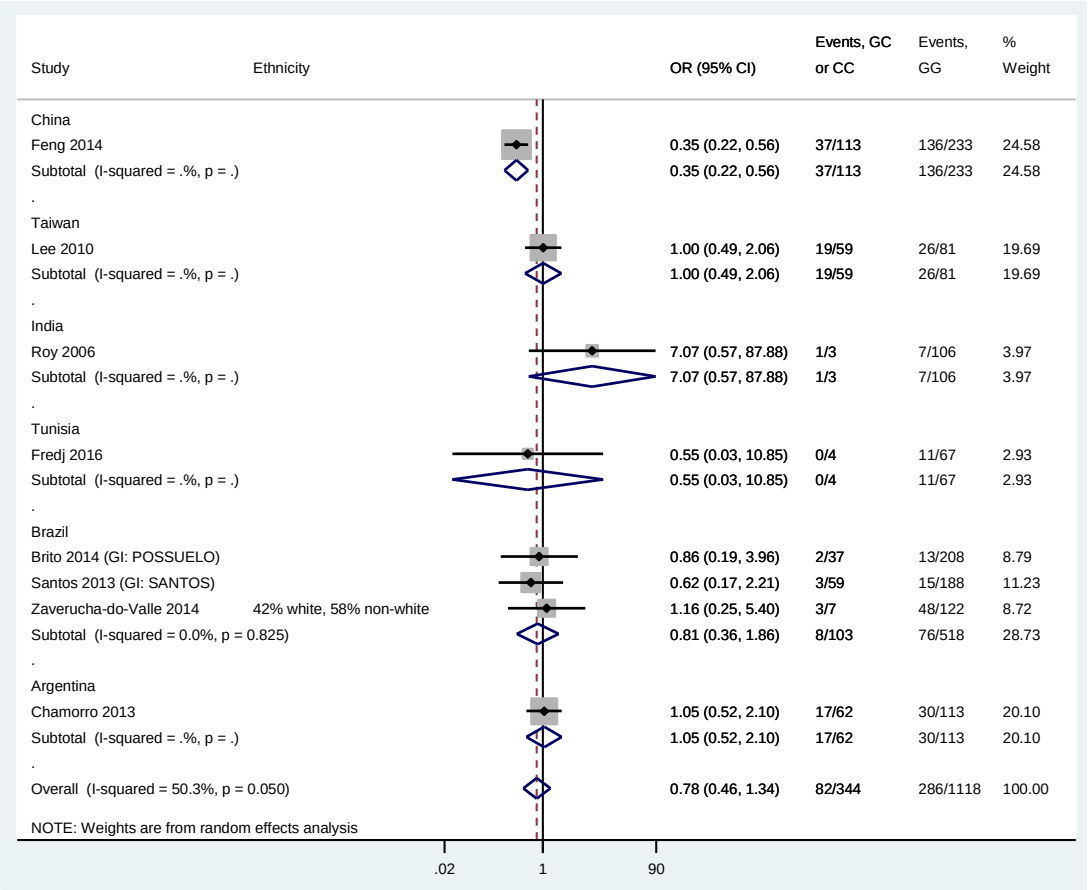

Figure 10 CYP2E1 Pst*I polymorphism and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (GC)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

The sensitivity analysis excluding one study[114] where genotypes deviated from HWE (Appendix 3) also showed no significant association (OR=0.74, 95% CI: 0.40 to 1.37, I^2=48.1%, 7 studies). The pairwise comparisons also showed no significant differences between genotype groups (GC *versus* GG: OR=1.05, 95% CI: 0.66 to 1.65, I^2=0.0%, 6 studies; CC *versus* GG: OR=1.04, 95% CI: 0.36 to 2.99, I^2=0.0%, 6 studies) (Appendix 3). A comparison of the results obtained by applying different approaches to the analysis of *CYP2E1 Pst*I polymorphism and hepatotoxicity, i.e. assuming a dominant mode of inheritance (CC/GC *versus* GG), excluding studies due to deviation from HWE from this dominant analysis approach, and performing pairwise comparisons, is provided in Figure 11.

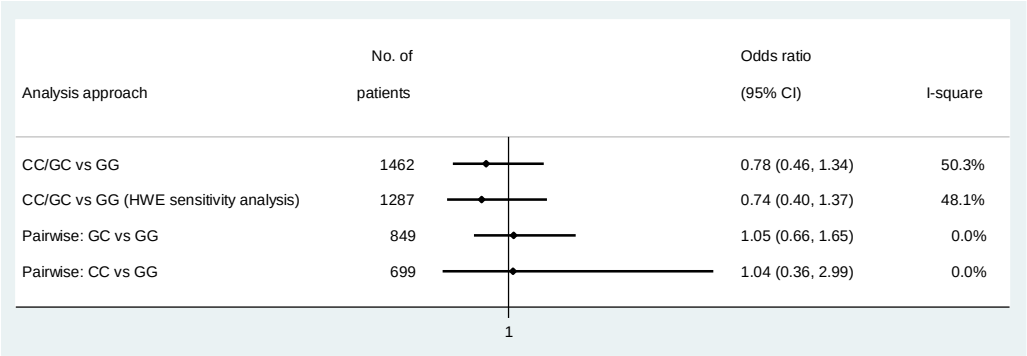

Figure 11 Summary of the results obtained by applying different approaches to the analysis of CYP2E1 PstI *polymorphism and hepatotoxicity*

CI: confidence interval; HWE: Hardy-Weinberg equilibrium

We produced a funnel plot for each of the primary analyses for the *CYP2E1* gene (Figure 12 to Figure 14). There was no evidence to suggest that small study effects were an issue of concern.

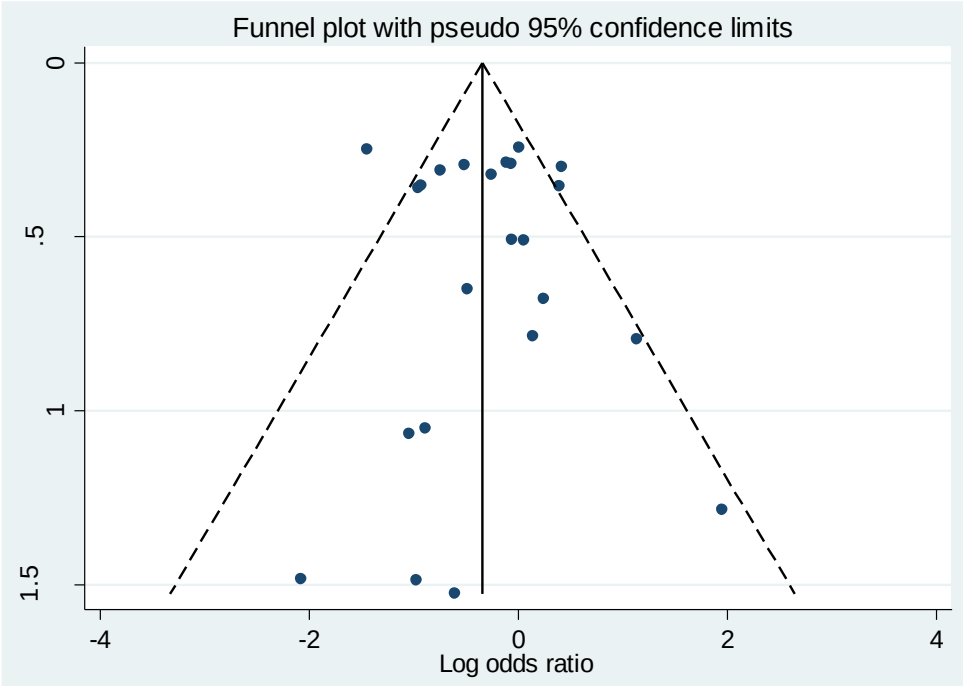

Figure 12 Funnel plot for the analysis of CYP2E1 RsaI *polymorphism and hepatotoxicity: homozygous mutant-type (TT) or heterozygous (CT)* versus *homozygous wild-type (CC)*

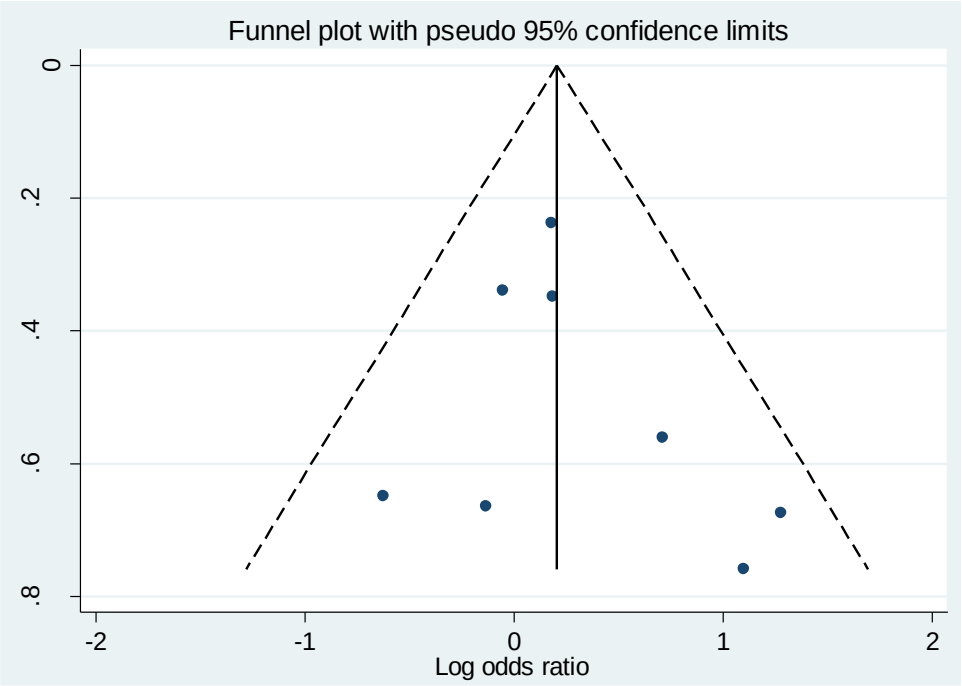

Figure 13 Funnel plot for the analysis of CYP2E1 Dra*I polymorphism and hepatotoxicity: homozygous mutant-type (AA) or heterozygous (AT)* versus *homozygous wild-type (TT)*

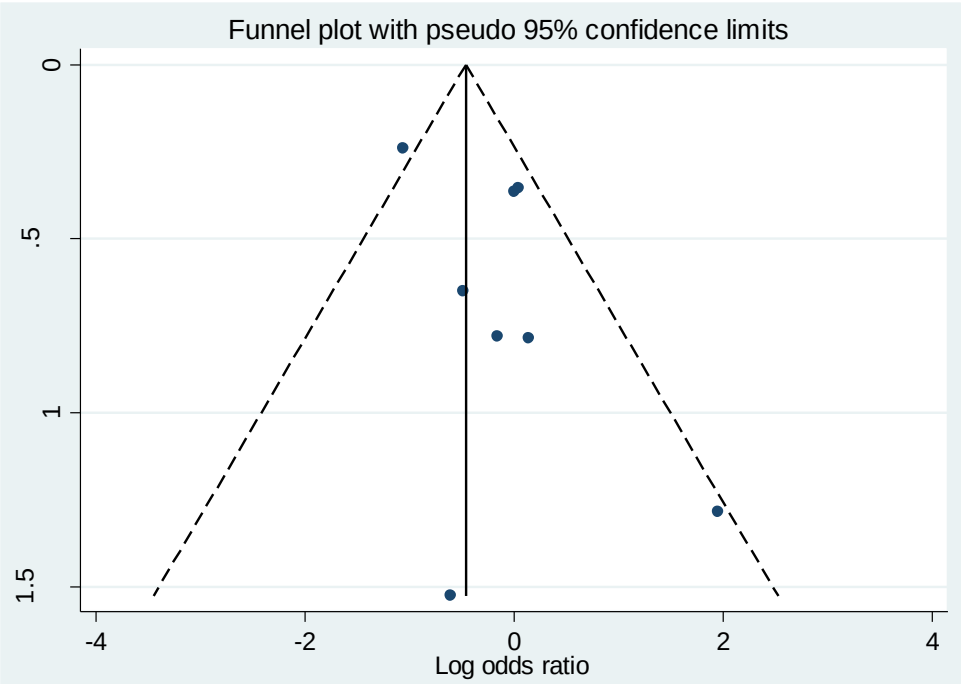

Figure 14 Funnel plot for the analysis of CYP2E1 Pst*I polymorphism and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (GC)* versus *homozygous wild-type (GG)*

Primary analyses: *GSTM1*/*GSTT1* and hepatotoxicity

Forest plots displaying the results of the primary analyses for *GSTM1* and *GSTT1* are provided in Figure 15 and Figure 17, respectively.

A total of 18 studies reported data for each *GSTM1* genotype group separately, or for combined homozygous present and heterozygous groups *versus* the homozygous null group, and were therefore included in the primary analysis for the *GSTM1* null polymorphism and hepatotoxicity. For *GSTM1*, patients with homozygous null genotype were significantly more likely to experience hepatotoxicity than patients with heterozygous or homozygous present genotype (OR=1.44, 95% CI: 1.15 to 1.82) (Figure 15). Moderate heterogeneity was observed in this analysis (I^2=51.2%).

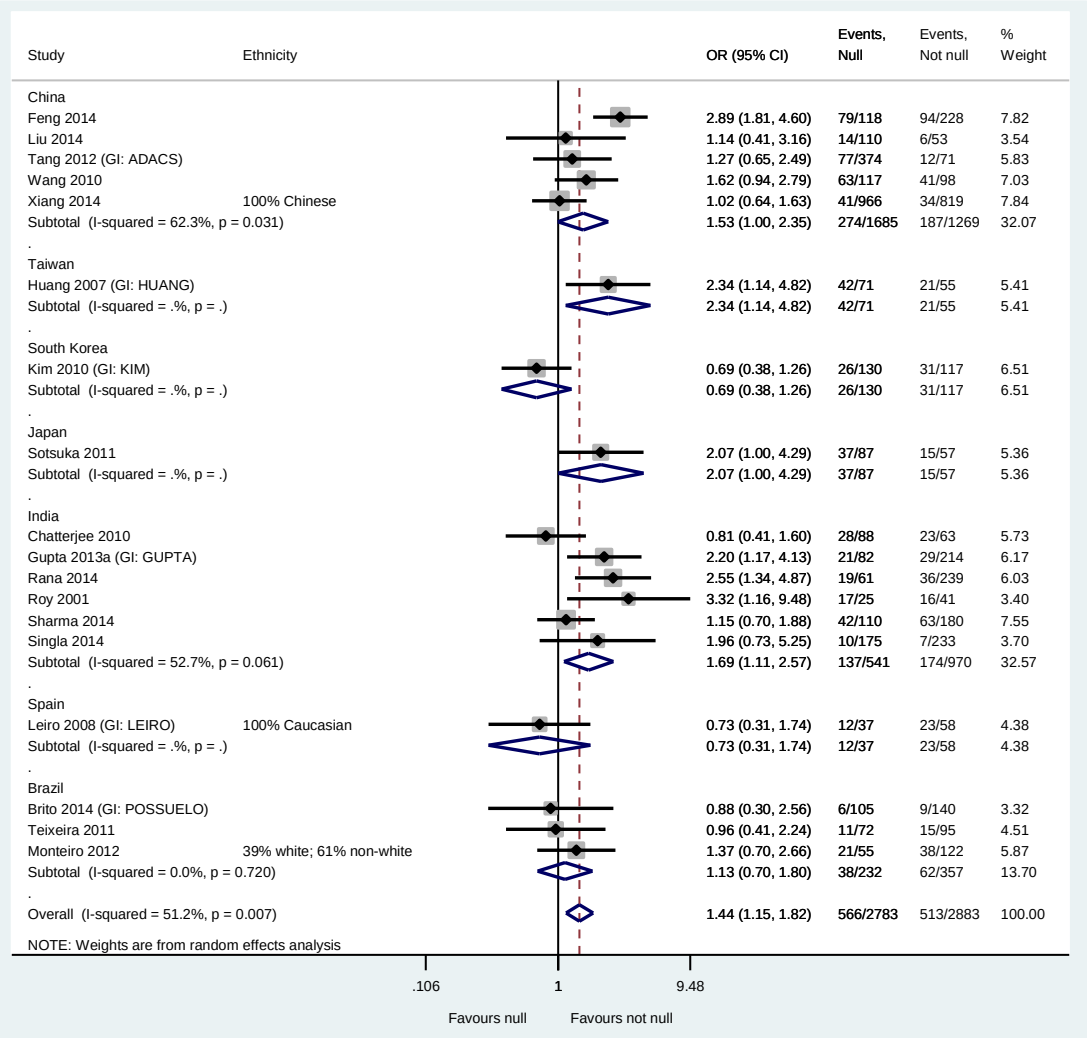

Figure 15 GSTM1 *null polymorphism and hepatotoxicity: homozygous null* versus *heterozygous or homozygous present*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

The results of the sensitivity analyses for the *GSTM1* null polymorphism and gene are provided in Appendix 3. It was only possible to test deviation from HWE for one study,[144] which provided data for each genotype group separately. Genotypes were found to deviate from HWE for this study, and so we conducted a sensitivity analysis excluding this study,

which also showed a statistically significant association (OR=1.46, 95% CI: 1.15 to 1.85, I^2=53.7%, 17 studies). No significant differences were observed for either pairwise comparison (heterozygous *versus* homozygous present: OR=0.42, 95% CI: 0.02 to 8.18, 1 study; homozygous null *versus* homozygous present: OR=0.97, 95% CI: 0.35 to 2.71, 1 study). Data for these pairwise comparisons came from a single study,[144] and so we did not calculate the I^2 statistic.

A comparison of the results obtained by applying different approaches to the analysis of *GSTM1* null polymorphism and hepatotoxicity, i.e. assuming a recessive mode of inheritance (homozygous null *versus* heterozygous or homozygous present), excluding studies due to deviation from HWE from this recessive analysis approach, and performing pairwise comparisons, is provided in Figure 16.

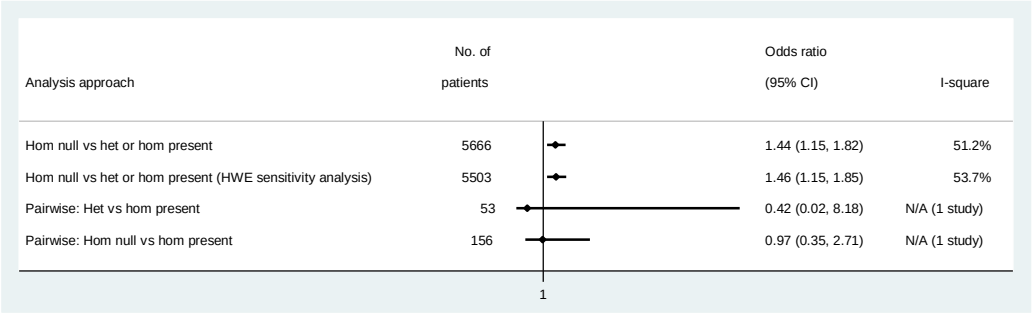

Figure 16 Summary of the results obtained by applying different approaches to the analysis of GSTM1 *null polymorphism and hepatotoxicity*

CI: confidence interval; het: heterozygous; hom: homozygous; HWE: Hardy-Weinberg equilibrium

A total of 16 studies reported data for each *GSTT1* genotype group separately, or for combined homozygous present and heterozygous groups *versus* the homozygous null group, and were therefore included in the primary analysis for the *GSTT1* null polymorphism and hepatotoxicity. For *GSTT1*, there was no significant difference in the likelihood of experiencing hepatotoxicity between patients with homozygous null genotype and patients with heterozygous or homozygous present genotype (OR=1.06, 95% CI: 0.85 to 1.32) (Figure 17). A relatively small amount of heterogeneity was observed in this analysis (I^2=27.1%).

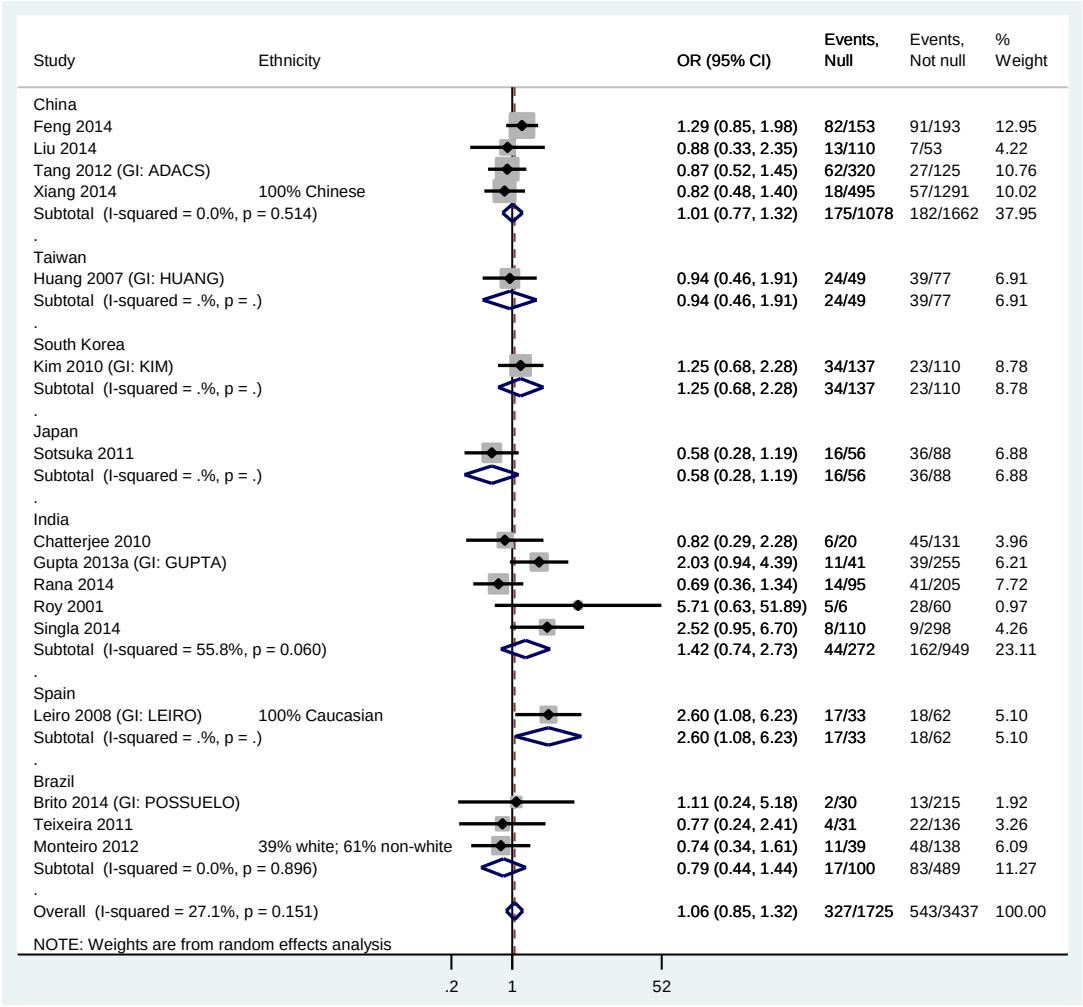

Figure 17 GSTT1 *null polymorphism and hepatotoxicity: homozygous null* versus *heterozygous or homozygous present*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

It was only possible to test deviation from HWE for two studies, Liu 2014[144] and Teixera 2011,[163] which both reported data for each genotype group separately. We did not identify any deviation from HWE, so we did not perform a sensitivity analysis to investigate the impact of deviation from HWE. The sensitivity analyses (two pairwise comparisons) also showed no significant differences between genotype groups (heterozygous *versus* homozygous present: OR=0.67, 95% CI: 0.29 to 1.55; I^2=0.0%, 2 studies; homozygous null *versus* homozygous present: OR=0.61, 95% CI: 0.22 to 1.66; I^2=0.0, 2 studies) (Appendix 3).

A comparison of the results obtained by applying different approaches to the analysis of *GSTT1* null polymorphism and hepatotoxicity, i.e. assuming a recessive mode of inheritance (homozygous null *versus* heterozygous or homozygous present) and performing pairwise comparisons, is provided in Figure 18.

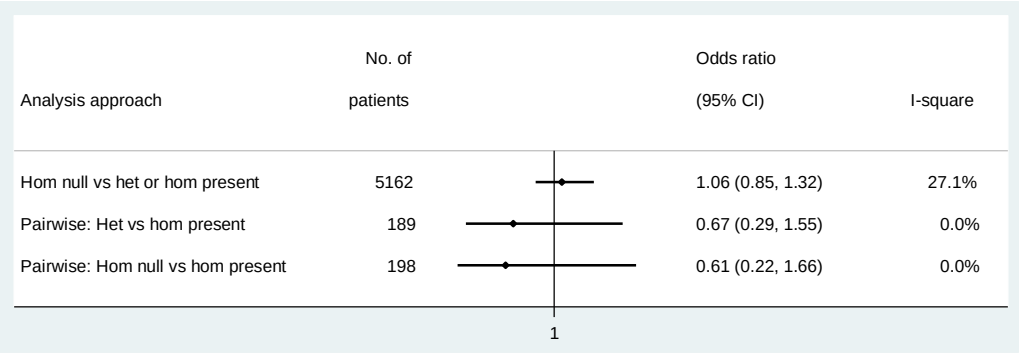

Figure 18 Summary of the results obtained by applying different approaches to the analysis of GSTT1 *null polymorphism and hepatotoxicity*

CI: confidence interval; het: heterozygous; hom: homozygous

We produced funnel plots for the primary analyses of the *GSTM1* and *GSTT1* genes (Figure 19 and Figure 20). There was no evidence to suggest that small study effects were an issue of concern.

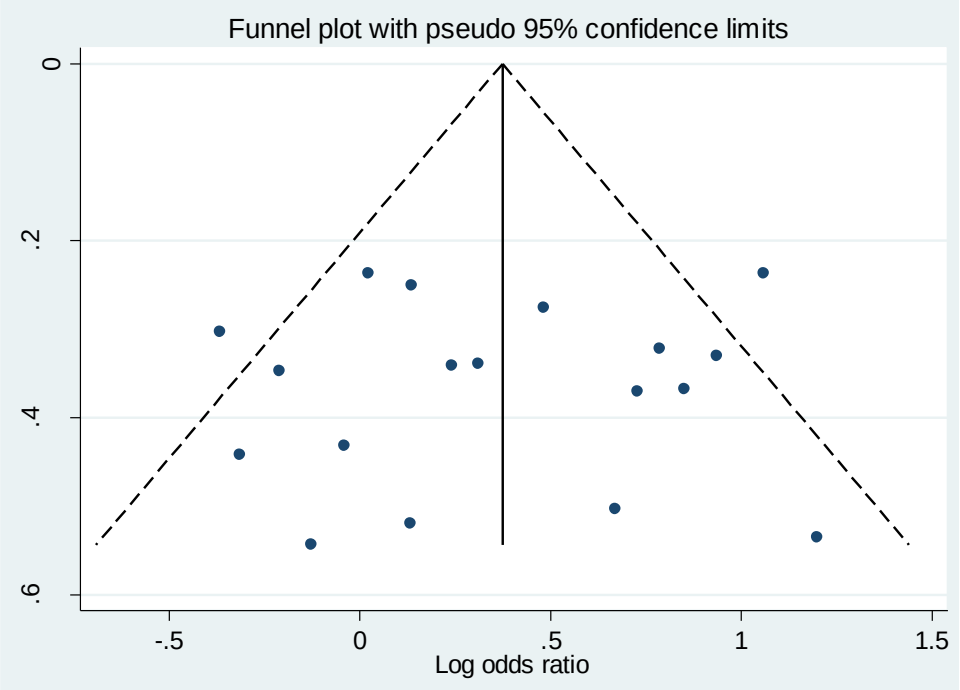

Figure 19 Funnel plot for the analysis of GSTM1 *null polymorphism and hepatotoxicity: homozygous null* versus *heterozygous or homozygous present*

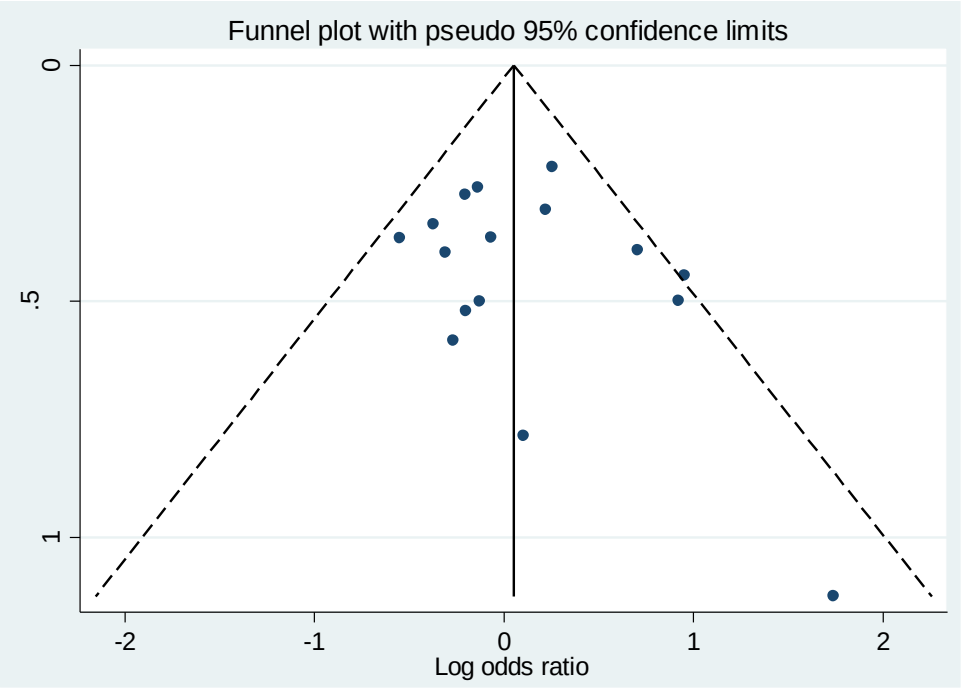

Figure 20 Funnel plot for the analysis of GSTT1 *null polymorphism and hepatotoxicity: homozygous null* versus *heterozygous or homozygous present*

Heterogeneity in the primary analyses

Moderate heterogeneity was observed in the primary analyses for the *CYP2E1 Rsa*I and *Pst*I polymorphisms, and in the primary analysis for the *GSTM1* null polymorphism. Furthermore, although no heterogeneity was observed in the primary analysis for *NAT2* acetylator status, heterogeneity was observed in the sensitivity analyses of slow *versus* rapid acetylator status, and slow *versus* rapid/intermediate acetylator status. Heterogeneity in these analyses may have been caused by the variable distribution of genotypes in different ethnic populations. Previous studies have suggested that there is a considerable amount of genetic diversity across different ethnic populations for *NAT2* acetylator status,[194] for the *CYP2E1 Rsa*I and *Pst*I polymorphisms[195] and for the *GSTM1* null polymorphism.[196] We stratified all analyses by country (as a proxy variable for ethnicity); however, it is is difficult to draw conclusions about the effect of country on the investigated genetic associations, due to the small numbers of studies conducted in each country.

Heterogeneity may also have been introduced by differences in the definitions of hepatotoxicity across included studies. Performing subgroup analyses according to definition of hepatotoxicity was not possible, considering the number of different definitions of hepatotoxicity (see Section 3.3), and a lack of any obvious way to categorise these definitions.

Finally, we followed advice given in the HuGENet HuGE Review Handbook[24] and explored the possibility that the 'Proteus phenomenon' might have occurred for any of the primary analyses (see Section 2.3). It is interesting that in the primary analyses of *NAT2* acetylator status and the *GSTM1* null polymorphism, the most extreme result was observed in the earliest studies (*NAT2* acetylator status, Ohno 2000;[150] *GSTM1* null polymorphism, Roy 2001[153]). However, in both analyses, the majority of subsequent studies suggested effect estimates in the same direction, so it does not seem likely that the associations are spurious.

There is no evidence to suggest that the Proteus phenomenon occurred for the primary analyses of the *CYP2E1 Rsa*I or *Dra*I polymorphisms; however for the analysis of the *CYP2E1 Pst*I polymorphism, the most extreme result is observed in the earliest study (Roy 2006[154]), and results from subsequent studies are very different. Similarly, in the analysis of the *GSTT1* null polymorphism, the most extreme result is observed in the earliest study (Roy 2001[153]), and all but one of the subsequent studies do not detect an association between homozygous null genotype and hepatotoxicity. Therefore, it is possible that these early findings were spurious, and consequently introduce heterogeneity to the analyses of the *CYP2E1 Pst*I polymorphism and the *GSTT1* null polymorphism.

Secondary analyses: *NAT2* SNPs and hepatotoxicity

The included studies reported data for the association between 12 *NAT2* SNPs and hepatotoxicity. A summary of this data is provided in Table 16. There were sufficient data to perform meta-analyses for six SNPs. Forest plots showing the results of these meta-analyses are provided in Appendix 4.

Table 16 Results of the secondary analyses: NAT2 *SNPs and hepatotoxicity*

NAT2 SNP	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
190C-T	Het (CT) *vs* Hom WT (CC)	China (1 study)[111]	NR	0.21 (0.01 to 4.38)	101	107	N/A
	Hom MT (TT) *vs* Hom WT (CC)	China (1 study)[111]	NR	Data excluded[c]			
191G-A (rs1801279)	Het (GA) *vs* Hom WT (GG)	Taiwan (1 study)[168]	NR	Data excluded[c]			
		Turkey (1 study)[113]	NR	Data excluded[c]			
	Hom MT (AA) *vs* Hom WT (GG)	Taiwan (1 study)[168]	NR	Data excluded[c]			
		Turkey (1 study)[113]	NR	Data excluded[c]			
282C-T (rs1041983)	Het (CT) *vs* Hom WT (CC)	China (1 study)[111]	NR	1.28 (0.67 to 2.44)	65	98	N/A
		Taiwan (1 study)[168]	NR	0.50 (0.06 to 4.06)	70	284	N/A
		Indonesia (1 study)[174]	100% Indonesian	1.25 (0.51 to 3.05)	27	148	N/A
		Brazil (1 study)[155]	NR	1.67 (0.56 to 5.00)	14	216	N/A
		All (4 studies)[111, 155, 168, 174]		**1.27 (0.80 to 2.02)**	**176**	**746**	**0.0%**
	Hom MT (TT) *vs* Hom WT (CC)	China (1 study)[111]	NR	7.00 (2.89 to 16.98)	60	51	N/A
		Taiwan (1 study)[168]	NR	1.33 (0.05 to 32.91)	69	277	N/A
		Indonesia (1 study)[174]	100% Indonesian	3.41 (1.38 to 8.40)	31	94	N/A
		Brazil (1 study)[155]	NR	2.07 (0.59 to 7.25)	12	185	N/A
		All (4 studies)[111, 155, 168, 174]		**3.95 (2.21 to 7.05)**	**172**	**607**	**5.5%**
341T-C (rs1801280)	Het (TC) *vs* Hom WT (TT)	China (1 study)[111]	NR	1.63 (0.45 to 5.94)	101	107	N/A
		Taiwan (2 studies)[141, 168]	NR	1.26 (0.58 to 2.75)	114	376	0.0%
		Indonesia (1 study)[174]	100% Indonesian	1.13 (0.54 to 2.35)	49	188	N/A
		Brazil (1 study)[155]	NR	0.66 (0.18 to 2.42)	10	187	N/A
		All (5 studies)[111, 141, 155, 168, 174]		**1.15 (0.72 to 1.82)**	**274**	**858**	**0.0%**
	Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[111]	NR	Data excluded[c]			
		Taiwan (2 studies)[141, 168]	NR	1.18 (0.08 to 16.93)	105	355	41.4%
		Indonesia (1 study)[174]	100% Indonesian	1.32 (0.13 to 13.01)	38	149	N/A
		Brazil (1 study)[155]	NR	1.75 (0.50 to 6.13)	12	122	N/A
		All (4 studies)[141, 155, 168, 174]		**1.54 (0.58 to 4.04)**	**155**	**626**	**0.0%**

NAT2 SNP	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
481C-T (rs1799929)	Het (CT) *vs* Hom WT (CC)	China (3 studies)[111, 145, 170]	1 study - 100% Chinese, 2 studies - NR	1.66 (1.11 to 2.48)	259	2027	0.0%
		Taiwan (1 study)[168]	NR	4.12 (0.25 to 66.63)	70	285	N/A
		Indonesia (1 study)[174]	100% Indonesian	1.01 (0.47 to 2.14)	49	188	N/A
		India (1 study)[126]	NR	1.82 (0.89 to 3.71)	39	154	N/A
		Tunisia (1 study)[146]	NR	1.33 (0.29 to 6.06)	8	42	N/A
		Turkey (1 study)[113]	NR	2.17 (0.88 to 5.36)	28	63	N/A
		Brazil (1 study)[155]	NR	0.44 (0.14 to 1.37)	14	216	N/A
		All (9 studies)[111, 113, 126, 145, 146, 155, 168, 170, 174]		**1.48 (1.12 to 1.97)**	**467**	**2975**	**0.0%**
	Hom MT (TT) *vs* Hom WT (CC)	China (3 studies[d])[111, 145, 170]	1 study - 100% Chinese, 2 studies - NR	0.81 (0.19 to 3.41)	41	1155	N/A
		Taiwan (1 study)[168]	NR	Data excluded[c]			
		Indonesia (1 study)[174]	100% Indonesian	1.28 (0.13 to 12.66)	39	149	N/A
		India (1 study)[126]	NR	5.38 (1.99 to 14.49)	27	97	N/A
		Tunisia (1 study)[146]	NR	3.60 (0.83 to 15.57)	10	34	N/A
		Turkey (1 study)[113]	NR	0.93 (0.17 to 5.08)	14	46	N/A
		Brazil (1 study)[155]	NR	1.19 (0.34 to 4.09)	13	132	N/A
		All (6 studies)[113, 126, 146, 155, 170, 174]		**1.91 (0.93 to 3.92)**	**144**	**1613**	**34.1%**
499G-A	Het (GA) *vs* Hom WT (GG)	China (1 study)[111]	NR	0.21 (0.01 to 4.38)	101	107	N/A
	Hom MT (AA) *vs* Hom WT (GG)	China (1 study)[111]	NR	Data excluded[c]			

NAT2 SNP	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
590G-A (rs1799930)	Het (GA) *vs* Hom WT (GG)	China (3 studies)[111, 145, 170]	1 study - 100% Chinese, 2 studies - NR	1.19 (0.86 to 1.66)	236	1921	15.6%
		Taiwan (2 studies)[141, 168]	NR	1.16 (0.74 to 1.82)	104	356	0.0%
		South Korea (1 study)[133]	NR	1.99 (1.06 to 3.74)	57	145	N/A
		Indonesia (1 study)[174]	100% Indonesian	1.17 (0.58 to 2.36)	38	173	N/A
		India (1 study)[126]	NR	1.38 (0.70 to 2.72)	45	137	N/A
		Tunisia (1 study)[146]	NR	0.77 (0.22 to 2.77)	12	50	N/A
		Turkey (1 study)[113]	NR	2.63 (1.00 to 6.87)	24	67	N/A
		Brazil (1 study)[155]	NR	2.36 (0.27 to 20.76)	18	247	N/A
		All (11 studies)[111, 113, 126, 133, 141, 145, 146, 155, 168, 170, 174]		**1.30 (1.06 to 1.59)**	**534**	**3096**	**0.0%**
	Hom MT (AA) *vs* Hom WT (GG)	China (3 studies)[111, 145, 170]	1 study - 100% Chinese, 2 studies - NR	1.63 (0.66 to 4.00)	165	1356	58.1%
		Taiwan (2 studies)[141, 168]	NR	1.52 (0.68 to 3.40)	74	250	0.0%
		South Korea (1 study)[133]	NR	5.26 (1.61 to 17.26)	39	107	N/A
		Indonesia (1 study)[174]	100% Indonesian	3.29 (1.34 to 8.08)	29	102	N/A
		India (1 study)[126]	NR	0.64 (0.22 to 1.88)	25	100	N/A
		Tunisia (1 study)[146]	NR	3.71 (0.44 to 31.26)	9	28	N/A
		Turkey (1 study)[113]	NR	9.11 (1.91 to 43.46)	15	44	N/A
		Brazil (1 study)[155]	NR	1.25 (0.07 to 23.62)	17	246	N/A
		All (11 studies)[111, 113, 126, 133, 141, 145, 146, 155, 168, 170, 174]		**2.05 (1.24 to 3.40)**	**373**	**2233**	**47.7%**
803A-G (rs1208)	Het (AG) *vs* Hom WT (AA)	China (1 study)[111]	NR	1.63 (0.45 to 5.94)	101	107	N/A
		Taiwan (1 study)[168]	NR	1.36 (0.14 to 13.30)	70	285	N/A
		Indonesia (1 study)[174]	100% Indonesian	1.15 (0.55 to 2.41)	49	187	N/A
		Brazil (1 study)[155]	NR	0.82 (0.27 to 2.52)	13	219	N/A
		All (4 studies)[111, 155, 168, 174]		**1.14 (0.67 to 1.96)**	**233**	**798**	**0.0%**
	Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[111]	NR	Data excluded[c]			
		Taiwan (1 study)[168]	NR	Data excluded[c]			
		Indonesia (1 study)[174]	100% Indonesian	0.99 (0.11 to 9.09)	38	150	N/A
		Brazil (1 study)[155]	NR	2.32 (0.69 to 7.78)	12	140	N/A
		All (2 studies)[155, 174]		**1.90 (0.66 to 5.52)**	**50**	**290**	**0.0%**

NAT2 SNP	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
857G-A (rs1799931)	Het (GA) *vs* Hom WT (GG)	China (3 studies)[111, 145, 170]	1 study - 100% Chinese, 2 studies - NR	1.28 (0.74 to 2.22)	254	2069	61.5%
		Taiwan (2 studies)[141, 168]	NR	1.13 (0.70 to 1.82)	103	368	0.0%
		South Korea (1 study)[133]	NR	1.11 (0.56 to 2.20)	65	150	N/A
		Indonesia (1 study)[174]	100% Indonesian	1.41 (0.72 to 2.75)	49	190	N/A
		Tunisia (1 study)[146]	NR	0.70 (0.03 to 15.34)	14	52	N/A
		Turkey (1 study)[113]	NR	3.39 (0.84 to 13.67)	29	69	N/A
		Brazil (1 study)[155]	NR	2.19 (0.73 to 6.55)	17	250	N/A
		All (10 studies)[111, 113, 133, 141, 145, 146, 155, 168, 170, 174]		**1.30 (1.03 to 1.64)**	**531**	**3148**	**0.9%**
	Hom MT (AA) *vs* Hom WT (GG)	China (3 studies)[111, 145, 170]	1 study - 100% Chinese, 2 studies - NR	0.98 (0.38 to 2.51)	184	1677	0.0%
		Taiwan (2 studies)[141, 168]	NR	5.05 (0.47 to 54.88)	82	268	74.2%
		South Korea (1 study)[133]	NR	1.18 (0.10 to 13.36)	50	118	N/A
		Indonesia (1 study)[174]	100% Indonesian	4.31 (0.26 to 70.80)	33	139	N/A
		Tunisia (1 study)[146]	NR	Data excluded[c]			
		Turkey (1 study)[113]	NR	2.71 (0.16 to 45.03)	25	66	N/A
		Brazil (1 study)[155]	NR	8.75 (0.74 to 103.44)	13	212	N/A
		All (9 studies)[111, 113, 133, 141, 145, 155, 168, 170, 174]		**1.99 (1.02 to 3.91)**	**387**	**2480**	**11.3%**
Rs1495741	Het (AG) *vs* Hom WT (AA)	Taiwan (1 study)[82]	NR	0.19 (0.07 to 0.52)	19	249	N/A
	Hom MT (GG) *vs* Hom WT (AA)	Taiwan (1 study)[82]	NR	0.07 (0.01 to 0.56)	14	152	N/A
Rs4646244	Het (TA) *vs* Hom WT (TT)	South Korea (1 study)[133]	NR	2.03 (1.09 to 3.78)	57	152	N/A
	Hom MT (AA) *vs* Hom WT (TT)	South Korea (1 study)[133]	NR	4.06 (1.36 to 12.13)	37	110	N/A
Rs4646267	Het (AG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.50 (0.25 to 0.98)	52	127	N/A
	Hom MT (GG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.63 (0.27 to 1.45)	35	66	N/A

Emboldened text is used for results from meta-analyses.

[a] Total number of patients who experienced hepatotoxicity across the two genotype groups being compared

[b] Total number of patients who did not experience hepatotoxicity across the two genotype groups being compared

[c] Data were excluded due to zero patients in one of the genotype groups.

[d] Data from 2[111, 145] of the 3 Chinese studies were excluded due to zero counts.

CI: confidence interval; Het: heterozygous; Hom: homozygous; MT: mutant-type; N/A: not applicable; NR: not reported; OR: odds ratio; SNP: single nucleotide polymorphism; WT: wild-type

The main findings from these meta-analyses are:

- For 282C-T, there was strong evidence against the null hypothesis of no association for the comparison of TT genotype *versus* CC genotype, with the observed OR indicating an almost 4-fold increase in the odds of experiencing hepatotoxicity among patients with TT genotype (OR=3.95, 95% CI: 2.21 to 7.05, p<0.001); minimal heterogeneity was observed between the four studies included in this analysis (I^2=5.5%). However, there was no evidence to support an important association for the comparison of CT genotype *versus* CC genotype for this same SNP (OR=1.27, 95% CI: 0.80 to 2.02, p=0.302, I^2=0.0%, 4 studies).
- For 481C-T, there was some evidence to suggest that individuals with the CT genotype were more likely to experience hepatotoxicity in comparison to individuals with the CC genotype (OR=1.48, 95% CI: 1.12 to 1.97, p=0.006, I^2=0.0%, 9 studies). The point estimate of the OR suggested a greater association for the comparison of TT genotype *versus* CC genotype (OR=1.91, 95% CI: 0.93 to 3.92, p=0.076, I^2=34.1%, 9 studies); however, the 95% CIs were wide due to the relatively small number of homozygous mutant-type (TT) patients (n=162) among the patients contributing data to this analysis (n=1757). The evidence of an association for this comparison is therefore weak.
- For 590G-A and 857G-A, there was some evidence of an association for both comparisons between heterozygous and homozygous wild-type genotypes, and between homozygous mutant-type and homozygous wild-type genotypes (590G-A: GA *versus* GG, OR=1.30, 95% CI: 1.06 to 1.59, p=0.011, I^2=0.0%, 11 studies; AA *versus* GG, OR=2.05, 95% CI: 1.24 to 3.40, p=0.005, I^2=47.7%, 11 studies; 857G-A: GA *versus* GG, OR=1.30, 95% CI: 1.03 to 1.64, p=0.026, I^2=0.9%, 10 studies; AA *versus* GG, OR=1.99, 95% CI: 1.02 to 3.91, p=0.045, I^2=11.3%, 10 studies).

Results were relatively homogeneous between studies for most comparisons, except for the comparison between the AA and GG genotypes for the 590G-A SNP (I^2=47.7%). This moderate heterogeneity may be due to the variable distribution of genotypes in different geographic areas, or due to differences in definitions of hepatotoxicity in the 11 included studies[111, 113, 126, 133, 141, 145, 146, 155, 168, 170, 174] (one study did not provide the definition of hepatotoxicity and the remaining 10 studies used 10 different definitions of hepatotoxicity).

Secondary analyses: *NAT2* variants and other toxicity outcomes

A summary of all data for the association between *NAT2* variants and toxicity outcomes (other than hepatotoxicity) is provided in Table 17. Each reported result is based on data

from a single study, as there were no comparisons where more than one study provided data.

Table 17 Results of the secondary analyses: NAT2 *genetic variants and other toxicity outcomes*

Outcome	Variant	Study	Country	Ethnicity	Comparison	OR (95% CI)	# cases[a]	# controls[b]
Peripheral neuropathy	Acetylator status	Azuma 2013[81]	Japan	NR	Intermediate *vs* rapid	1.36 (0.32 to 5.75)	8	104
					Slow *vs* rapid	4.29 (0.66 to 27.8)	6	67
	191G-A (rs1801279)	Dhoro 2013[120]	Zimbabwe	NR	Het (GA) *vs* Hom WT (GG)	0.69 (0.33 to 1.41)	102	56
					Hom MT (AA) *vs* Hom WT (GG)	2.48 (0.12 to 53.02)	79	38
	341T-C (rs1801280)	Dhoro 2013[120]	Zimbabwe	NR	Het (TC) *vs* Hom WT (TT)	1.01 (0.50 to 2.07)	84	48
					Hom MT (CC) *vs* Hom WT (TT)	1.34 (0.32 to 5.62)	54	30
Adverse DIH outcome	Acetylator status	Bose 2011[112]	India	NR	Slow *vs* rapid/intermediate	3.31 (1.03 to 10.62)	16	202
ADRs	Acetylator status	Costa 2012[119]	Brazil	84% Black/mixed race, 16% other	Slow *vs* rapid/intermediate	3.20 (1.31 to 7.80)	40	47
Skin rash	Acetylator status	Higuchi 2007 (GI: HIGUCHI)[128]	Japan	NR	Intermediate *vs* rapid	0.83 (0.32 to 2.19)	22	68
					Slow *vs* rapid	1.21 (0.27 to 5.46)	15	41
Eosinophilia	Acetylator status	Higuchi 2007 (GI: HIGUCHI)[128]	Japan	NR	Intermediate *vs* rapid	1.44 (0.60 to 3.45)	31	59
					Slow *vs* rapid	0.98 (0.22 to 4.35)	17	39
ATD-induced MPE	R197Q (590G-A, rs1799930)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (AA) or Het (GA) *vs* Hom WT (GG)	0.96 (0.50 to 1.84)	58	150
	G286E (857G-A, rs1799931)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (AA) or Het (GA) *vs* Hom WT (GG)	1.65 (0.86 to 3.18)	59	152
	-9796T-A (rs4646244)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (AA) or Het (TA) *vs* Hom WT (TT)	1.08 (0.59 to 2.00)	62	159
	-9601A-G (rs4646267)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (GG) or Het (AG) *vs* Hom WT (AA)	0.65 (0.33 to 1.27)	61	159
Gastrointestinal ADRs	Acetylator status	Possuelo 2008 (GI: POSSUELO)[151]	Brazil	57% White	Slow *vs* rapid/intermediate	1.18 (0.51 to 2.70)	33	207

ATD: anti-tuberculosis drug; ADR: adverse drug reaction; CI: confidence interval; DIH: drug-induced hepatotoxicity; GI: group identifier; Het: heterozygous; Hom: homozygous; MPE: maculopapular eruption; MT: mutant-type; NR: not reported; OR: odds ratio; WT: wild-type

[a] Total number of cases across the genotype groups being compared

[b] Total number of controls across the genotype groups being compared

Some evidence existed to suggest that slow acetylators are more likely to experience "adverse drug-induced hepatotoxicity outcomes" (definition unclear; OR=3.31, 95% CI: 1.03 to 10.62, 1 study) and "adverse drug reactions" (ADRs, defined as at least one of the following: gastric, joint, neuromuscular or skin reactions, hepatotoxicity; OR=3.20, 95% CI: 1.31 to 7.80, 1 study). The point estimates of the ORs suggest that these associations are large in magnitude; however 95% CIs around these estimates were wide due to fairly small numbers of patients contributing data to these analyses (adverse drug-induced hepatotoxicity outcomes, n=218; ADRs, n=87). Otherwise, the strength of the evidence to support associations between *NAT2* genetic variants and anti-TB drug-induced toxicity outcomes (other than hepatotoxicity) was weak or non-existent.

Secondary analyses: *CYP* genetic variants and hepatotoxicity

The included studies reported data for the association between 24 SNPs of eight *CYP* genes (in addition to the key *CYP2E1* SNPs reported in the primary analyses) and hepatotoxicity. A summary of all data for the association between *CYP* genetic variants (other than the key *CYP2E1* SNPs) and hepatotoxicity is provided in Table 18. There were sufficient data to perform meta-analyses for three SNPs, and forest plots showing the results of these meta-analyses are provided in Appendix 4.

Table 18 Results of the secondary analyses: CYP genetic variants and hepatotoxicity

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
CYP2E1	rs2080672	Het (AG) *vs* Hom WT (AA)	China (1 study)[161]	NR	1.16 (0.72 to 1.89)	86	334	N/A
		Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[161]	NR	0.69 (0.19 to 2.42)	54	228	N/A
	rs915908	Het (GA) *vs* Hom WT (GG)	China (1 study)[161]	NR	0.89 (0.47 to 1.69)	79	318	N/A
		Hom MT (AA) *vs* Hom WT (GG)	China (1 study)[161]	NR	1.09 (0.52 to 2.32)	75	292	N/A
	rs8192775	Het (GA) *vs* Hom WT (GG)	China (1 study)[161]	NR	1.17 (0.72 to 1.90)	85	333	N/A
		Hom MT (AA) *vs* Hom WT (GG)	China (1 study)[161]	NR	0.76 (0.25 to 2.29)	55	234	N/A
	rs2515641	Het (CT) *vs* Hom WT (CC)	China (1 study)[161]	NR	1.20 (0.73 to 1.99)	85	342	N/A
		Hom MT (TT) *vs* Hom WT (CC)	China (1 study)[161]	NR	1.31 (0.41 to 4.18)	60	252	N/A
	rs2515644	Het (CA) *vs* Hom WT (CC)	China (1 study)[161]	NR	1.26 (0.74 to 2.15)	73	285	N/A
		Hom MT (AA) *vs* Hom WT (CC)	China (1 study)[161]	NR	1.04 (0.52 to 2.08)	42	186	N/A
	rs2070672	Het (AG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	1.74 (0.93 to 3.25)	63	149	N/A
		Hom MT (GG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.94 (0.18 to 4.85)	41	116	N/A
	rs2070673[c]	Het (TA) *vs* Hom WT (TT)	South Korea (1 study)[133]	NR	0.88 (0.48 to 1.63)	59	134	N/A
		Hom MT (AA) *vs* Hom WT (TT)	South Korea (1 study)[133]	NR	0.75 (0.28 to 1.96)	37	84	N/A
	96-bp (deletion-insertion SNP)	Het (DI) *vs* Hom WT (DD)	India (1 study)[154]	NR	1.13 (0.22 to 5.88)	6	98	N/A
			Brazil (1 study)[155]	NR	0.25 (0.01 to 4.26)	18	228	N/A
			All (2 studies)[154, 155]		**0.77 (0.19 to 3.21)**	**24**	**326**	**0.0%**
		Hom MT (II) *vs* Hom WT (DD)	India (1 study)[154]	NR	11.56 (1.37 to 97.67)	5	55	N/A
			Brazil (1 study)[155]	NR	3.72 (0.15 to 94.60)	18	207	N/A
			All (2 studies)[154, 155]		**8.20 (1.38 to 48.68)**	**23**	**262**	**0.0%**

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
CYP2C9	rs4918758[d]	Het (TC) *vs* Hom WT (TT)	China (1 study)[162]	NR	0.78 (0.46 to 1.33)	69	285	N/A
			South Korea (1 study)[133]	NR	1.66 (0.85 to 3.23)	59	127	N/A
			All (2 studies)[133, 162]		**1.11 (0.53 to 2.31)**	**128**	**412**	**66.7%**
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[162]	NR	0.94 (0.49 to 1.80)	51	188	N/A
			South Korea (1 study)[133]	NR	0.72 (0.27 to 1.95)	24	80	N/A
			All (2 studies)[133, 162]		**0.87 (0.51 to 1.50)**	**75**	**268**	**0.0%**
	rs9332098	Het (GA) *vs* Hom WT (GG)	China (1 study)[162]	NR	0.32 (0.07 to 1.38)	88	354	N/A
		Hom MT (AA) *vs* Hom WT (GG)	China (1 study)[162]	NR	Data excluded[e]			
	rs9332096	Het (CT) *vs* Hom WT (CC)	South Korea (1 study)[133]	NR	0.63 (0.27 to 1.47)	66	156	N/A
		Hom MT (TT) *vs* Hom WT (CC)	South Korea (1 study)[133]	NR	0.73 (0.03 to 18.24)	58	129	N/A
	rs1057910	Het (AC) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	1.00 (0.34 to 2.97)	64	154	N/A
		Hom MT (CC) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	Data excluded[e]			
CYP2B6	rs3745274	Het (GT) *vs* Hom WT (GG)	Brazil (1 study)[122]	NR	1.57 (0.71 to 3.45)	30	176	N/A
			Ethiopia (1 study)[173]	NR	1.42 (0.68 to 2.98)	35	145	N/A
			All (2 studies)[122, 173]		**1.49 (0.87 to 2.55)**	**65**	**321**	**0.0%**
		Hom MT (TT) *vs* Hom WT (GG)	Brazil (1 study)[122]	NR	0.58 (0.07 to 4.81)	13	103	N/A
			Ethiopia (1 study)[173]	NR	1.98 (0.66 to 5.87)	22	94	N/A
			All (2 studies)[122, 173]		**1.51 (0.55 to 4.13)**	**35**	**197**	**4.2%**
CYP3A4	rs12333983	Het (TA) *vs* Hom WT (TT)	China (1 study)[162]	NR	1.33 (0.81 to 2.18)	78	312	N/A
		Hom MT (AA) *vs* Hom WT (TT)	China (1 study)[162]	NR	1.33 (0.62 to 2.86)	47	204	N/A
	-392A-G	Het (GA) *vs* Hom WT (AA)	Brazil (1 study)[175]	42% white, 58% non-white	0.69 (0.32 to 1.47)	45	69	N/A
		Hom MT (GG) *vs* Hom WT (AA)	Brazil (1 study)[175]	42% white, 58% non-white	0.91 (0.31 to 2.70)	34	45	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
CYP2C19	rs11568732	Het (TG) *vs* Hom WT (TT)	China (1 study)[162]	NR	0.54 (0.25 to 1.19)	87	350	N/A
		Hom MT (GG) *vs* Hom WT (TT)	China (1 study)[162]	NR	0.93 (0.10 to 8.47)	80	229	N/A
	rs4986894	Het (TC) *vs* Hom WT (TT)	China (1 study)[162]	NR	0.95 (0.57 to 1.59)	72	302	N/A
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[162]	NR	1.11 (0.53 to 2.32)	48	191	N/A
	rs17878465	Het (CT) *vs* Hom WT (CC)	South Korea (1 study)[133]	NR	0.99 (0.50 to 1.94)	65	153	N/A
		Hom MT (TT) *vs* Hom WT (CC)	South Korea (1 study)[133]	NR	0.33 (0.02 to 6.58)	49	118	N/A
	rs4986893	Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[133]	NR	0.69 (0.31 to 1.56)	66	156	N/A
		Hom MT (AA) *vs* Hom WT (GG)	South Korea (1 study)[133]	NR	0.74 (0.03 to 18.42)	57	128	N/A
CYP3A5	rs776746	Het (AG) *vs* Hom WT (AA)	Brazil (1 study)[122]	NR	1.84 (0.83 to 4.05)	31	189	N/A
		Hom MT (GG) *vs* Hom WT (AA)	Brazil (1 study)[122]	NR	Data excluded[e]			
	Number of *CYP3A5*1*	One copy *vs* zero copies	Ethiopia (1 study)[173]	NR	1.56 (0.76 to 3.20)	39	151	N/A
		Two copies *vs* zero copies	Ethiopia (1 study)[173]	NR	1.02 (0.21 to 5.05)	24	110	N/A
CYP1A1	*Msp*I	Hom MT or Het *vs* Hom WT	China (1 study)[127]	NR	1.33 (0.81 to 2.19)	127	127	N/A
CYP2D6	rs1080983	Het (GA) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.83 (0.43 to 1.61)	65	152	N/A
		Hom MT (GG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.56 (0.06 to 5.11)	50	113	N/A
	rs1080989	Het (GA) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.89 (0.45 to 1.74)	50	121	N/A
		Hom MT (GG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	1.03 (0.47 to 2.27)	36	80	N/A

Emboldened text is used for results from meta-analyses

[a] Total number of patients who experienced hepatotoxicity across the genotype groups being compared

[b] Total number of patients who did not experience hepatotoxicity across the genotype groups being compared

[c] The paper (Kim 2009)[133] reports WT to be A and MT to be T, but data suggest that WT is T and MT is A.

[d] One of the studies (Kim 2009)[133] reports WT to be C and MT to be T, but the other study (Tang 2013b),[162] and the data, suggest that WT is T and MT is C

[e] Data were excluded due to zero counts in one of the genotype groups.

CI: confidence interval; Het: heterozygous; Hom: homozygous; MT: mutant-type; N/A: not applicable; NR: not reported; OR: odds ratio; WT: wild-type

For the 96-bp deletion-insertion SNP of the *CYP2E1* gene, there was some evidence to suggest that individuals with the II genotype were more likely to experience hepatotoxicity in comparison to individuals with the DD genotype (OR=8.20, 95% CI: 1.38 to 48.68, p=0.021, I^2=0.0%, 2 studies). The point estimate of the OR suggests that this association is large in magnitude; however the 95% CI around the estimate is extremely wide, and it is therefore difficult to draw firm conclusions about this association. There was no evidence of an association for the comparison of DI genotype *versus* DD genotype for this same SNP (OR=0.77, 95% CI: 0.19 to 3.21, p=0.722, I^2=0.0%, 2 studies). The meta-analyses conducted for the rs4918758 SNP of the *CYP2C9* gene and the rs3745274 SNP of the *CYP2B6* gene also provided no evidence of associations between these genetic variants and hepatotoxicity.

Heterogeneity was minimal in the meta-analyses conducted for these three SNPS, with the exception of the TC *versus* TT comparison for the rs4918758 SNP of the *CYP2C9* gene (I^2=66.7%). This heterogeneity may be due to the variable distribution of genotypes in different geographic areas, or due to the fact that the two studies (Kim 2009[133] and Tang 2013b[162]) included in this meta-analysis used different definitions of hepatotoxicity (see Table 13).

Secondary analyses: *CYP* genetic variants and other toxicity outcomes

A summary of all data for the association between *CYP* genetic variants and toxicity outcomes (other than hepatotoxicity) is provided in Table 19. It was not possible to perform meta-analyses for any toxicity outcomes other than hepatotoxicity as there were no comparisons for which more than one study provided data, so each reported result is based on data from a single study.

Table 19 Results of the secondary analyses: CYP genetic variants and other toxicity outcomes

Outcome	Gene	Variant	Study	Country	Ethnicity	Comparison	OR (95% CI)	# cases[a]	# controls[b]
Adverse DIH outcome	*CYP2E1*	7632T-A/*Dra*I (rs6413432)	Bose 2011[112]	India	NR	Hom MT (AA) or Het (TA) *vs* Hom WT (TT)	0.40 (0.13 to 1.23)	16	202
ADR	*CYP2E1*	7632T-A/*Dra*I (rs6413432)	Costa 2012[119]	Brazil	84% Black/mixed race, 16% other	Het (TA) *vs* Hom WT (TT)	Data excluded[c]		
						Hom MT (AA) *vs* Hom WT (TT)	Data excluded[c]		
		1293G-C/*Pst*I (rs3813867)	Costa 2012[119]	Brazil	84% Black/mixed race, 16% other	Het (GC) *vs* Hom WT (GG)	Data excluded[c]		
						Hom MT (CC) *vs* Hom WT (GG)	Data excluded[c]		
ATD-induced MPE	*CYP2E1*	1053C-T/*Rsa*I (rs2031920)[d]	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (TT) or Het (CT) *vs* Hom WT (CC)	0.87 (0.46 to 1.63)	60	153
		-352A-G (rs2070672)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (GG) or Het (AG) *vs* Hom WT (AA)	1.02 (0.53 to 1.96)	61	155
		-333A-T (rs2070673)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (AA) or Het (TA) *vs* Hom WT (TT)	1.12 (0.61 to 2.03)	61	154
	CYP2C9	-1188C-T (rs4918758)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (CC) or Het (TC) *vs* Hom WT (TT)	0.95 (0.50 to 1.81)	60	156
		-1565C-T (rs9332096)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (TT) or Het (CT) *vs* Hom WT (CC)	0.23 (0.07 to 0.78)	61	157
		I359L (rs1057910)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (CC) or Het (AC) *vs* Hom WT (AA)	1.31 (0.47 to 3.68)	60	154
	CYP2C19	W212X (rs4986893)	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (AA) or Het (GA) *vs* Hom WT (GG)	0.30 (0.10 to 0.88)	61	157
		-1418C-T	Kim 2011 (GI: KIM)[135]	South Korea	NR	Hom MT (TT) or Het (CT) *vs* Hom WT (CC)	0.57 (0.27 to 1.23)	59	156

[a] Total number of cases across the two genotype groups being compared

[b] Total number of controls across the two genotype groups being compared

[c] Data were excluded due to zero counts.

[d] The study (Kim 2011)[135] refers to this SNP as -1055C-T.

ATD: anti-tuberculosis drug; ADR: adverse drug reaction; CI: confidence interval; DIH: drug-induced hepatotoxicity; GI: group identifier; Het: heterozygous; Hom: homozygous; MPE: maculopapular eruption; MT: mutant-type; NR: not reported; OR: odds ratio; SNP: single nucleotide polymorphism; WT: wild-type

For the rs9332096 SNP of the *CYP2C9* gene and the rs4986893 SNP of the *CYP2C19* gene, some evidence existed to suggest that individuals with homozygous mutant-type or heterozygous genotype were less likely to experience anti-TB drug-induced maculopapular eruption in comparison to individuals with homozygous wild-type genotype (rs9332096: OR=0.23, 95% CI: 0.07 to 0.78, 1 study; rs4986893: OR=0.30, 95% CI: 0.10 to 0.88, 1 study). The point estimates of these ORs are indicative of a considerable protective effect; however 95% CIs around these estimates were wide due to the fairly small number of patients contributing data to these analyses (n=218 for both analyses). Otherwise, the strength of the evidence to support associations between *CYP* genetic variants and anti-TB drug-induced toxicity outcomes (other than hepatotoxicity) was weak or non-existent.

Secondary analyses: Other genetic variants and hepatotoxicity

The included studies reported data for 27 genes (83 SNPs) other than the *NAT2*, *CYP* and *GST* genes. A summary of all data for associations between other genetic variants and hepatotoxicity is provided in Table 20. There were sufficient data to perform meta-analyses for 14 SNPs of six different genes. Forest plots showing the results of these meta-analyses are provided in Appendix 4.

Table 20 Results of the secondary analyses: other genetic variants and hepatotoxicity

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
ABCB1	Rs1045642[c]	Het (CT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.13 (0.61 to 2.09)	58	137	N/A
			Ethiopia (1 study)[173]	NR	0.43 (0.18 to 1.06)	36	154	N/A
			All (2 studies)[137, 173]		**0.74 (0.29 to 1.89)**	**94**	**291**	**67.0%**
		Hom MT (TT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.01 (0.40 to 2.56)	35	88	N/A
			Ethiopia (1 study)[173]	NR	2.84 (0.81 to 10.00)	34	105	N/A
			All (2 studies)[137, 173]		**1.55 (0.57 to 4.23)**	**69**	**193**	**41.0%**
	Rs10261685	Het (TG) *vs* Hom WT (TT)	South Korea (1 study)[137]	NR	0.72 (0.32 to 1.62)	65	159	N/A
		Hom MT (GG) *vs* Hom WT (TT)	South Korea (1 study)[137]	NR	6.93 (0.28 to 172.71)	57	130	N/A
ABCC2	1774G-del[d]	Het (G/-) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	0.60 (0.32 to 1.14)	58	137	N/A
		Hom MT (-/-) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	0.69 (0.26 to 1.79)	33	63	N/A
	Rs1885301	Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	1.18 (0.65 to 2.15)	61	149	N/A
		Hom MT (AA) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	1.87 (0.56 to 6.29)	39	96	N/A
	Rs717620	Het (CT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.10 (0.61 to 2.01)	64	149	N/A
		Hom MT (TT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	0.81 (0.16 to 4.18)	40	98	N/A
	Rs2804400	Het (CT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.27 (0.70 to 2.30)	63	152	N/A
		Hom MT (TT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.53 (0.42 to 5.56)	38	98	N/A
	Rs2273697	Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	0.48 (0.17 to 1.32)	66	157	N/A
		Hom MT (AA) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	0.44 (0.02 to 9.25)	61	136	N/A
	Rs3740070	Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	1.00 (0.34 to 2.97)	66	159	N/A
		Hom MT (AA) *vs* Hom WT (GG)	South Korea (1 study)[137]	NR	Data excluded[e]	61	147	N/A
	Rs3740066	Het (CT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.21 (0.66 to 2.21)	61	147	N/A
		Hom MT (TT) *vs* Hom WT (CC)	South Korea (1 study)[137]	NR	1.44 (0.45 to 4.61)	40	100	N/A
BACH1	rs2300301	Het (AG) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.40 (0.13 to 1.29)	14	73	N/A
		Hom MT (GG) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	1.33 (0.31 to 5.72)	11	30	N/A
	rs1153285	Het (GA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	1.21 (0.39 to 3.75)	15	76	N/A
		Hom MT (AA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	2.83 (0.55 to 14.54)	9	40	N/A
	rs2070401	Het (TC) *vs* Hom WT (TT)	Japan (1 study)[148]	NR	1.13 (0.37 to 3.50)	15	81	N/A
		Hom MT (CC) *vs* Hom WT (TT)	Japan (1 study)[148]	NR	17.00 (1.59 to 182.14)	12	52	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
GSTP1	Ile105Val	Hom WT *vs* Het or Hom MT	China (1 study)[127]	NR	1.84 (1.10 to 3.07)	127	127	N/A
HMOX1	rs2071746	Het (TA) *vs* Hom WT (TT)	Japan (1 study)[148]	NR	2.67 (0.69 to 10.31)	15	70	N/A
		Hom MT (AA) *vs* Hom WT (TT)	Japan (1 study)[148]	NR	2.33 (0.41 to 13.26)	6	40	N/A
	rs2071749	Het (GA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	2.44 (0.77 to 7.67)	16	78	N/A
		Hom MT (AA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	4.10 (0.59 to 28.38)	7	45	N/A
	rs5755720	Het (AG) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.81 (0.26 to 2.55)	16	61	N/A
		Hom MT (GG) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.32 (0.06 to 1.76)	8	41	N/A
HSPA1L	rs2227956	Het (TC) *vs* Hom WT (TT)	China (1 study)[167]	100% Chinese	1.98 (1.21 to 3.23)	82	321	N/A
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[167]	100% Chinese	1.16 (0.45 to 3.00)	42	223	N/A
IL4	rs2243289	Het (GA) *vs* Hom WT (GG)	China (1 study)[169]	100% Chinese	0.84 (0.50 to 1.41)	85	341	N/A
		Hom MT (AA) *vs* Hom WT (GG)	China (1 study)[169]	100% Chinese	1.43 (0.37 to 5.54)	64	240	N/A
	rs2243250	Het (TC) *vs* Hom WT (TT)	China (1 study)[169]	100% Chinese	0.82 (0.49 to 1.38)	86	346	N/A
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[169]	100% Chinese	1.62 (0.41 to 6.46)	65	242	N/A
	rs2070874	Het (TC) *vs* Hom WT (TT)	China (1 study)[169]	100% Chinese	0.80 (0.47 to 1.35)	85	343	N/A
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[169]	100% Chinese	1.62 (0.41 to 6.44)	65	241	N/A
IL6	rs2066992	Het (GT) *vs* Hom WT (GG)	China (1 study)[167]	100% Chinese	1.15 (0.69 to 1.92)	79	335	N/A
		Hom MT (TT) *vs* Hom WT (GG)	China (1 study)[167]	100% Chinese	2.35 (1.03 to 5.36)	60	242	N/A
	rs2069837	Het (AG) *vs* Hom WT (AA)	China (1 study)[167]	100% Chinese	1.41 (0.85 to 2.33)	87	335	N/A
		Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[167]	100% Chinese	0.57 (0.13 to 2.57)	59	259	N/A
	rs1524107	Het (CT) *vs* Hom WT (CC)	China (1 study)[167]	100% Chinese	1.18 (0.71 to 1.96)	80	333	N/A
		Hom MT (TT) *vs* Hom WT (CC)	China (1 study)[167]	100% Chinese	1.99 (0.85 to 4.63)	59	241	N/A
IL10	rs1800896	Het (AG) *vs* Hom WT (AA)	China (1 study)[169]	100% Chinese	0.90 (0.43 to 1.87)	88	353	N/A
		Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[169]	100% Chinese	3.96 (0.25 to 64.0)	79	310	N/A
	rs1800871	Het (TC) *vs* Hom WT (TT)	China (1 study)[169]	100% Chinese	1.11 (0.68 to 1.82)	79	316	N/A
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[169]	100% Chinese	1.13 (0.52 to 2.46)	52	213	N/A
	rs1800872	Het (AC) *vs* Hom WT (AA)	China (1 study)[169]	100% Chinese	1.12 (0.68 to 1.84)	78	314	N/A
		Hom MT (CC) *vs* Hom WT (AA)	China (1 study)[169]	100% Chinese	1.18 (0.54 to 2.57)	51	210	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
KEAP1	rs1048290	Het (GC) vs Hom WT (GG)	Japan (1 study)[148]	NR	1.12 (0.30 to 4.13)	12	64	N/A
		Hom MT (CC) vs Hom WT (GG)	Japan (1 study)[148]	NR	1.92 (0.47 to 7.83)	10	41	N/A
	rs11545829	Het (CT) vs Hom WT (CC)	Japan (1 study)[148]	NR	0.97 (0.31 to 3.06)	14	71	N/A
		Hom MT (TT) vs Hom WT (CC)	Japan (1 study)[148]	NR	1.82 (0.45 to 7.39)	11	46	N/A
MAFF	rs2413508	Het (CG) vs Hom WT (CC)	Japan (1 study)[148]	NR	1.04 (0.32 to 3.41)	13	72	N/A
		Hom MT (GG) vs Hom WT (CC)	Japan (1 study)[148]	NR	2.83 (0.71 to 11.27)	11	44	N/A
	rs2267373	Het (TC) vs Hom WT (TT)	Japan (1 study)[148]	NR	1.86 (0.55 to 6.27)	13	67	N/A
		Hom MT (CC) vs Hom WT (TT)	Japan (1 study)[148]	NR	2.40 (0.60 to 9.52)	10	51	N/A
	rs2235264	Het (AG) vs Hom WT (AA)	Japan (1 study)[148]	NR	1.49 (0.36 to 6.14)	11	67	N/A
		Hom MT (GG) vs Hom WT (AA)	Japan (1 study)[148]	NR	3.73 (0.83 to 16.71)	10	39	N/A
	rs4821765	Het (TC) vs Hom WT (TT)	Japan (1 study)[148]	NR	0.56 (0.17 to 1.88)	17	82	N/A
		Hom MT (CC) vs Hom WT (TT)	Japan (1 study)[148]	NR	11.89 (0.46 to 308.42)	14	53	N/A
MAFK	rs4720833	Het (GA) vs Hom WT (GG)	Japan (1 study)[148]	NR	4.03 (1.30 to 12.52)	18	74	N/A
		Hom MT (AA) vs Hom WT (GG)	Japan (1 study)[148]	NR	0.49 (0.02 to 9.64)	5	53	N/A
	rs3808337	Het (TC) vs Hom WT (TT)	Japan (1 study)[148]	NR	2.28 (0.76 to 6.81)	17	74	N/A
		Hom MT (CC) vs Hom WT (TT)	Japan (1 study)[148]	NR	0.85 (0.09 to 8.09)	7	49	N/A
NFE2L2	rs2886161	Het (CT) vs Hom WT (CC)	Japan (1 study)[148]	NR	1.25 (0.41 to 3.75)	16	65	N/A
		Hom MT (TT) vs Hom WT (CC)	Japan (1 study)[148]	NR	0.54 (0.10 to 2.88)	9	49	N/A
	rs4243387	Het (TC) vs Hom WT (TT)	Japan (1 study)[148]	NR	0.76 (0.24 to 2.37)	17	76	N/A
		Hom MT (CC) vs Hom WT (TT)	Japan (1 study)[148]	NR	0.68 (0.07 to 6.20)	13	55	N/A
	rs6726395	Het (GA) vs Hom WT (GG)	Japan (1 study)[148]	NR	0.41 (0.12 to 1.40)	15	72	N/A
		Hom MT (AA) vs Hom WT (GG)	Japan (1 study)[148]	NR	1.04 (0.24 to 4.44)	14	48	N/A
	rs2001350	Het (AG) vs Hom WT (AA)	Japan (1 study)[148]	NR	0.64 (0.19 to 2.19)	16	76	N/A
		Hom MT (GG) vs Hom WT (AA)	Japan (1 study)[148]	NR	1.39 (0.25 to 7.76)	14	56	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
NOS2A	rs10459953	Het (CG) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.55 (0.17 to 1.79)	14	59	N/A
		Hom MT (GG) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.52 (0.13 to 2.04)	11	44	N/A
	rs3794764	Het (GA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	1.05 (0.36 to 3.04)	17	80	N/A
		Hom MT (AA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	2.40 (0.20 to 29.10)	11	50	N/A
	rs12944039	Het (GA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	1.22 (0.42 to 3.58)	16	80	N/A
		Hom MT (AA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	5.50 (0.67 to 44.90)	10	46	N/A
	rs11080344	Het (CT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.40 (0.13 to 1.19)	17	69	N/A
		Hom MT (TT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.20 (0.02 to 1.74)	12	42	N/A
	rs2314810	Het (GC) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	0.91 (0.32 to 2.58)	18	80	N/A
		Hom MT (CC) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	0.83 (0.04 to 18.41)	11	49	N/A
	rs3729966	Het (CT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.73 (0.25 to 2.14)	16	76	N/A
		Hom MT (TT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	1.33 (0.23 to 7.89)	10	38	N/A
	rs944722	Het (TC) *vs* Hom WT (TT)	Japan (1 study)[148]	NR	0.84 (0.29 to 2.42)	17	79	N/A
		Hom MT (CC) *vs* Hom WT (TT)	Japan (1 study)[148]	NR	1.43 (0.13 to 15.26)	11	46	N/A
	rs2255929	Het (AT) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.68 (0.23 to 2.02)	16	73	N/A
		Hom MT (TT) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.84 (0.15 to 4.59)	11	43	N/A
	rs3794756	Het (CT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.31 (0.08 to 1.20)	15	79	N/A
		Hom MT (TT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	3.67 (0.65 to 20.54)	15	47	N/A
NQO1	609C-T (rs1800566)	Het (CT) *vs* Hom WT (CC)	Taiwan (1 study)[130]	NR	0.93 (0.40 to 2.16)	53	49	N/A
			Japan (1 study)[148]	NR	1.45 (0.47 to 4.41)	16	71	N/A
			All (2 studies)[130, 148]		**1.09 (0.56 to 2.14)**	**69**	**120**	**0.0%**
		Hom MT (TT) *vs* Hom WT (CC)	Taiwan (1 study)[130]	NR	0.63 (0.22 to 1.83)	27	29	N/A
			Japan (1 study)[148]	NR	1.00 (0.18 to 5.70)	8	44	N/A
			All (2 studies)[130, 148]		**0.72 (0.29 to 1.78)**	**35**	**73**	**0.0%**
	rs689452	Het (CG) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	1.29 (0.44 to 3.82)	16	71	N/A
		Hom MT (GG) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.91 (0.17 to 4.91)	10	51	N/A
	rs2917669	Het (CT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	1.37 (0.46 to 4.05)	16	71	N/A
		Hom MT (TT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	0.93 (0.17 to 5.03)	10	52	N/A
	rs10517	Het (CT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	1.71 (0.56 to 5.22)	16	71	N/A
		Hom MT (TT) *vs* Hom WT (CC)	Japan (1 study)[148]	NR	1.09 (0.19 to 6.20)	8	47	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
PXR	Rs3814055	Het (CT) *vs* Hom WT (CC)	Taiwan (2 studies[f])[168]	NR	1.19 (0.75 to 1.88)	99	414	0.0%
			Indonesia (1 study)[176]	NR	1.07 (0.43 to 2.66)	30	69	N/A
			All (3 studies)[168, 176]		**1.16 (0.77 to 1.75)**	**129**	**483**	**0.0%**
		Hom MT (TT) *vs* Hom WT (CC)	Taiwan (2 studies[f])[168]	NR	1.19 (0.43 to 3.32)	68	298	0.0%
			Indonesia (1 study)[176]	NR	5.88 (1.05 to 32.85)	25	49	N/A
			All (3 studies)[168, 176]		**1.86 (0.65 to 5.33)**	**93**	**347**	**23.4%**
	Rs12488820	Het (CT) *vs* Hom WT (CC)	Taiwan (2 studies[f])[168]	NR	Data excluded[e]			
		Hom MT (TT) *vs* Hom WT (CC)	**Taiwan (2 studies[f])[168]**	**NR**	**0.93 (0.31 to 2.81)**	**103**	**432**	**0.0%**
	Rs2461823	Het (GA) *vs* Hom WT (GG)	**Taiwan (2 studies[f])[168]**	**NR**	**0.84 (0.52 to 1.36)**	**83**	**377**	**0.0%**
		Hom MT (AA) *vs* Hom WT (GG)	**Taiwan (2 studies[f])[168]**	**NR**	**1.60 (0.86 to 2.98)**	**59**	**211**	**0.0%**
	Rs7643645	Het (AG) *vs* Hom WT (AA)	**Taiwan (2 studies[f])[168]**	**NR**	**1.29 (0.59 to 2.80)**	**74**	**334**	**48.6%**
		Hom MT (GG) *vs* Hom WT (AA)	**Taiwan (2 studies[f])[168]**	**NR**	**1.64 (0.89 to 3.04)**	**52**	**226**	**0.0%**
	Rs6785049	Het (GA) *vs* Hom WT (GG)	**Taiwan (2 studies[f])[168]**	**NR**	**1.12 (0.70 to 1.80)**	**94**	**360**	**0.0%**
		Hom MT (AA) *vs* Hom WT (GG)	**Taiwan (2 studies[f])[168]**	**NR**	**0.56 (0.26 to 1.19)**	**44**	**213**	**0.0%**
	Rs3814057	Het (AC) *vs* Hom WT (AA)	**Taiwan (2 studies[f])[168]**	**NR**	**1.98 (1.06 to 3.69)**	**78**	**343**	**0.0%**
		Hom MT (CC) *vs* Hom WT (AA)	**Taiwan (2 studies[f])[168]**	**NR**	**2.18 (1.07 to 4.44)**	**40**	**193**	**0.0%**
SLC10A1	rs4646285	Het (GA) *vs* Hom WT (GG)	China (1 study)[117]	NR	1.05 (0.61 to 1.80)	86	348	N/A
		Hom MT (AA) *vs* Hom WT (GG)	China (1 study)[117]	NR	2.05 (0.50 to 8.41)	67	268	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
SLCO1B1	Rs4149013	Het (AG) *vs* Hom WT (AA)	China (1 study)[117]	NR	1.61 (0.95 to 2.73)	88	349	N/A
			South Korea (1 study)[137]	NR	1.76 (0.93 to 3.31)	66	150	N/A
			All (2 studies)[117, 137]		**1.67 (1.12 to 2.50)**	**154**	**499**	**0.0%**
		Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[117]	NR	0.89 (0.10 to 7.78)	63	282	N/A
			South Korea (1 study)[137]	NR	0.24 (0.01 to 4.46)	43	120	N/A
			All (2 studies)[117, 137]		**0.56 (0.10 to 3.19)**	**106**	**402**	**0.0%**
	Rs4149014[g]	Het (TG) *vs* Hom WT (TT)	China (1 study)[117]	NR	0.86 (0.54 to 1.39)	87	309	N/A
			South Korea (1 study)[137]	NR	0.78 (0.43 to 1.45)	61	147	N/A
			All (2 studies)[117, 137]		**0.83 (0.57 to 1.21)**	**148**	**456**	**0.0%**
		Hom MT (GG) *vs* Hom WT (TT)	China (1 study)[117]	NR	0.15 (0.03 to 0.63)	48	197	N/A
			South Korea (1 study)[137]	NR	1.21 (0.38 to 3.87)	43	92	N/A
			All (2 studies)[117, 137]		**0.44 (0.05 to 3.82)**	**91**	**289**	**81.5%**
	Rs2306283[h]	Het (GA) *vs* Hom WT (GG)	China (2 studies)[90, 117]	NR	1.10 (0.76 to 1.57)	188	471	0.0%
			South Korea (1 study)[137]	NR	1.12 (0.61 to 2.06)	59	145	N/A
			Ethiopia (1 study)[173]	NR	0.69 (0.32 to 1.50)	32	140	N/A
			All (4 studies)[90, 117, 137, 173]		**1.03 (0.77 to 1.38)**	**279**	**756**	**0.0%**
		Hom MT (AA) *vs* Hom WT (GG)	China (2 studies)[90, 117]	NR	1.31 (0.71 to 2.41)	135	330	0.0%
			South Korea (1 study)[137]	NR	1.40 (0.48 to 4.11)	39	96	N/A
			Ethiopia (1 study)[173]	NR	1.59 (0.60 to 4.21)	24	73	N/A
			All (4 studies)[90, 117, 137, 173]		**1.38 (0.87 to 2.21)**	**198**	**499**	**0.0%**
	Rs4149056	Het (TC) *vs* Hom WT (TT)	China (2 studies)[90, 117]	NR	1.58 (0.42 to 5.89)	204	505	88.8%
			South Korea (1 study)[137]	NR	1.23 (0.65 to 2.32)	66	153	N/A
			Ethiopia (1 study)[173]	NR	1.05 (0.50 to 2.21)	40	156	N/A
			All (4 studies)[90, 117, 137, 173]		**1.35 (0.74 to 2.43)**	**310**	**814**	**69.4%**
		Hom MT (CC) *vs* Hom WT (TT)	China (2 studies)[90, 117]	NR	3.98 (0.64 to 24.66)	158	417	0.0%
			South Korea (1 study)[137]	NR	0.35 (0.02 to 6.88)	46	116	N/A
			Ethiopia (1 study)[173]	NR	0.99 (0.11 to 9.23)	28	111	N/A
			All (4 studies)[90, 117, 137, 173]		**1.62 (0.45 to 5.79)**	**232**	**644**	**0.0%**
	Rs2291075	Het (CT) *vs* Hom WT (CC)	China (1 study)[117]	NR	0.95 (0.54 to 1.65)	68	262	N/A
		Hom MT (TT) *vs* Hom WT (CC)	China (1 study)[117]	NR	0.81 (0.42 to 1.56)	45	185	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
SOD1	rs2070424	Hom MT (GG) or Het (GA) *vs* Hom WT (AA)	South Korea (1 study)[139]	100% Korean	2.28 (1.16 to 4.48)	84	236	N/A
SOD2 (MnSOD)	47T-C (rs4880)	Hom MT (CC) or Het (CT) *vs* Hom WT (TT)	Taiwan (1 study)[130]	NR	2.40 (1.12 to 5.16)	63	63	N/A
			South Korea (1 study)[139]	100% Korean	1.27 (0.71 to 2.27)	83	237	N/A
			All (2 studies)[130, 139]		**1.66 (0.89 to 3.08)**	**146**	**300**	**40.9%**
		Het (TC) *vs* Hom WT (TT)	Taiwan (1 study)[130]	NR	2.56 (1.12 to 5.81)	59	60	N/A
		Hom MT (CC) *vs* Hom WT (TT)	Taiwan (1 study)[130]	NR	1.78 (0.37 to 8.44)	40	51	N/A
SOD3	rs1799895	Hom MT (AA) or Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[139]	100% Korean	1.03 (0.62 to 1.71)	83	234	N/A
	rs2536512	Hom MT (GG) or Het (CG) *vs* Hom WT (CC)	South Korea (1 study)[139]	100% Korean	1.81 (0.72 to 4.54)	84	237	N/A
STAT3	rs1053004	Het (TC) *vs* Hom WT (TT)	China (1 study)[167]	100% Chinese	0.90 (0.54 to 1.48)	78	300	N/A
		Hom MT (CC) *vs* Hom WT (TT)	China (1 study)[167]	100% Chinese	0.75 (0.36 to 1.58)	49	191	N/A
	rs1053023	Het (AG) *vs* Hom WT (AA)	China (1 study)[167]	100% Chinese	1.26 (0.72 to 2.19)	64	293	N/A
		Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[167]	100% Chinese	2.15 (1.14 to 4.07)	49	187	N/A
	rs1053005	Het (AG) *vs* Hom WT (AA)	China (1 study)[167]	100% Chinese	1.00 (0.61 to 1.63)	80	310	N/A
		Hom MT (GG) *vs* Hom WT (AA)	China (1 study)[167]	100% Chinese	0.79 (0.36 to 1.76)	49	199	N/A
TNF-alpha	-308G-A	Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[138]	100% Korean	1.80 (0.95 to 3.43)	75	228	N/A
		Hom MT (AA) *vs* Hom WT (GG)	South Korea (1 study)[138]	100% Korean	6.81 (0.61 to 76.44)	59	195	N/A

Gene	Variant	Comparison	Country (# studies)	Ethnicity	OR (95% CI)	# cases[a]	# controls[b]	I^2
UGT1A1	211G-A (rs4148323)	Het (GA) *vs* Hom WT (GG)	Taiwan (1 study)[115]	NR	0.99 (0.31 to 3.12)	17	81	N/A
			South Korea (1 study)[133]	NR	1.03 (0.55 to 1.93)	61	145	N/A
			All (2 studies)[115, 133]		**1.02 (0.59 to 1.77)**	**78**	**226**	**0.0%**
		Hom MT (AA) *vs* Hom WT (GG)	Taiwan (1 study)[115]	NR	Data excluded[e]			
			South Korea (1 study)[133]	NR	1.20 (0.39 to 3.73)	45	106	N/A
	rs3755319	Het (CA) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	1.55 (0.86 to 2.80)	63	150	N/A
		Hom MT (CC) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	0.83 (0.16 to 4.21)	33	97	N/A
	rs2003569	Het (GA) *vs* Hom WT (GG)	South Korea (1 study)[133]	NR	0.96 (0.49 to 1.87)	64	155	N/A
		Hom MT (AA) *vs* Hom WT (GG)	South Korea (1 study)[133]	NR	1.20 (0.11 to 13.53)	49	117	N/A
	686C-A	Het (CA) *vs* Hom WT (CC)	Taiwan (1 study)[115]	NR	5.27 (0.69 to 40.35)	17	81	N/A
		Hom MT (AA) *vs* Hom WT (CC)	Taiwan (1 study)[115]	NR	Data excluded[e]			
	$TA_6 \rightarrow TA_7$ at the promoter region	Het (TA_7TA_6) *vs* Hom WT (TA_6TA_6)	Taiwan (1 study)[115]	NR	1.12 (0.28 to 4.46)	17	81	N/A
		Hom MT (TA_7TA_7) *vs* Hom WT (TA_6TA_6)	Taiwan (1 study)[115]	NR	Data excluded[e]			
	1091C-T	Het (CT) *vs* Hom WT (CC)	Taiwan (1 study)[115]	NR	5.00 (0.30 to 84.17)	17	81	N/A
		Hom MT (TT) *vs* Hom WT (CC)	Taiwan (1 study)[115]	NR	Data excluded[e]			
UGT1A3	rs2008584	Het (GA) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	1.62 (0.89 to 2.93)	63	149	N/A
		Hom MT (GG) *vs* Hom WT (AA)	South Korea (1 study)[133]	NR	1.26 (0.31 to 5.17)	34	98	N/A
	rs6431625	Het (TC) *vs* Hom WT (TT)	South Korea (1 study)[133]	NR	1.36 (0.64 to 2.89)	54	143	N/A
		Hom MT (CC) *vs* Hom WT (TT)	South Korea (1 study)[133]	NR	8.42 (0.34 to 210.83)	42	116	N/A
UGT2B7	Rs7662029[i]	Het (GA) *vs* Hom WT (AA)	Ethiopia (1 study)[173]	NR	0.69 (0.32 to 1.50)	36	124	N/A
		Hom MT (GG) *vs* Hom WT (AA)	Ethiopia (1 study)[173]	NR	0.38 (0.12 to 1.15)	19	74	N/A
XPO1	rs7606167	Het (GC) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	0.90 (0.29 to 2.80)	15	73	N/A
		Hom MT (CC) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	1.56 (0.35 to 6.92)	12	51	N/A
	rs11125883	Het (AC) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.33 (0.10 to 1.07)	16	64	N/A
		Hom MT (CC) *vs* Hom WT (AA)	Japan (1 study)[148]	NR	0.27 (0.05 to 1.38)	13	45	N/A
	rs1050567	Het (GA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	0.59 (0.18 to 1.91)	14	76	N/A
		Hom MT (AA) *vs* Hom WT (GG)	Japan (1 study)[148]	NR	2.89 (0.67 to 12.42)	13	45	N/A

Emboldened text is used for results from meta-analysis.
[a] Total number of patients who experienced hepatotoxicity across the genotype groups being compared
[b] Total number of patients who did not experience hepatotoxicity across the genotype groups being compared
[c] One of the studies (Kim 2012b)[137] reports WT to be T and MT to be C, but the other study (Yimer 2011),[173] and the data, suggest that WT is C and MT is T.
[d] The study (Kim 2012b)[137] reports WT to be del (-) and MT to be G, but the data suggest that WT is G and MT is del (-).
[e] Data were excluded due to zero counts.
[f] Reported in the same paper (Wang 2015b),[168] but two separate cohorts of patients
[g] One of the studies (Kim 2012b)[137] reports WT to be G and MT to be T, but the other study (Chen 2015),[117] and the data, suggest that WT is T and MT is G.
[h] Three of the studies (Chen 2015,[117] Li 2012[90] and Yimer 2011[173]) report WT to be A and MT to be G, but the other study (Kim 2012b),[137] and the data, suggest that WT is G and MT is A.
[i] The paper (Yimer 2011)[173] reports WT to be G and MT to be A, but the data suggest that WT is A and MT is G.
CI: confidence interval; Het: heterozygous; Hom: homozygous; MT: mutant-type; N/A: not applicable; NR: not reported; OR: odds ratio; WT: wild-type

For the rs3814057 polymorphism of the *PXR* gene, there was some evidence to suggest that individuals with AC and CC genotypes are more likely to experience hepatotoxicity in comparison to individuals with AA genotype (AC *versus* AA: OR=1.98, 95% CI: 1.06 to 3.69, p=0.032, I^2=0.0%, 2 patient cohorts reported in the same article; CC *versus* AA: OR=2.18, 95% CI: 1.07 to 4.44, p=0.032, I^2=0.0%, 2 patient cohorts reported in the same article). The meta-analysis for the rs4149013 polymorphism of the *SLCO1B1* gene also provided evidence that individuals with AG genotype are more likely to experience hepatotoxicity in comparison to individuals with AA genotype (OR=1.67, 95% CI: 1.12 to 2.50, p=0.013, I^2=0.0%, 2 studies). However, the evidence for the comparison between GG genotype and AA genotype at this SNP was inconclusive, with the 95% CI for the OR including both considerable protective and harmful effects (OR=0.56, 95% CI: 0.10 to 3.19, p=0.515, I^2=0.0%, 2 studies). The strength of the evidence for associations from the meta-analyses for the remaining 12 SNPs and hepatotoxicity was either weak or non-existent.

Substantial heterogeneity was observed in the meta-analyses comparing: i) CT *versus* CC genotype for the rs1045642 SNP of the *ABCB1* gene; ii) GG *versus* TT genotype for the rs4149014 SNP of the *SLCO1B1* gene; iii) TC *versus* TT genotype for the rs4149056 SNP of the *SLCO1B1* gene. In each of these meta-analyses, each included study applied a different definition of hepatotoxicity. Variable frequencies of genotypes across the countries where studies were conducted may also be a source of heterogeneity in these analyses.

Secondary analyses: *GSTM1*/*GSTT1* and other toxicity outcomes

Data for the association between *GST* genetic variants and toxicity outcomes (other than hepatotoxicity) are summarised in Table 21. Each reported result is based on data from a single study.

One study[119] investigated associations between *GSTM1* and *GSTT1* homozygous null genotypes and ADRs (defined as the presence of at least one of the following symptoms during the follow-up period: gastric, joint, neuromuscular, or skin reactions; and hepatotoxicity). Another study[134] examined whether *GSTM1* and *GSTT1* homozygous null genotypes increased the likelihood of experiencing anti-TB drug-induced cutaneous reactions. Neither study provided evidence of an important association between *GSTM1* or *GSTT1* homozygous null genotypes and these toxicity outcomes.

Secondary analyses: Other genetic variants and other toxicity outcomes

Data for the association between other genetic variants and toxicity outcomes (other than hepatotoxicity) are summarised in Table 21. Each reported result is based on data from a single study.

One study[136] investigated associations between polymorphisms of the *ABCB1* and *ABCC2* genes and anti-TB drug-induced maculopapular eruption. For the rs1885301 and rs2804400 polymorphisms of the *ABCC2* gene, there was some evidence to suggest that homozygous mutant-type individuals are more likely to experience anti-TB drug-induced maculopapular eruption in comparison to homozygous wild-type individuals (rs1885301: OR=3.39, 95% CI: 1.13 to 10.14, 1 study; rs2804400: OR=3.47, 95% CI: 1.16 to 10.36, 1 study). The point estimates of these ORs suggest important effects; however, 95% CIs around these estimates were wide due to the fairly small number of patients contributing data to these analyses (rs1885301, n=134; rs2804400, n=136). There was no evidence to suggest that heterozygous genotype has an important impact on anti-TB drug-induced maculopapular eruption in comparison to homozygous wild-type genotype for either polymorphism. The strength of the evidence from this study[136] to support associations between other genetic variants of the *ABCB1* and *ABCC2* genes and anti-TB drug-induced maculopapular eruption was weak or non-existent.

Table 21 Results of the secondary analyses: association between GST and other genetic variants and toxicity outcomes (other than hepatotoxicity)

Study	Country	Ethnicity	Outcome	Gene	Variant	Comparison	OR (95% CI)	# cases[a]	# controls[b]
Costa (2012)[119]	Brazil	84% Black/ mixed race, 16% other	ADRs	*GSTM1*	Null or present	Hom null *vs* Het or Hom present	1.06 (0.42 to 2.72)	43	45
				GSTT1	Null or present	Hom null *vs* Het or Hom present	0.71 (0.29 to 1.73)	43	45
Kim 2012a (GI: KIM)[136]	South Korea	NR	ATD-induced MPE	*ABCB1/ MDR1*	I1145I (rs1045642)	Het (CT) *vs* Hom WT (CC)	1.02 (0.55 to 1.89)	57	137
						Hom MT (TT) *vs* Hom WT (CC)	0.49 (0.15 to 1.55)	32	88
					-114918T-G (rs10261685)	Het (TG) *vs* Hom WT (TT)	1.10 (0.52 to 2.32)	61	159
						Hom MT (GG) *vs* Hom WT (TT)	7.91 (0.32 to 197.42)	50	130
				ABCC2	1774G-del[c]	Het (G/-) *vs* Hom WT (GG)	0.62 (0.32 to 1.18)	54	137
						Hom MT (-/-) *vs* Hom WT (GG)	0.75 (0.29 to 1.96)	31	63
					-1549G-A (rs1885301)	Het (GA) *vs* Hom WT (GG)	1.19 (0.63 to 2.23)	54	149
						Hom MT (AA) *vs* Hom WT (GG)	3.39 (1.13 to 10.14)	38	96
					-24C-T (rs717620)	Het (CT) *vs* Hom WT (CC)	1.12 (0.60 to 2.10)	56	149
						Hom MT (TT) *vs* Hom WT (CC)	2.79 (0.84 to 9.25)	39	98
					IVS3-49C-T (rs2804400)	Het (CT) *vs* Hom WT (CC)	1.19 (0.64 to 2.23)	54	152
						Hom MT (TT) *vs* Hom WT (CC)	3.47 (1.16 to 10.36)	38	98
					V417L (rs2273697)	Het (GA) *vs* Hom WT (GG)	1.01 (0.44 to 2.32)	61	157
						Hom MT (AA) *vs* Hom WT (GG)	1.29 (0.11 to 14.52)	53	136
					S978S (rs3740070)	Het (GA) *vs* Hom WT (GG)	0.84 (0.26 to 2.73)	62	159
						Hom MT (AA) *vs* Hom WT (GG)	Data excluded[d]		
					I1324I (rs3740066)	Het (CT) *vs* Hom WT (CC)	1.40 (0.75 to 2.63)	54	147
						Hom MT (TT) *vs* Hom WT (CC)	2.79 (0.99 to 7.89)	37	100
Kim 2010 (GI: KIM)[134]	South Korea	NR	ATD-induced cutaneous reactions	*GSTM1*	Null or present	Hom null *vs* Het or Hom present	1.22 (0.74 to 2.01)	94	190
				GSTT1	Null or present	Hom null *vs* Het or Hom present	1.19 (0.72 to 1.96)	94	190

[a] Total number of cases across the genotype groups being compared

[b] Total number of controls across the genotype groups being compared

[c] The study (Kim 2012a)[136] reports WT to be del (-) and MT to be G, but the data suggest that WT is G and MT is del (-).

[d] Data were excluded due to zero counts.

ADR: adverse drug reaction; ATD: anti-tuberculosis drug; CI: confidence interval; GI: group identifier; Het: heterozygous; Hom: homozygous; MPE: maculopapular eruption; MT: mutant-type; NR: not reported; OR: odds ratio; WT: wild-type

4.3 Discussion

Our systematic review and meta-analysis has demonstrated that performing robust synthesis of the evidence base for associations between genetic variants and toxicity outcomes related to anti-TB drugs is challenging. The included studies varied in terms of the genetic variants selected for investigation, choice and definition of outcomes, how genotype groups were combined for analysis, ethnicity of participants and methodological quality. While conducting our review, we carefully considered these challenges, and stratified our meta-analyses by genetic variant, choice of genotype groups to be combined, and outcome. We also stratified further by the country in which the study was conducted as a proxy for ethnicity, which was not widely reported.

Meta-analyses

Where possible, meta-analyses were undertaken to improve power to estimate pharmacogenetic associations. It is important to note that we were only able to perform meta-analyses for the outcome of hepatotoxicity; there were insufficient data to perform pooled analyses for any other outcome. Included studies investigated the influence of genetic variants on various outcomes other than hepatotoxicity, including maculopapular eruption, peripheral neuropathy, skin rash, eosinophilia, and gastrointestinal ADRs. Very few statistically significant associations were observed; however, it is possible that some true associations were not identified due to the small sample sizes of the included studies. More studies investigating outcomes other than hepatotoxicity, and synthesis of the total evidence base would improve power to detect particular magnitudes of association as statistically significant, and also allow researchers to investigate the possibility that statistically significant associations identified in the initial studies may be spurious.

Considering the meta-analyses performed for hepatotoxicity, we found that *NAT2* slow/intermediate acetylators were significantly more likely to experience this outcome than *NAT2* rapid acetylators. This result is consistent with the findings of several meta-analyses,[92, 95, 98, 101] but is not consistent with the findings of the meta-analysis reported by Sun et al.,[99] who did not identify a significant association between slow acetylator status and hepatotoxicity. However, the search date for Sun et al.[99] (May 2007) is several years earlier than the search dates for the other meta-analyses, and many relevant studies have been published in recent years. As more studies are published, the power to detect particular magnitudes of association as statistically significant increases.

Interestingly, the result from a sensitivity analysis comparing the likelihood of hepatotoxicity for slow *versus* rapid/intermediate acetylators resulted in a stronger estimate of association (OR=3.12, 95% CI: 2.45 to 3.97) than our primary analysis (of slow/intermediate *versus* rapid acetylator status, OR=1.59, 95% CI: 1.26 to 2.01). This suggests that the way in which acetylator status groups are combined can have a considerable impact on the effect estimate observed. It is also worth noting that more studies combined rapid and intermediate acetylators than combined slow and intermediate acetylators, so the sensitivity analysis included data for a larger number of participants (n=7916) than the primary analysis for acetylator status (n=5036), and so was even more well-powered.

A recently conducted GWAS[197] in a Thai population also identified a strong association between *NAT2* acetylator status (where phenotypes were inferred by using *NAT2* rs1495741 as a tag SNP) and ATDH when assuming a recessive mode of inheritance, i.e. slow *versus* intermediate/rapid acetylators (OR=6.01, 95% CI: 3.42 to 10.57, p=6.86e-11). There were no other genes for which associations were identified that met the threshold for genome-wide significance (p<5e-8); this study therefore identified *NAT2* as the most important risk factor for ATDH in Thai TB patients.

Considering previously conducted meta-analyses of *CYP2E1* polymorphisms, three of these meta-analyses[94, 97, 102] performed analyses for the *Rsa*I and *Pst*I polymorphisms combined, presumably because these polymorphisms have been reported to be in linkage disequilibrium.[83] The approach taken for the analysis of *CYP2E1* polymorphisms in two other meta-analyses[92, 99] was unclear. However, we identified studies reporting data for these two polymorphisms separately,[86, 121] so we performed separate meta-analyses for each polymorphism.

We found that patients with homozygous wild-type (TT) or heterozygous (CT) genotype at the *CYP2E1 Rsa*I polymorphism were significantly less likely to experience hepatotoxicity than patients with homozygous mutant-type (CC) genotype. This result is consistent with the findings of four previously conducted meta-analyses.[94, 97, 99, 102] We observed no significant association for the *CYP2E1 Dra*I polymorphism and hepatotoxicity, a result which is consistent with previous meta-analyses.[97, 102] We also observed no significant association for the *CYP2E1 Pst*I polymorphism and hepatotoxicity; this result is not consistent with the findings of previously conducted meta-analyses.[94, 97, 102] This may be because we only included studies that explicitly stated that results were for the *Pst*I polymorphism or the

*Rsa*I/*Pst*I polymorphisms combined (if these alleles were in complete linkage disequilibrium in the study population). The previous meta-analyses synthesised data for the *Rsa*I/*Pst*I polymorphisms combined, and therefore may have additionally included data from studies that stated that reported data were for the *Rsa*I polymorphism (without mentioning the *Pst*I polymorphism). The number of studies contributing data to the analysis of the *CYP2E1 Pst*I polymorphism was relatively small (n=8) compared with the number of studies contributing data to the analysis of the *CYP2E1 Rsa*I polymorphism (n=23).

We found that for *GSTM1*, patients with homozygous null genotype were significantly more likely to experience hepatotoxicity than patients with heterozygous or homozygous present genotype. We also found that for *GSTT1*, there was no significant difference in the likelihood of experiencing hepatotoxicity between patients with homozygous null genotype and those with heterozygous or homozygous present genotype. These results are consistent with the findings of previously conducted meta-analyses, which all identified a significant association between *GSTM1* null genotype and hepatotoxicity, and no significant association between *GSTT1* null genotype and hepatotoxicity.[92, 93, 96, 99, 100] In particular, Cai et al.[92] and Tang et al.[100] reported very similar ORs for both the *GSTM1* and *GSTT1* homozygous null genotypes to those reported in the current review.

We also identified: for the 96-bp deletion-insertion SNP of the *CYP2E1* gene, there was some evidence to suggest that homozygous mutant-type (II) individuals are more likely to experience hepatotoxicity than homozygous wild-type (DD) individuals; for the rs3814057 polymorphism of the *PXR* gene, both the heterozygous (AC) and homozygous mutant-type (CC) genotypes were associated with an increased likelihood of hepatotoxicity risk in comparison to homozygous wild-type (AA) genotype. The data for the analysis of the *PXR* rs3814057 SNP came from two independent cohorts (a derivation cohort and a validation cohort) from a single study.[168] By combining results from these two cohorts in meta-analyses, the power to detect associations increases. However, it is important to note that the number of hepatotoxicity cases included in the analyses of both the 96-bp deletion-insertion SNP of the *CYP2E1* gene and the rs3814057 polymorphism of the *PXR* gene were fairly small, and therefore 95% CIs were wide. More data are required before it is possible to draw firm conclusions on the association between these two SNPs and hepatotoxicity.

To the best of our knowledge, no meta-analyses on genetic variants other than variants of the *NAT2*, *CYP2E1*, *GSTM1* and *GSTT1* genes have been published in this field, and therefore

our results add to the existing understanding of the association between genetic variants and ATDH.

Limitations

Firstly, it could be argued that there are limitations to performing primary analyses for variants of four different genes (*NAT2, CYP2E1, GSTM1* and *GSTT1*). Considering the existing literature, researchers have hypothesised that variants of each of these genes may play a role in the in the biological pathway between anti-TB treatment and hepatotoxicity; however, the NAT2 enzyme is the main isoniazid metabolising enzyme, and its activity has clearly been shown to be affected by common *NAT2* polymorphisms. Evidence for the relevance of *CYP2E1, GSTM1 and GSTT1* gene polymorphisms is less clear. In line with this existing knowledge, the strongest evidence of association between a genetic variant and hepatotoxicity identified in our meta-analyses was for *NAT2* acetylator status, and this ought to be considered as the main finding of our systematic review.

As discussed in Section 3.3, most included studies did not report the ethnic background of participants. We therefore performed analyses stratified by the country in which the study was conducted as a proxy variable for ethnicity. It is clear that this approach is not ideal as the population of any given country is often ethnically diverse. However, stratifying by country was deemed the most suitable approach in the absence of definitive information on ethnicity.

Our quality assessment using criteria developed by Jorgensen and Williamson[18] highlighted that there were many areas of methodological concern across the included studies, which we considered could introduce bias into the results of our analyses. As the quality assessment was qualitative rather than quantitative, it was not possible to exclude studies from our meta-analyses based on a single summary score. However, there were no studies which satisfied all criteria of the quality assessment checklist, and therefore, we advise that caution be applied when interpreting the results of the meta-analyses.

We also found that it was often difficult to identify distinct patient cohorts from the included articles. If multiple articles report data for the same patient cohort, data for this patient cohort must only be included in any given meta-analysis once, otherwise a unit-of-analysis error occurs.[25] We found that it was often not possible to determine from articles alone whether patient cohorts were distinct, overlapping, or identical. We contacted several study authors for clarification. For three articles,[108, 109, 152] we did not receive a response and, consequently, data from the older articles[108, 109] were excluded from meta-

analyses. If these three articles reported data for distinct patient cohorts, then information would have been lost by excluding two articles. Furthermore, there may have been cases of multiple articles reporting outcomes for the same patient cohorts that we did not identify; if this was the case, some patients may have been double-counted in the meta-analyses. It would be highly beneficial (for both systematic reviewers and the general reader) if pharmacogenetic researchers were to highlight in their study reports when the included patient cohort overlaps or is identical to the patient cohort of other articles.

An additional challenge encountered was the inconsistent use of SNP nomenclature, which made gathering data for meta-analyses problematic. For example, the *CYP2E1* SNPs considered in the primary analyses were referred to in various ways in the included studies. In particular, the *CYP2E1* SNP identified by rs2031920 was referred to in articles using one or more of the following: rs2031920, '*Rsa*I polymorphism', '*CYP2E1* 1053C-T', '*CYP2E1* -1019C-T', '*CYP2E1* -1055C-T'. Since rs numbers are unique to each SNP, in the first instance we identified studies reporting data for the same SNPs by using the rs numbers. If an article did not report the rs number, then we searched the literature to match the reported SNP (whatever nomenclature was used) to the rs number for that SNP. This process was especially challenging, as often the nomenclature used did not appear in standard databases, such as dbSNP[198] or pharmGKB.[6] We would recommend that where possible, authors of pharmacogenetic studies should state the unique rs number of each genotyped SNP. As discussed in the STREGA (STrengthening the REporting of Genetic Association studies) Statement,[199] guidelines are available for how to define variants that are not listed in dbSNP.[200, 201]

Furthermore, the considerable variability in the definitions of hepatotoxicity across the included studies (see Section 3.4) introduced heterogeneity into the meta-analyses. Interpretation of meta-analyses that include data for a variety of different outcome definitions is difficult, as there is not a single outcome definition that the estimate of association relates to. As previously discussed in Section 3.4, the development of a core outcome set for studies of anti-TB drug-related toxicity would be valuable; consistent outcome definitions would help to reduce heterogeneity and increase the robustness of findings from meta-analyses.

Finally, for the meta-analyses presented in this chapter, we either performed two pairwise comparisons, or assumed a specific mode of inheritance, so that standard methods of meta-analysis could be applied (see Section 2.3). However, these methods are not without

their limitations. If a specific mode of inheritance is assumed incorrectly, power to detect pharmacogenetic associations will be reduced. Furthermore, performing pairwise comparisons ignores any correlation between the effect estimates induced by the common baseline group; this is therefore an inefficient approach since the two effect estimates do not borrow strength from one another as they would in a bivariate analysis. We explore more complex methods of meta-analysis that have the potential to overcome the limitations of standard methods of meta-analysis in Chapter 5.

4.4 Conclusion

This review showed that slow acetylators were significantly more likely to experience ATDH than intermediate/rapid acetylators. We also observed significant associations between the *CYP2E1 Rsa*I and *GSTM1* null polymorphisms and ATDH. Therefore, pharmacogenetic testing may be useful in clinical practice in terms of risk stratification for ATDH during treatment of TB. Whilst the findings from our meta-analyses alone lack the strength of evidence required to support a stratified medicine approach at this time, they suggest that comprehensive genotyping in a wider range of populations (including African populations) is required to establish the value of pharmacogenetic testing in the treatment of TB. More studies are needed to overcome the reported methodological limitations of the existing studies and to assess if a stratified medicine approach might be feasible and cost effective.

Throughout the process of conducting this review, we observed that improvements in the reporting of pharmacogenetic studies would firstly, and importantly, enable readers to fully assess the strengths and weaknesses of individual studies, and interpret the results accordingly. However, improvements in the reporting of pharmacogenetic studies would also enable systematic reviewers to gain more insight into the methodological quality of included studies, include more data in meta-analyses, and tailor their analysis approaches based on information reported in study reports, such as ethnic background and patient cohort overlap.

5 Further analyses of the association between genetic variants of the *NAT2* data and hepatotoxicity

5.1 Introduction

For the meta-analyses described in Chapter 4, we used the following approach for our analyses:

- For genetic variants where all studies presented data for each genotype group separately, we performed two separate pairwise comparisons, as we had no reason to make an assumption about the underlying mode of inheritance.
- For genetic variants where all studies presented data for combined genotype groups, we performed one comparison of the combined genotype groups (i.e. we assumed a recessive or a dominant model, depending on which genotype groups were combined), as the evidence base suggested that assuming a particular mode of inheritance would be valid.
- For genetic variants where the approach varied between studies, we performed one comparison of the combined genotype groups, and also separate pairwise comparisons, which included studies that reported data for each genotype group separately. This approach allowed us to consider the robustness of our findings to the chosen analysis approach.

However, these analyses are subject to limitations. Specifically, if a specific mode of inheritance is assumed incorrectly, power to detect pharmacogenetic associations will be reduced. Furthermore, performing pairwise comparisons ignores any correlation between the effect estimates induced by the common baseline group; this is therefore an inefficient approach since the two effect estimates do not borrow strength from one another as they would in a bivariate analysis.

In 2005, Minelli et al.[61] introduced the genetic model-free method of meta-analysis for pharmacogenetic studies. This method does not require any assumptions about the underlying mode of inheritance, nor does it ignore the correlation between effect estimates induced by the common baseline group. This method is therefore potentially able to overcome the limitations of the standard methods of meta-analysis.

In this chapter, we introduce the genetic model-free method and describe how we used the method to analyse data from our systematic review of genetic variants and anti-TB drug-related toxicity. We then present the results of these analyses, and compare these results

to those from simple pairwise comparisons. We applied the genetic model-free method of meta-analysis for variants of the *NAT2* gene only; these data were sufficient to allow us to investigate the impact of meta-analysis approach on findings from a systematic review.

In addition to allowing comparisons between standard and more complex methods of meta-analysis, we also hoped that exploring alternative methods of meta-analysis would allow us to gain further insight into the process of synthesising pharmacogenetic evidence. In particular, we aimed to identify whether any barriers exist that might prevent researchers from performing meta-analyses of pharmacogenetic studies using the optimal statistical approach.

5.2 Genetic model-free and bivariate meta-analysis approaches

Minelli et al.[61] have developed a method of meta-analysis that does not specify any particular mode of inheritance in advance; the mode of inheritance is instead estimated from data from the included studies. This method is known as the genetic model-free approach. However, this method does require the assumption that the unknown genetic model is common to all studies included in the meta-analysis. If the assumption of a common mode of inheritance across all studies is violated, a bivariate approach to meta-analysis may be more suitable. The bivariate approach models the two effect sizes (i.e. heterozygotes *versus* homozygous wild-types, 'Gg' *versus* 'GG', and homozygous mutant-types *versus* homozygous wild-types, 'gg' *versus* 'GG') separately whilst still accounting for the correlation between them.[202] Before specifying the genetic model-free approach, it is useful to first specify the bivariate method of meta-analysis.

Bivariate approach specification

The following notation is used:

G: wild-type allele at SNP of interest

g: mutant-type allele at SNP of interest

μ_{ji}: true effect size for comparison j (j=1, comparison of Gg *versus* GG; j=2, comparison of gg *versus* GG) for study i

z_{ji}: effect size for comparison j (e.g. log OR) estimated in study i

Assuming approximate bivariate normality, the model can be written as:

$$1)\ \begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \end{bmatrix}, \begin{bmatrix} v_{1i} & v_{12i} \\ v_{12i} & v_{2i} \end{bmatrix} \right\},$$

where v_{ji} is the within-study variance for effect size μ_{ji} and v_{12i} is the within-study covariance between the two effect sizes for study i. The variances and covariances can be derived from the number of cases and controls in each genotype group in study i as follows:

$$v_{1i} = \frac{1}{\mathrm{GG}_{\mathrm{cases}}} + \frac{1}{\mathrm{Gg}_{\mathrm{cases}}} + \frac{1}{\mathrm{GG}_{\mathrm{controls}}} + \frac{1}{\mathrm{Gg}_{\mathrm{controls}}},$$

$$v_{2i} = \frac{1}{\mathrm{GG}_{\mathrm{cases}}} + \frac{1}{\mathrm{gg}_{\mathrm{cases}}} + \frac{1}{\mathrm{GG}_{\mathrm{controls}}} + \frac{1}{\mathrm{gg}_{\mathrm{controls}}},$$

$$v_{12i} = \frac{1}{\mathrm{GG}_{\mathrm{cases}}} + \frac{1}{\mathrm{GG}_{\mathrm{controls}}};$$

where $\mathrm{GG}_{\mathrm{cases}}$, $\mathrm{Gg}_{\mathrm{cases}}$, and $\mathrm{gg}_{\mathrm{cases}}$ are the number of cases in the homozygous wild-type, heterozygous, and homozygous mutant-type groups of study i, respectively, and $\mathrm{GG}_{\mathrm{controls}}$, $\mathrm{Gg}_{\mathrm{controls}}$, and $\mathrm{gg}_{\mathrm{controls}}$ are the number of controls in the homozygous wild-type, heterozygous, and homozygous mutant-type groups of study i, respectively.

If we assume that the studies come from a population in which the true effect sizes are also normally distributed, then:

$$2)\ \begin{bmatrix}\mu_{1i}\\ \mu_{2i}\end{bmatrix} \sim N\left\{\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\tau_1 & \tau_{12}\\ \tau_{12} & \tau_2\end{bmatrix}\right\},$$

where τ_j is the between-study variance for effect size j and τ_{12} is the between-study covariance between the two effect sizes. With this approach, the between-study variance is therefore allowed to differ between the two contrasts.

Combining 1) and 2) gives:

$$3)\ \begin{bmatrix}z_{1i}\\ z_{2i}\end{bmatrix} \sim N\left\{\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}v_{1i}+\tau_1 & v_{12i}+\tau_{12}\\ v_{12i}+\tau_{12} & v_{2i}+\tau_2\end{bmatrix}\right\},$$

from which a likelihood can be formed and maximised to estimate the parameters.

Genetic model-free approach specification

The genetic model-free approach is similar to the bivariate model. A parameter, λ, is introduced, which determines the relationship between μ_1 and μ_2. The relationship between the two effect sizes is given by:

$$\mu_1 = \lambda\mu_2$$

The value of λ is determined by the underlying mode of inheritance, i.e. a value of 0 represents a recessive mode of inheritance, a value of 0.5 represents an additive mode of inheritance, and a value of 1 represents a dominant mode of inheritance. So, the genetic model-free approach assumes that:

$$4)\ \begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \lambda\mu_{2i} \\ \mu_{2i} \end{bmatrix}, \begin{bmatrix} v_{1i} & v_{12i} \\ v_{12i} & v_{2i} \end{bmatrix} \right\}.$$

The variances and covariances can be derived from the number of cases and controls in each genotype group in study i as described previously for the bivariate model.

As for the bivariate model, we assume that the studies come from a population in which the true effect sizes are also normally distributed. Specifically, we assume that μ_{2i} is normally distributed with mean μ_2 and variance τ. Since the following mathematical properties are true:

i. the variance of a variable multiplied by a constant, λ, is equal to the variance of the raw variable multiplied by λ^2; and
ii. the covariance between a variable and that variable multiplied by a constant, λ, is equal to the variance of the variable, multiplied by λ,

we can derive the following:

$$5)\ \begin{bmatrix} \lambda\mu_{2i} \\ \mu_{2i} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \lambda\mu_2 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \lambda^2\tau & \lambda\tau \\ \lambda\tau & \tau \end{bmatrix} \right\}.$$

Here, it is assumed that the effect sizes share a common random component of variance, τ, due to the relationship between the effect sizes defined by λ ($\mu_1 = \lambda\mu_2$). Combining 4) and 5), we have:

$$6)\ \begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} \sim N\left\{ \begin{bmatrix} \lambda\mu_2 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} v_{1i} + \lambda^2\tau & v_{12i} + \lambda\tau \\ v_{12i} + \lambda\tau & v_{2i} + \tau \end{bmatrix} \right\},$$

and once again the likelihood can be formed and the parameters estimated.

It is important to note that it is possible to implement the genetic model-free method in a Bayesian framework.[203] The Bayesian approach may be preferable to a frequentist approach when there is external information relating to the magnitude of the effect sizes or

underlying mode of inheritance, such as expert opinion or data from studies not included in the meta-analysis.[61] In this chapter, we implemented all analyses in the frequentist framework, as no such external information was available. Furthermore, estimates of effect size obtained from the frequentist approach would be more directly comparable with those obtained in Chapter 4.

5.3 Methods

Model selection

In order to decide whether to use the genetic model-free or the bivariate approach for each genetic variant, we first needed to investigate whether the assumption of a common mode of inheritance across all studies included in the meta-analysis was valid. To do this, we plotted the log OR for Gg *versus* GG (z_{1i}) against the log OR for gg *versus* GG (z_{2i}) for each study. If the relationship between the two true effect sizes was of the form $\mu_1=\lambda\mu_2$, we would expect the plot to demonstrate a linear relationship between the two log ORs. This plot allowed us to assess the extent of heterogeneity in λ between studies, as shown by deviation from the linear relationship. We also plotted the value of λ and its bootstrapped 95% CI for each study. This allowed us to determine whether heterogeneity in λ between studies could be attributed to sampling error, or if the assumption of a constant λ across studies was violated.[61]

We generated and plotted bootstrapped 95% CIs for λ using the gplots package[204] in R 3.5.1.[205] Firstly, for each study, we generated 100,000 estimates of μ_1 and μ_2 ($\widehat{\mu_1}$ and $\widehat{\mu_2}$). We assumed that the number of events occurring in each genotype group of each study followed a binomial distribution, with the parameters 'p' (the observed proportion of patients experiencing an event in the genotype group) and 'n' (the total number of patients in the genotype group). For each simulation, we generated values for the number of events occurring in each genotype group from these binomial distributions, and calculated $\widehat{\mu_1}$ and $\widehat{\mu_2}$ using the following formulae:

$$\widehat{\mu_1} = ln\frac{(\mathrm{a}+0.5)(\mathrm{B}-\mathrm{b}+0.5)}{(\mathrm{A}-\mathrm{a}+0.5)(\mathrm{b}+0.5)}, \widehat{\mu_2} = \ln\frac{(\mathrm{a}+0.5)(\mathrm{C}-\mathrm{c}+0.5)}{(\mathrm{A}-\mathrm{a}+0.5)(\mathrm{c}+0.5)};$$

where a and A are the number of events and patients, respectively, in the homozygous wild-type group; b and B are the number of events and patients, respectively, in the heterozygous group; and c and C are the number of events and patients, respectively, in the homozygous group. A continuity correction of 0.5 was applied in these formulae to ensure that $\widehat{\mu_1}$ and $\widehat{\mu_2}$ were always defined.

For each of the 100,000 simulations, we calculated:

$$\hat{\lambda} = \frac{\widehat{\mu_1}}{\widehat{\mu_2}}$$

We then ordered these $\hat{\lambda}$ values, and used the 2,500th and 97,500th values to be the lower and upper limit of the CI, respectively. However, we encountered some difficulties when producing these bootstrapped CIs. If any of the simulated values of $\widehat{\mu_2}$ were equal to zero, this meant that $\hat{\lambda}$ would be undefined, and the bootstrapped CIs could not be calculated. To overcome this problem, we tried both 1) excluding zero values of $\widehat{\mu_2}$, and 2) replacing them with a small value (0.001). For all of our datasets, we observed that CIs were similar between the two approaches when few zeros were observed. As the number of zero values observed increased, we saw that CIs became increasingly wider when we replaced zero values with a small number, in comparison to when we excluded zero values. We therefore used the method of excluding the zeros, as this was the more conservative approach for investigating whether the assumption of constant λ across studies was valid.

For each genetic variant, if we decided that the assumption of a constant λ across studies was reasonable, we analysed the data using the genetic model-free approach. However, we also performed analyses using the bivariate approach, in order to investigate the robustness of our results, as the assessments of the appropriateness of the constant λ assumption were subjective. If we had found sufficient evidence to suggest that the constant λ assumption was violated, we would have performed the analysis using the bivariate approach only. For genetic variants where the validity of the constant λ assumption was unclear, we performed analyses using both the genetic model-free and bivariate approaches.

Fixed or random effects

We modelled μ_2 as having a random effect for all our analyses using the genetic model-free approach due to heterogeneity between studies in terms of study design, quality of methods, ethnic background of participants, and outcome definitions. Technically, it is possible to assume that the mode of inheritance varies slightly between studies, by assuming λ to have a random effect. However, it is usually not possible to model both μ_2 and λ as random-effects parameters, as it is difficult to disentangle the heterogeneity of λ from the heterogeneity of μ_2 without obtaining additional information .[61] We therefore treated λ as a fixed-effects parameter in our analyses.

For the bivariate approach, we modelled both μ_1 and μ_2 as having random effects, once again due to clinical and methodological heterogeneity between studies.

Heterosis

If the occurrence of heterosis (where the effect of the heterozygous genotype is greater than that of the homozygous mutant-type genotype) is biologically implausible, it is recommended that λ is restricted to lie within the interval [0-1].[61] We did not have sufficient information on the *NAT2* genetic variants to rule out the possibility of heterosis, so we did not specify limits within which λ had to lie.

Continuity correction

In the systematic review described in Chapter 3 and Chapter 4, there were sufficient data to perform meta-analyses for only one outcome, hepatotoxicity. We summarised hepatotoxicity data for each study according to the following contingency table:

Table 22 Contingency table to be completed for each study

Study ID	Genetic variant	Hepatotoxicity			No hepatotoxicity		
		No. of Hom WT patients	No. of Het patients	No. of Hom MT patients	No. of Hom WT patients	No. of Het patients	No. of Hom MT patients
Study X							

Het: heterozygous; Hom: homozygous; MT: mutant-type; WT: wild-type

If any of the cells in the contingency table were zero, we added a continuity correction of 0.5 to each cell in the table before using the genetic model-free or bivariate methods of meta-analysis.

Sensitivity analyses

Prior to undertaking a meta-analysis for a SNP, a test for HWE was undertaken for each study separately. We did not test HWE for *NAT2* acetylator status, as acetylator status is a trait defined by several *NAT2* SNPs. Where genotypes for a study were found to deviate from HWE (p<0.05), a sensitivity analysis was conducted excluding that study.

Software and code

We used Stata 14[192] to perform the genetic model-free and bivariate analyses; parameters were estimated in all analyses by maximum likelihood using the ml command. We used code provided in online supplementary materials to an article published by Bagos,[206] which reviews methods for multivariate meta-analysis of genetic association studies.

5.4 Results

Model selection

We examined the plot of z_{1i} *versus* z_{2i}, and the bootstrapped 95% CIs for λ for each study (Appendix 5) for each *NAT2* genetic variant. For *NAT2* acetylator status, *NAT2* 481C-T, *NAT2* 590G-A, and *NAT2* 857G-A, we concluded that the assumption of constant λ seemed reasonable. For *NAT2* 282C-T, *NAT2* 341T-C, *NAT2* 803A-G, there were few studies contributing data to each of these analyses, and considerable variability in the estimates of λ. We therefore concluded that the validity of the assumption of constant λ was unclear. Results from both the genetic model-free and bivariate approaches are provided for all genetic variants investigated.

Association between *NAT2* genetic variants and hepatotoxicity

The results of the genetic model-free analyses and bivariate analyses are presented in Table 23 to Table 29. We also present the results of simple pairwise comparisons for comparative purposes.

It should be noted that some results from simple pairwise comparisons presented in this chapter differ slightly to those presented in Chapter 4. These differences are due to a change in how we handled studies that did not identify any patients belonging to one of the relevant genotype groups. For the pairwise comparisons presented in Chapter 4, we excluded these studies from meta-analysis. For the pairwise comparisons presented in this chapter, we applied continuity correction (by adding 0.5 to each cell in the contingency table) so that the results were directly comparable with those obtained from the genetic model-free and bivariate approaches.

NAT2 *acetylator status*

Results from the genetic model-free and bivariate analyses were very similar to those from the simple pairwise comparisons. All analyses suggested that slow acetylators were significantly more likely to experience hepatotoxicity than rapid acetylators, and that there was no significant difference between intermediate and rapid acetylators (Table 23). The estimated value of λ from the genetic model-free analysis is close to zero (λ=0.12, 95% CI: -0.08 to 0.32), and therefore there is evidence to suggest that the mode of inheritance for *NAT2* acetylator status may be recessive.

Table 23 Association between NAT2 *acetylator status and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for intermediate *vs* rapid acetylators (95% CI)	I^2	OR for slow *vs* rapid acetylators (95% CI)	I^2
Genetic model-free	0.12 (-0.08 to 0.32)	1.17 (0.88 to 1.56)	0.0%	3.76 (2.32 to 6.09)	60.0%
Bivariate	N/A	1.17 (0.87 to 1.57)	0.0%	3.76 (2.32 to 6.08)	60.0%
Pairwise comparisons	N/A	1.12 (0.87 to 1.45)	0.0%	3.68 (2.23 to 6.09)	60.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

A graphical comparison of the results obtained in this chapter for the analysis of *NAT2* acetylator status and hepatotoxicity, with the analyses performed in Chapter 4 (assuming a dominant mode of inheritance, i.e. slow/intermediate *versus* rapid acetylators, and a recessive mode of inheritance, i.e. slow *versus* intermediate/rapid acetylators), is provided in Figure 21.

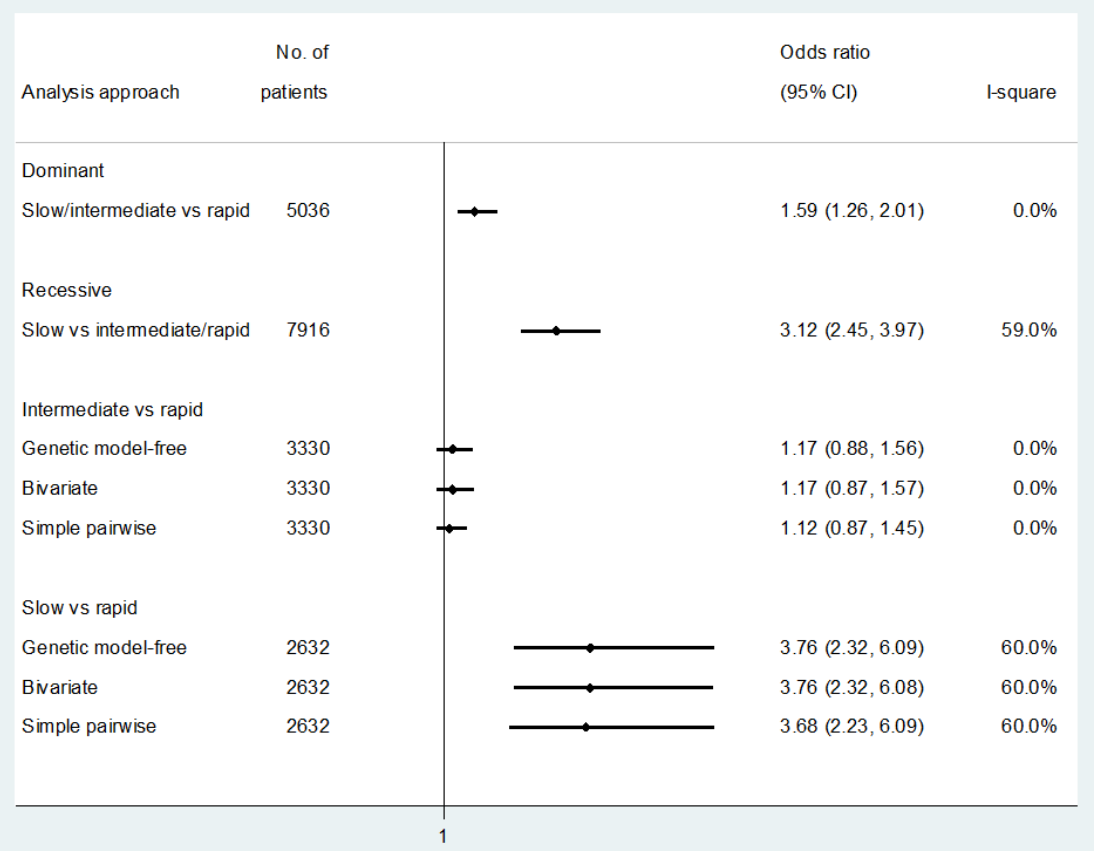

Figure 21 Summary of the results obtained by applying different approaches to the analysis of NAT2 *acetylator status and hepatotoxicity*

CI: confidence interval

NAT2 *282C-T*

Results from the genetic model-free analysis and bivariate analysis were very similar to those from the simple pairwise comparisons. All three analyses suggested that homozygous

mutant-type (TT) patients were significantly more likely to experience hepatotoxicity than homozygous wild-type (CC) patients, and that there was no significant difference between heterozygous (CT) patients and homozygous wild-type (CC) patients (Table 24). The estimated value of λ from the genetic model-free analysis is close to zero (λ=0.16, 95% CI: -0.15 to 0.47), and therefore there is evidence to suggest that the mutant-type allele, T, may be recessive.

Table 24 Association between NAT2 *282C-T and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for CT *vs* CC (95% CI)	I^2	OR for TT *vs* CC (95% CI)	I^2
Genetic model-free	0.16 (-0.15 to 0.47)	1.24 (0.79 to 1.95)	0.0%	3.83 (2.24 to 6.54)	5.5%
Bivariate	N/A	1.24 (0.79 to 1.95)	0.0%	3.83 (2.24 to 6.54)	5.5%
Pairwise comparisons	N/A	1.27 (0.80 to 2.02)	0.0%	3.95 (2.21 to 7.05)	5.5%

CI: confidence interval; N/A: not applicable; OR: odds ratio

We identified that genotype data from Santos 2013[155] deviated from HWE, and so performed a sensitivity analysis excluding data from this study. The exclusion of Santos 2013[155] had little impact on the results (Appendix 6).

NAT2 *341T-C*

Results from the genetic model-free analysis and bivariate analysis were very similar to those from the simple pairwise comparisons. All three analyses suggested that there were no significant differences between either heterozygous (TC) patients and homozygous wild-type (TT) patients, or homozygous mutant-type (CC) patients and homozygous wild-type (TT) patients (Table 25). The most noticeable difference was between the OR calculated for CC *versus* TT by the simple pairwise comparison (OR=1.51, 95% CI: 0.59 to 3.85) and the ORs calculated for CC *versus* TT by the genetic model-free approach (OR=1.96, 95% CI: 0.85 to 4.49) and the bivariate approach (OR=1.96, 95% CI: 0.85 to 4.48). The estimated value of λ from the genetic model-free analysis is close to zero (λ=0.19, 95% CI: -0.47 to 0.86), and therefore there is some evidence to suggest that the mutant-type allele, C, may be recessive. However, the 95% CI for λ is wide, and it is therefore difficult to draw firm conclusions about the true mode of inheritance of the *NAT2* 341T-C SNP.

Table 25 Association between NAT2 *341T-C and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for TC *vs* TT (95% CI)	I^2	OR for CC *vs* TT (95% CI)	I^2
Genetic model-free	0.19 (-0.47 to 0.86)	1.14 (0.72 to 1.79)	0.0%	1.96 (0.85 to 4.49)	0.0%
Bivariate	N/A	1.13 (0.72 to 1.78)	0.0%	1.96 (0.85 to 4.48)	0.0%
Pairwise comparisons	N/A	1.15 (0.72 to 1.82)	0.0%	1.51 (0.59 to 3.85)	0.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

We identified that genotype data from Lee 2010[141] deviated from HWE, and so performed a sensitivity analysis excluding data from this study. The exclusion of Lee 2010[141] had little impact on the results (Appendix 6).

NAT2 *481C-T*

The ORs for heterozygous (CT) patients *versus* homozygous wild-type (CC) patients calculated by the genetic model-free and bivariate analyses were almost identical to that calculated by the simple pairwise comparison. All analyses identified a significant difference between homozygous mutant-type (TT) patients and homozygous wild-type (CC) patients. The estimated value of λ from the genetic model-free analysis is close to 0.5 (λ=0.47, 95% CI: 0.08 to 0.86), and therefore there is some evidence to suggest that the mode of inheritance for this SNP may be additive. However, the 95% CI for λ is wide, and it is therefore difficult to draw firm conclusions about the true mode of inheritance of the *NAT2* 481C-T SNP.

Table 26 Association between NAT2 *481C-T and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for CT *vs* CC (95% CI)	I^2	OR for TT *vs* CC (95% CI)	I^2
Genetic model-free	0.47 (0.08 to 0.86)	1.49 (1.12 to 1.97)	0.0%	2.32 (1.36 to 3.94)	0.0%
Bivariate	N/A	1.48 (1.10 to 2.00)	0.0%	2.16 (1.14 to 4.11)	0.0%
Pairwise comparisons	N/A	1.48 (1.12 to 1.97)	0.0%	2.13 (1.23 to 3.67)	0.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

NAT2 *590G-A*

Results from the genetic model-free and bivariate analyses were very similar to those from the simple pairwise comparisons. All analyses suggested that heterozygous (AG) patients were significantly more likely to experience hepatotoxicity than homozygous wild-type (GG) patients, and that homozygous mutant-type (AA) patients were significantly more likely to experience hepatotoxicity than homozygous wild-type (GG) patients (Table 27). The

estimated value of λ from the genetic model-free analysis (λ=0.35, 95% CI: 0.09 to 0.62) is not indicative of a particular mode of inheritance for the *NAT2* 590G-A SNP.

Table 27 Association between NAT2 *590G-A and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for AG *vs* GG (95% CI)	I^2	OR for AA *vs* GG (95% CI)	I^2
Genetic model-free	0.35 (0.09 to 0.62)	1.30 (1.03 to 1.64)	0.0%	2.11 (1.31 to 3.39)	47.7%
Bivariate	N/A	1.33 (1.06 to 1.68)	0.0%	2.01 (1.21 to 3.34)	47.7%
Pairwise comparisons	N/A	1.30 (1.06 to 1.59)	0.0%	2.05 (1.24 to 3.40)	47.7%

CI: confidence interval; N/A: not applicable; OR: odds ratio

Excluding data from the Xiang 2014[170] and Santos 2013[155] studies, which we found to deviate from HWE, had little impact on the results of these analyses (Appendix 6).

NAT2 *803A-G*

Results from the genetic model-free analysis and bivariate analysis were similar to those from the simple pairwise comparisons. There were no statistically significant differences between either heterozygous (GA) patients and homozygous wild-type (AA) patients, or homozygous mutant-type (GG) patients and homozygous wild-type (AA) patients for any of the three analyses (Table 28). Once again, the most noticeable difference was between the OR calculated for GG *versus* AA by the simple pairwise comparison (OR=1.93, 95% CI: 0.71 to 5.21) and the ORs calculated for GG *versus* AA by the genetic model-free approach (OR=2.21, 95% CI: 0.86 to 5.70) and the bivariate approach (OR=2.21, 95% CI: 0.86 to 5.69). The estimated value of λ from the genetic model-free analysis is close to zero (λ=0.18, 95% CI: -0.47 to 0.84), and therefore there is some evidence to suggest that the mutant-type allele, G, may be recessive. However, the 95% CI for λ is wide, and it is therefore difficult to draw firm conclusions about the true mode of inheritance of the *NAT2* 803A-G SNP.

Table 28 Association between NAT2 803A-G *and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for GA *vs* AA (95% CI)	I^2	OR for GG *vs* AA (95% CI)	I^2
Genetic model-free	0.18 (-0.47 to 0.84)	1.16 (0.68 to 1.96)	0.0%	2.21 (0.86 to 5.70)	0.0%
Bivariate	N/A	1.15 (0.68 to 1.96)	0.0%	2.21 (0.86 to 5.69)	0.0%
Pairwise comparisons	N/A	1.14 (0.67 to 1.96)	0.0%	1.93 (0.71 to 5.21)	0.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

NAT2 *857G-A*

Results from all three analyses suggested that heterozygous (GA) patients were significantly more likely to experience hepatotoxicity than homozygous wild-type (GG) patients (Table 29).The results for the comparison of homozygous mutant-type (AA) and homozygous wild-type (GG) patients were also similar between the three analysis approaches, although the result from the genetic model-free analysis did not reach statistical significance.

The estimated value of λ from the genetic model-free analysis is close to 0.5 (λ=0.46, 95% CI: -0.16 to 1.08), and therefore there is some evidence to suggest that the mode of inheritance for this SNP may be additive. However, the 95% CI for λ is wide, and it is therefore difficult to draw firm conclusions about the true mode of inheritance of the *NAT2* 857G-A SNP. It is interesting to note here that the 95% CI includes 1, and therefore does not rule out the possibility of heterosis.

Table 29 Association between NAT2 *857G-A and hepatotoxicity: results of the genetic model-free analysis, bivariate analysis and simple pairwise comparisons*

Analysis approach	λ (95% CI)	OR for GA *vs* GG (95% CI)	I^2	OR for AA *vs* GG (95% CI)	I^2
Genetic model-free	0.46 (-0.16 to 1.08)	1.31 (1.04 to 1.65)	0.9%	1.80 (0.95 to 3.41)	1.0%
Bivariate	N/A	1.29 (1.01 to 1.65)	0.9%	1.96 (1.01 to 3.83)	1.0%
Pairwise comparisons	N/A	1.30 (1.03 to 1.64)	0.9%	1.97 (1.06 to 3.64)	1.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

5.5 Discussion

In this chapter, we applied the genetic model-free method of meta-analysis to data for the association between genetic variants of the NAT2 gene and hepatotoxicity, obtained as part of the systematic review and meta-analysis described in Chapter 3 and Chapter 4. The analyses presented in this chapter ought to be considered as exploratory, as no adjustments have been made to account for multiple analysis approaches. Conclusions regarding the association between *NAT2* genetic variants and ATDH ought to be based on results from the primary and secondary analyses presented in Chapter 4.

Impact of meta-analysis approach

Generally, results from the genetic model-free analyses and bivariate analyses were very similar to those obtained from the pairwise comparisons. There was no obvious pattern in terms of how the size or precision of the effect estimate was impacted by changing the analysis approach.

The largest impact in terms of effect size was seen for the analysis of the *NAT2* 341T-C SNP; the genetic model-free approach and bivariate approach both estimated an increase in odds of 96% for homozygous mutant-type (CC) individuals in comparison to homozygous wild-type (TT) individuals, while the simple pairwise analysis estimated an increase in odds of 51% for the same comparison. However, neither effect was shown to be statistically significant. When comparing results presented in this chapter, it is important to note that there were differences in how heterogeneity was estimated between the multivariate (bivariate and genetic-model free) and univariate (pairwise comparisons) approaches. Specifically, for the genetic model-free and the bivariate analyses, a maximum likelihood approach was implemented to estimate heterogeneity, in accordance with the methods implemented in the paper that outlines the genetic-model free approach.[61] Alternatively, the pairwise comparisons made use of the DerSimonian and Laird 'moment-based approach' to estimate heterogeneity. It is possible that differences between the results of the multivariate methods and the univariate method for the *NAT2* 341T-C SNP may at least in part be attributed to differences between the methods used to estimate heterogeneity.

Langan et al.[207] conducted empirical research to explore the impact that choice of heterogeneity estimator may have on observed results from meta-analyses. The study identified that differences between heterogeneity estimators rarely had a major impact on the observed effect estimate. However, this study did not specifically consider the maximum likelihood approach, as this method has been reported to be inferior to the restricted maximum likelihood approach,[208] which was included in the study. It is possible that differences between the maximum likelihood and DerSimonian and Laird 'moment-based' approaches may be more pronounced.

Riley et al.[209] demonstrated that the extent to which results of bivariate random-effects meta-analyses differ from results of univariate random-effects meta-analyses is likely to vary according to the amount of missing data in the data set. This is due to the fact that when data are missing for one of the comparisons of interest (say, j=1), the differences between the within-study variances for this comparison of interest (the v_{1i}s) are likely to be large; this allows more scope for the bivariate meta-analysis to 'borrow strength' from data for the related comparison (j=2). On examination of the *NAT2* SNP data sets, we did not identify an obvious relationship between the amount of missing data and the extent to which results differed between the univariate and bivariate meta-analyses. We therefore concur with Riley and colleagues, who concluded that further work is required to

investigate how, and under what conditions, the results of bivariate and univariate random-effects meta-analyses differ for a variety of missing data scenarios.

For the analysis of the *NAT2* 857G-A SNP; the simple pairwise analysis and bivariate analysis demonstrated a statistically significant effect for the comparison of homozygous mutant-type (AA) individuals versus homozygous wild-type (GG) individuals, while the result from the genetic model-free analysis did not reach statistical significance. Here, it seems unlikely that differences in the method used to estimate heterogeneity could explain the difference between the observed results, as results from the bivariate and pairwise analyses (which implement different methods of estimating heterogeneity) were very similar. It is more likely that differences between the observed results were due to the underlying assumptions of the genetic model-free approach, i.e. that the two effect sizes are related according to λ, and consequently share a common random component of variance.

Based on our observations that differences between the genetic model-free, bivariate and univariate approaches may have a considerable impact on the magnitude of the observed effect size, and may also alter the statistical significance of an effect size, we consider it to be plausible that for some systematic reviews, the choice of meta-analysis method may impact the overall conclusions.

Limitations

When applying the genetic model-free method, the first step is to investigate whether the assumption of constant λ is reasonable. For the *NAT2* 282C-T, *NAT2* 341T-C, and *NAT2* 803A-G SNPs, it was unclear whether the assumption of constant λ was valid. We consider that a limitation of the genetic model-free approach is that for genetic variants where data are sparse, it may often be difficult to make firm conclusions about the suitability of the genetic model-free approach. Even when data were not sparse, and we judged the assumption of constant λ to be reasonable (for *NAT2* acetylator status, *NAT2* 481C-T, *NAT2* 590G-A, and *NAT2* 857G-A), these judgements were subjective. It was therefore useful to examine the impact that using the bivariate approach had on the results for all the investigated genetic variants. Results were mostly similar between the bivariate and genetic model-free approaches; this provided reassurance that the results obtained were robust to changes in the assumption of constant λ.

A further limitation of the analyses presented in this chapter is that we were only able to include data from studies that reported outcome data for each genotype group separately. For *NAT2* acetylator status, 13 studies reported data for the rapid and intermediate

acetylator groups combined, and one study reported data for the slow and intermediate acetylator groups combined. It was not possible to include data on acetylator status from any of these 14 studies in the analyses presented in this chapter, and consequently information has been lost.

Salanti and Higgins[73] have developed methodology which allows the synthesis of data from studies, regardless of whether data has been reported for combined or separate genotype groups (as previously mentioned in Section 2.5). As the focus of this chapter was to investigate the impact of applying the genetic model-free approach, we have not applied the Salanti and Higgins[73] method to our data. As the loss of data in some of the analyses conducted as part of our systematic review was considerable due to merged genotype groups, an exploration of the Salanti and Higgins[73] method may have provided valuable insight into the robustness of our findings.

As previously outlined, it is possible that differences between the methods used to estimate heterogeneity between the three analysis approaches may at least partially explain some of the differences in observed results. Future exploration of how results varied under the different analysis approaches when assuming a common method of estimating heterogeneity would provide additional insight into how much impact the different underlying assumptions of each analysis approach has on observed meta-analysis results.

5.6 Conclusion

We found that results were often very similar between the genetic model-free, bivariate and simple pairwise methods of meta-analysis. However, we also observed that results occasionally varied to the extent that conclusions of a review may change depending on the analysis approach chosen. It is therefore important that review authors carefully consider the available methods of meta-analysis, and pre-specify their analysis strategy in a detailed protocol. It may also be informative to investigate the impact of meta-analysis approach in sensitivity analyses.

Improvements in the reporting of pharmacogenetic studies would help systematic review authors who wish to use the genetic model-free approach in their meta-analyses. If study authors routinely reported outcomes for each genotype group separately, this would give systematic reviewers greater freedom in terms of their analysis approach, and would reduce the likelihood of relevant data being excluded from meta-analyses.

6 Development of the STROPS (STrengthening the Reporting Of Pharmacogenetic Studies) guideline

6.1 Introduction

While conducting the work outlined in Chapter 3, Chapter 4 and Chapter 5 of this thesis, we hoped to identify challenges that systematic reviewers encounter when synthesising evidence from pharmacogenetic studies. A key theme that we observed was that improvements in the reporting of pharmacogenetic studies would enable systematic reviewers to conduct better-quality and more strongly powered meta-analyses. For example, lack of reporting of key patient characteristics, such as ethnicity, can severely hinder investigations of heterogeneity, which form a key part of any systematic review and/or meta-analysis. Additionally, if authors of a pharmacogenetic study do not report the number of participants in each genotype group and outcomes for each genotype group separately, systematic review authors are limited in terms of the analysis approaches available to them, and the study may even be excluded from meta-analyses.

Cobos et al.[210] also identified that there is room for improvement in the reporting of pharmacogenetic studies in their review of the methodological quality of pharmacogenetic studies with binary treatment response outcomes. In particular, study design characterisation was lacking in 43 (66%) of 65 included studies, while the majority (97%) of included studies did not report the sample size calculation or planned sample size. Furthermore, the issue of multiple testing was applicable to 59 included studies, although only 11 (19%) reported methods for addressing multiplicity. Finally, tests for HWE were reported fully (with population tested, test method, and *p*-value all provided) in 4 of 42 papers (10%) that analysed allele frequencies. The authors of this review concluded that specific guidance on the reporting of pharmacogenetic studies would be valuable.

Consequently, we decided to shift the focus of our work to the reporting of pharmacogenetic studies. Although reporting guidelines are available for observational studies[11] and genetic association studies,[199] there is currently no guideline for the reporting of pharmacogenetic studies that has been developed using widely accepted and robust methodology. Pharmacogenetic studies have different characteristics to other types of observational and indeed, genetic association studies. Although some items from existing guidelines can be applied to pharmacogenetic studies, there are many additional pharmacogenetic-specific characteristics that could be reported; clear guidance on which items are essential to report is needed.

We therefore decided to develop a reporting guideline for pharmacogenetic studies (and accompanying explanation and elaboration [E+E] document) using methodology proposed by EQUATOR (Enhancing the QUAlity and Transparency Of health Research).[211] In this chapter, we describe the methods we used to develop the reporting guideline, and present the final version, referred to as the STROPS (STrengthening the Reporting of Pharmacogenetic Studies) guideline. The STROPS guideline is applicable to studies designed to identify pharmacogenetic associations, i.e. the study designs described in Table 1 of Chapter 1. The STROPS guideline will set a robust standard of reporting for pharmacogenetic studies, and consequently will facilitate the conduct of high-quality systematic reviews and meta-analyses.

6.2 Methods

The protocol outlining the pre-specified methods of this project has been published.[212] The steering committee for the STROPS project included the following six individuals: Marty Chaplin (researcher into meta-analysis of pharmacogenetic studies), Jamie Kirkham (researcher into consensus methodology and developer of reporting guidelines), Kerry Dwan (researcher into systematic review methodology), Derek Sloan (clinical infectious diseases researcher), Gerry Davies (clinical pharmacogenetic researcher in infectious diseases) and Andrea Jorgensen (researcher into statistical methods for pharmacogenetics, including evidence synthesis methods).

In accordance with the EQUATOR methodology for guideline development,[211] we (the steering committee) developed the STROPS guideline according to the following stages:

1. Establish a preliminary checklist of reporting items to be considered for inclusion in the reporting guideline for pharmacogenetic studies (Stage 1).
2. Conduct a Delphi survey to gain consensus opinion on reporting items to be considered within a reporting guideline for pharmacogenetic studies (Stage 2).
3. Hold a consensus meeting to consider the results of the Delphi survey and to finalise the list of items for the reporting guideline (Stage 3).
4. Develop and publish the STROPS guideline and a detailed E+E document (Stage 4).

Preliminary checklist of reporting items

To establish a preliminary checklist of reporting items, we firstly included items from existing relevant guidelines. We considered all guidelines listed on the EQUATOR website[213] under the clinical area of genetics. Two authors (MC and ALJ) assessed guidelines to be

relevant if they were applicable to pharmacogenetics studies. Two authors (MC and ALJ) discussed whether items from these guidelines would ensure transparency of reporting of pharmacogenetic studies, and consequently decided whether to include each item in the preliminary checklist. For example, the GRIPS statement[214] includes some items that can be applied to pharmacogenetic studies; however, we did not include all items from this guideline as many items are only relevant to studies where a genetic risk prediction model is being developed, and these studies are outside the remit of our guideline. We also considered items from the STROBE guideline,[11] which is not listed on the EQUATOR website under the clinical area of genetics, but was judged to be applicable to pharmacogenetic studies by members of the steering committee. MC and ALJ modified some items from existing guidelines; the majority of these wording modifications were intended to make items more relevant to pharmacogenetic studies.

Secondly, we supplemented this list with additional items thought to be important. These items were either suggested by steering committee members based on our own experience in pharmacogenetic research, or were drafted by MC and ALJ to cover issues identified by Jorgensen and Williamson[18] that relate specifically to the conduct of pharmacogenetic research. Finally, we drafted help text for each item, to ensure that the language used was comprehensible to all Delphi participants. All steering committee members approved this preliminary checklist (including the wording of each item) before the Delphi survey began.

Delphi survey

Participants

In March-April 2019, we invited three groups of stakeholders to participate in the Delphi survey. Stakeholder groups were chosen to encompass all aspects of pharmacogenetic research.

1. Those who undertake primary pharmacogenetic research

We asked co-ordinators of ten national and international pharmacogenetics networks to forward the survey on to network members. We performed searches to ensure that all major networks across the globe were identified.

2. Those who systematically review pharmacogenetic research data

We identified 89 contact authors of systematic reviews of pharmacogenetics studies identified by searching PubMed, using appropriate search terms such as ‘pharmacogenetics’, ‘pharmacogenomics’, ‘systematic review’ and ‘meta-analysis’. An

information specialist helped us to design the search strategy. We used a snowball technique, asking contact authors to complete the survey and to forward the survey on to their co-authors.

3. Those who publish pharmacogenetic research

We contacted 210 editors-in-chief of 168 journals that may publish pharmacogenetic studies. We used a snowball technique, asking editors-in-chief to participate in the survey, and also to forward the survey on to editors at their journal. We performed searches using Google to identify journals using search terms 'pharmacogenetics', 'pharmacogenomics', 'precision medicine', 'personalised/personalized medicine' and 'journal'. We also considered journals listed on the 'SCImago Journal & Country Rank' website[215] under the category 'Genetics'.

Design

The Delphi process consisted of two rounds of electronic-based survey, response and feedback. The first round survey (Round 1, March-May 2019) included scoring of reporting guideline items from the preliminary list formed at Stage 1 and invited additional items not included in this list. A second round survey (Round 2, May-July 2019) was then undertaken providing feedback from Round 1 and inviting participants to re-score the items. Any additional reporting items identified by participants in Round 1 (and approved by the steering committee) were included for scoring by participants in Round 2 of the Delphi process.

The Delphi survey was conducted using DelphiManager,[216] a web-based system designed by the COMET Initiative[189] to facilitate the building and management of Delphi surveys.

Recruitment process and ethical considerations

We e-mailed the individuals listed above with information about the STROPS project and the Delphi process and an invitation to complete Round 1 of the Delphi survey within three weeks. We informed invitees that participation in the survey was optional, and that we would assume informed consent if an invitee responded to Round 1 of the survey. We informed invitees that all data would be anonymised; we allocated a unique identification number to each participant in the Delphi survey.

We sent a reminder e-mail at the end of the second week to prompt completion of the survey. All participants who completed Round 1 of the Delphi survey were invited to

participate in Round 2. However, we informed invitees that completion of Round 1 did not necessitate completion of Round 2.

The University of Liverpool ethics committee was consulted and confirmed ethical approval for this study (Reference: 3586).

Participant characteristics

We asked participants to provide their name, e-mail address and consent to be acknowledged as a participant in the Delphi survey in publications arising from this project.

Delphi scoring

Participants were asked to score each of the reporting guideline items listed using a scale of 1 to 9, with 1 to 3 labelled 'not important for inclusion in the guideline', 4 to 6 labelled 'important but not critical for inclusion in the guideline' and 7 to 9 labelled 'critical for inclusion in the guideline'.[217] Participants were also given the option to score a reporting guideline item as 'unable to score' if they were unable to offer an opinion on the importance of the item.

Delphi Round 1

Reporting guideline items were presented in the order in which they would be addressed in the pharmacogenetic study report and were grouped under relevant headings (i.e. title and abstract, introduction, methods, results, discussion and other information). Participants were asked to score each item as described previously. Participants were given the opportunity to provide free text feedback on each item; this feature ensured participants were able to comment on the wording of items or provide justification for their score. Participants were also given the chance to suggest items that they believed should be included in the reporting guideline.

Round 1 analysis

For each item, the number of participants who scored the item and the distribution of scores were summarised. Participants who scored an item as 'unable to score' were excluded from the analysis for that particular item. The steering committee reviewed all additional reporting items suggested by participants. If items were not already covered by the existing list, we added these items to the list of reporting items presented in Round 2, or we covered the item as part of the E+E text for existing items. We also reviewed free text participant feedback provided for each item, and amended the wording of items or the accompanying E+E text where appropriate. The wordings of additional or amended items

were constructed by MC and ALJ, and were approved by members of the steering committee.

Delphi Round 2

In Round 2, each participant was shown the number of respondents and distribution of scores for each item from Round 1, for each stakeholder group separately. Participants were also reminded how they personally scored each item in Round 1. Participants were asked to consider the responses from other Delphi participants, and to re-score the items. Additional items identified as part of Round 1 were scored by participants in Round 2.

Round 2 analysis

For each item, the number of respondents and the distribution of scores was summarised. Participants who scored an item as 'unable to score' were excluded from the analysis for that particular item.

If participants that did not respond to Round 2 have different opinions to participants from the same stakeholder group who completed both rounds, then attrition bias has occurred and the results of the Delphi survey may be affected. We investigated the risk of attrition bias by calculating average Round 1 scores for each participant, and then producing boxplots to show the distributions of these average scores according to whether participants completed Round 2 or not, for each stakeholder group. We visually examined these plots to assess the likelihood of attrition bias.

Consensus definition

We defined that each stakeholder group had reached consensus for an item if at least 70% of members of that group scored the item as 'critical for inclusion in the guideline'.

Consensus meeting

The steering committee and representatives from the stakeholder groups met to consider the results of the Delphi survey and to finalise the list of items for the reporting guideline. We aimed to include one or two representatives (with at least one being non-UK based) from each stakeholder group in the consensus meeting. We invited individuals to the consensus meeting using the following broad principles: i) Delphi participants who completed both rounds; (ii) a balance across the three stakeholder groups; (iii) a reasonable geographic spread. If an individual could not attend, they were replaced by someone else from the same stakeholder group. The meeting was conducted via the web conferencing software, Zoom.[218]

Prior to and during the meeting, participants were shown a summary of how each stakeholder group scored each reporting item at Round 2 of the Delphi survey, and the number of stakeholder groups who achieved consensus. Meeting attendees discussed each reporting item in turn, and made a decision on whether to include the item in the reporting guideline or not. Where necessary, attendees voted on whether an item ought to be included in the guideline using TurningPoint polling software;[219] the item was retained if at least 70% of attendees voted for its inclusion. Items were considered in the order they were presented in the Delphi survey.

Post-consensus meeting development

We drafted the reporting guideline and E+E document concurrently. The purpose of the E+E document is to provide the rationale for and meaning of each reporting item alongside examples of good reporting practice. We also provided the origin of each reporting item in the E+E document.

6.3 Results

Preliminary checklist of reporting items

We considered 12 guidelines listed on the EQUATOR website[213] under the clinical area of genetics; we assessed two[199, 214] to be applicable to pharmacogenetics studies. The preliminary checklist consisted of 92 reporting items (Appendix 7, Table 60). The items are labelled 1 to 85, as some items have subitems, for example, 52a, 52b, 52c. The checklist includes: 38 items from the STROBE statement,[11] 22 items from the STREGA statement,[199] 20 items drafted by MC and ALJ to cover issues identified by Jorgensen and Williamson,[18] 10 items suggested by members of the steering committee, and two items from the GRIPS statement.[214] A total of 22 items from existing guidelines were modified; the majority of these modifications were intended to make items more relevant to pharmacogenetic studies.

Delphi survey

In Round 1 of the Delphi survey, participants were asked to score 92 reporting items (the preliminary checklist of reporting items). A total of 71 individuals completed Round 1: 15 journal editors, 41 primary researchers and 15 systematic reviewers. A total of 10 participants suggested 31 additional items at this stage. In addition, during Round 1 of the survey, Delphi invitees made us aware of two publications[220, 221] which contained potentially relevant reporting items. After reviewing the additional reporting items suggested by participants and items from the two relevant publications, we included seven additional

items in the list of reporting items presented in Round 2 (Appendix 7, Table 2). We also covered some suggested reporting items by including additional detail in the E+E text for existing items. No items were amended based on participants' free text comments, although alterations were made to the E+E text accompanying some items to address the given feedback.

A total of 52 individuals scored 99 reporting items in Round 2 of the Delphi survey: 10 journal editors, 31 primary researchers, and 11 systematic reviewers. The anonymised data from both rounds of the Delphi survey are available in the supplementary materials to the publication of the STROPS guideline.[222]

As we asked network co-ordinators, systematic reviewers and journal editors to contact individuals on our behalf, it is impossible to determine a response rate to Round 1 of the Delphi survey. However, we considered the response received to Round 2 to be reasonable (overall: 52/71, 73%; journal editors: 10/15, 67%; systematic reviewers: 31/41, 76%; primary researchers: 11/15, 73%). Considering the boxplots presented in Figure 22 to Figure 25, the distributions of Round 1 scores were similar between those who completed both rounds of the Delphi survey, and those who completed Round 1 only. There was therefore no evidence to suggest that attrition bias occurred.

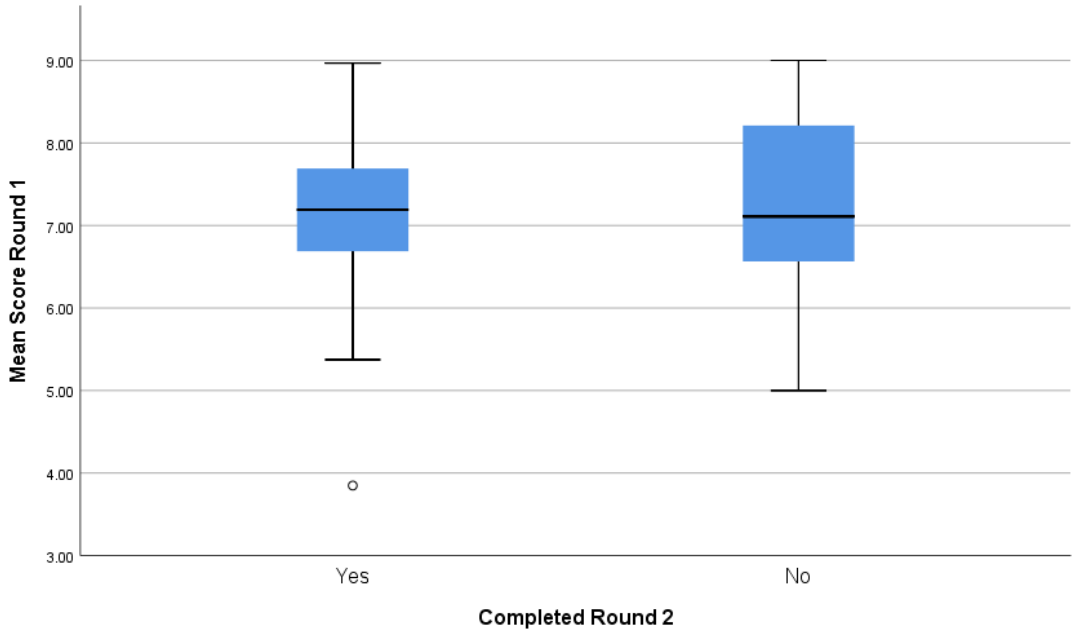

Figure 22 Attrition bias: All stakeholder groups

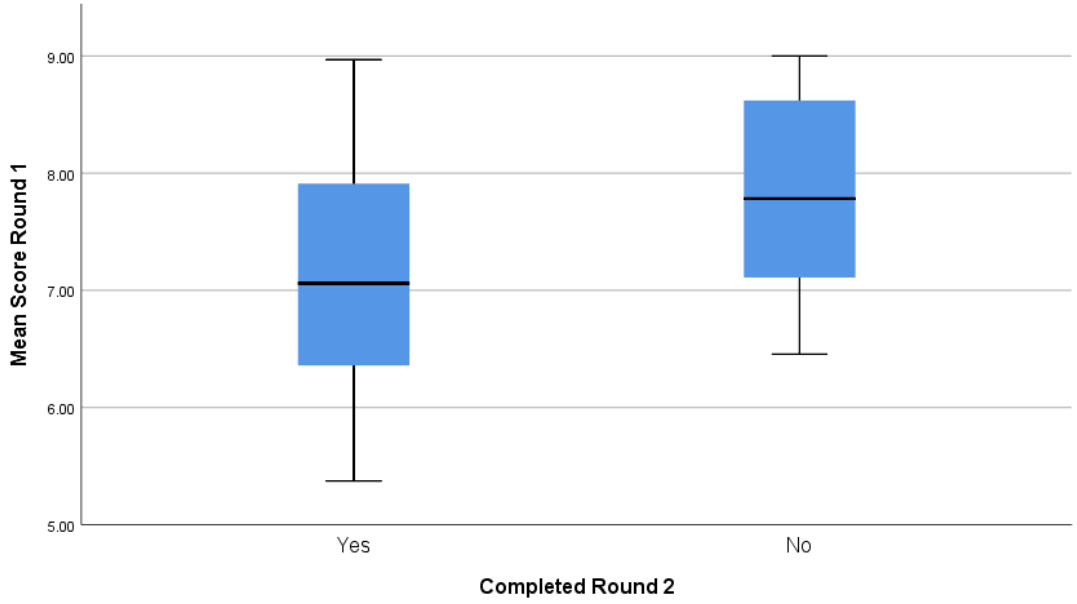

Figure 23 Attrition bias: Journal editors

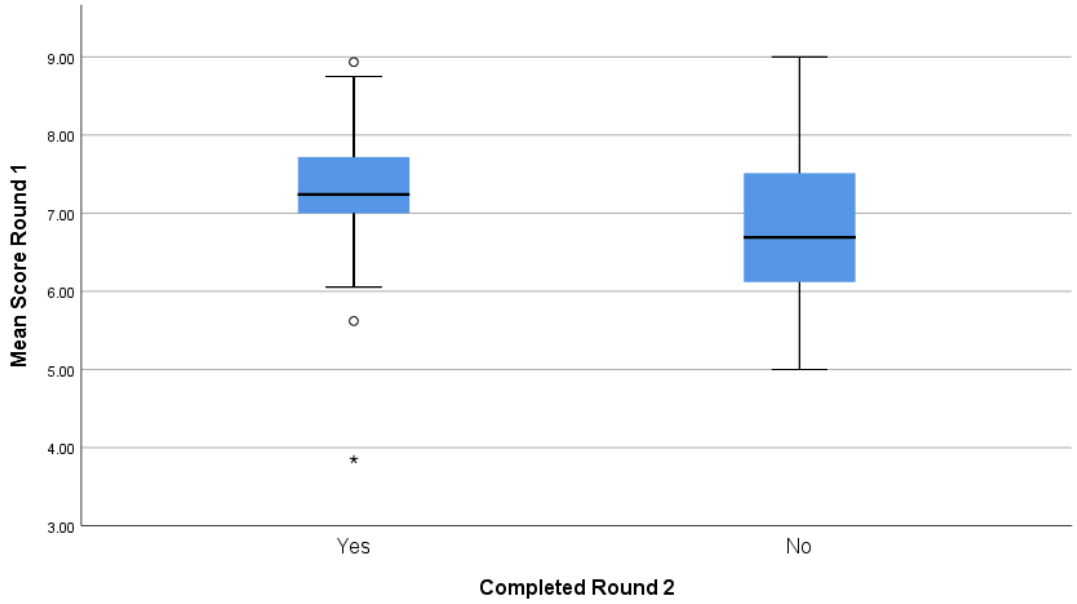

Figure 24 Attrition bias: Primary researchers

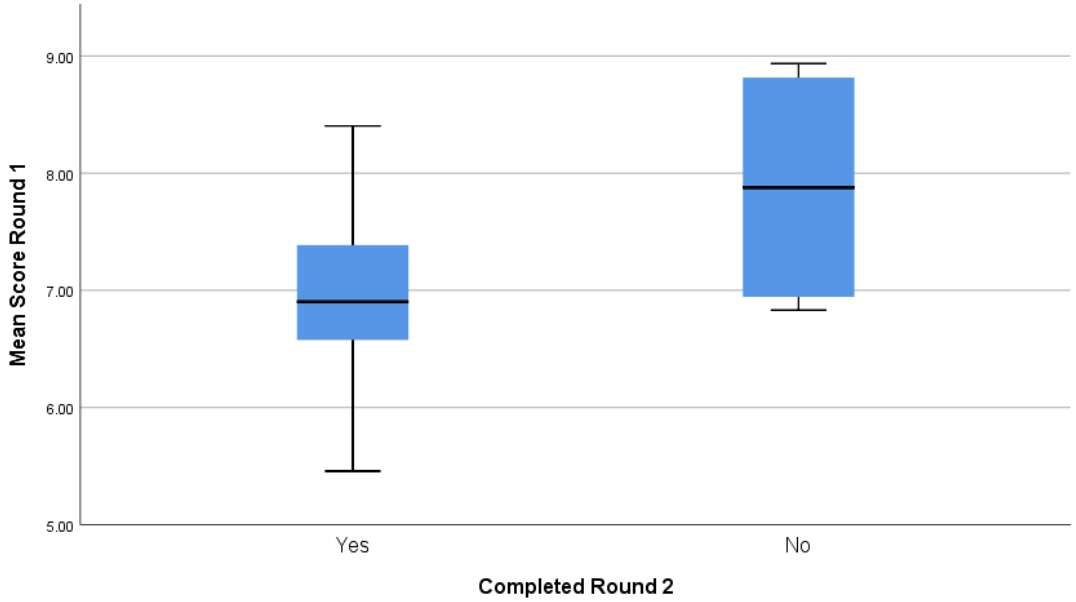

Figure 25 Attrition bias: Systematic reviewers

Consensus meeting

The consensus meeting took place in November 2019, including six steering committee members, and four representatives of stakeholder groups (one journal editor, based in Germany; one primary researcher, based in Switzerland; two systematic reviewers, based in the UK and Spain). Two steering committee members did not participate in voting (JK chaired and KD took notes), so there were eight voting individuals in attendance.

Prior to the meeting, we provided consensus meeting attendees with a consensus matrix (Appendix 8), which documents the percentage of individuals within each stakeholder group scoring each item as 'critical' for inclusion in the guideline, at Round 1 and at Round 2 of the Delphi Survey. A copy of the consensus meeting slides can be found in the supplementary materials to the publication of the STROPS guideline.[222]

A summary of the decisions made at the consensus meeting (alongside the number of stakeholder groups reaching consensus in the Delphi survey for each item) are provided in Table 30. We decided whether to include or exclude items, and whether to combine multiple items under a single item. Where a vote was taken, this is indicated in the table. Otherwise, no vote was undertaken and the decisions made were based solely on consideration of the Delphi results and discussion.

Table 30 Summary of decisions made at consensus meeting

Category	#	Criteria	Consensus at Delphi	Decision at consensus meeting
Title and abstract				
Title and abstract	1	Indicate the study's pharmacogenetic design in the title and the abstract.	1 group	Exclude after vote
	2	Provide in the abstract an informative and balanced summary of what was done and what was found.	3 groups	Include - add to E+E text that study design should be reported in abstract
Introduction				
Background/ rationale	3	Explain the scientific background and rationale for the investigation being reported.	3 groups	Include
	4	Provide reasons for choosing the genes and SNPs genotyped.	3 groups	Include
	5	If reasons for (4) include previous association studies, provide key details from these studies (effect size and standard error/confidence interval).	1 group	Exclude after vote
Objectives	6	State specific objectives, including any pre-specified hypotheses.	3 groups	Include
	7	State if the study is the first report of a pharmacogenetic association, a replication effort, or both.	2 groups	Include after vote
Methods				
Study design	8	Present key elements of study design early in the paper.	3 groups	Include
Setting	9	Describe the setting, locations and relevant dates, including periods of recruitment, follow-up, and data collection.	0 groups	Include after vote
Participants	10	Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.	3 groups	Combine #10, #11 and #12
	11	Case-control study – Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. State whether true controls or population controls were used. Give the rationale for the choice of cases and controls.	3 groups	Combine #10, #11 and #12
	12	Cross-sectional study – Give the eligibility criteria, and the sources and methods of selection of participants.	3 groups	Combine #10, #11 and #12
	13	Report the drug and regime participants were exposed to, and the length of exposure.	3 groups	Include
	14	Cohort study – For matched studies, give matching criteria and number in each genotype group.	3 groups	Combine #14, #15
	15	Case-control study – For matched studies, give matching criteria and the number of controls per case.	3 groups	Combine #14, #15

Category	#	Criteria	Consensus at Delphi	Decision at consensus meeting
	16	Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	3 groups	Include
	17	If other publications report results for the same patient cohort, or a subset of the patient cohort, provide information on this patient cohort overlap and references to the relevant publications.	2 groups	Include after vote
	18	Report disease/clinical indication of patients using a standardised ontology.	1 group	Include and add 'where possible' to E+E text (standardised ontology may not always be available); vote taken on whether to include the item with the change to E+E text
	19	Confirm whether patients were blinded to their genotyping result.	1 group	Exclude after vote
Variables	20	Provide justification for choice of outcomes.	3 groups	Include; reverse order of #20 and #21
	21	Clearly define all outcomes, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	3 groups	Include; reverse order of #20 and #21
	22	Clearly define genetic exposures (genetic variants) using a widely used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin).	3 groups	Include
	23	Report the rs number of each genotyped SNP.	3 groups	Include
	24	Report whether the outcomes measured (including definitions) are in line with core/preferred outcome sets for the particular topic of interest.	1 group	Exclude
	25	Clearly state how haplotypes or star alleles were defined.	3 groups	Include
	26	Clearly state on which chromosomal strand the alleles are reported.	1 group	Exclude
	27	If referring to the minor, wild-type or mutant allele of a variant, state which allele this is and for which given population/cohort.	2 groups	Include after vote
	28	If studying drug metabolites, provide references and links to structures and database identifiers.	0 groups	Exclude
Data sources /measurement	29	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.	1 group	Include after vote

Category	#	Criteria	Consensus at Delphi	Decision at consensus meeting
	30	Describe laboratory methods, including source and storage of DNA, genotyping methods and *platforms* (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.	3 groups	Include
	31	If study is case-control, confirm whether patients were genotyped in mixed batches.	0 groups	Cover by adding to E+E text for #30
	32	Confirm whether genotyping personnel were blinded to outcome status.	0 groups	Exclude
	33	Describe the primers used.	0 groups	Exclude
	34	Describe genotype quality control methods.	1 group	Include after vote
	35	Describe findings of genotype quality control methods.	0 groups	Combine with #34
Bias	36	Describe any efforts to address potential sources of bias.	1 group	Exclude after vote
	37	For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.	3 groups	Include
	38	Report how adherence to treatment was assessed, and report the results of the assessment.	3 groups	Include
Study size	39	Explain how the study size was arrived at, or provide details of the *a priori* power to detect effect sizes of varying degrees.	3 groups	Include
Quantitative variables	40	Explain how quantitative variables (confounders and effect modifiers) were handled in the analyses. If applicable, describe which groupings were chosen, and why.	3 groups	Include
	41	If applicable, describe how effects of treatment on quantitative outcome variables were dealt with.	2 groups	Exclude
Statistical methods	42	Describe all statistical methods, including those used to control for confounding.	3 groups	Include
	43	State software version used and options (or settings) chosen.	0 groups	Exclude
	44	Describe any methods used to examine subgroups and interactions.	1 group	Include after vote as subitem to #42
	45	Explain how missing data were addressed.	3 groups	Include as subitem to #42
	46	Report any methods used to assess the assumption of missingness at random and the finding of such assessments.	1 group	Cover by adding to E+E text of item #45
	47	Cohort study – If applicable, explain how loss to follow-up was addressed.	1 group	Include after vote as subitem to #42
	48	Case-control study – If applicable, explain how matching of cases and controls was addressed.	2 groups	Include after vote as subitem to #42

Category	#	Criteria	Consensus at Delphi	Decision at consensus meeting
	49	Cross-sectional study – If applicable, describe analytical methods taking account of sampling strategy.	1 group	Include after vote as subitem to #42
	50	Describe any sensitivity analyses.	0 groups	Include after vote as subitem to #42
	51	State whether Hardy-Weinberg equilibrium (HWE) was considered and, if so, how.	2 groups	Combine #51, #52 (vote taken on whether to include or exclude this combined item)
	52	Where HWE test is undertaken, quote the *p*-value threshold applied to determine deviation from HWE.	2 groups	Combine #51, #52 (vote taken on whether to include or exclude this combined item)
	53	Describe any methods used for inferring genotypes or haplotypes.	2 groups	Include
	54	Describe any methods used to assess or address population stratification.	1 group	Include after vote
	55	Describe any methods used to assess and correct for relatedness among subjects. Report results of assessments for relatedness.	1 group	Include after vote; after the vote meeting participants decided to cover #55 by adding to the E+E text for #34
	56	Describe any assumptions made regarding mode of inheritance.	1 group	Exclude after vote
	57	Provide justification for assumption of mode of inheritance or if no mode is assumed.	1 group	Exclude
	58a	Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating multiple genetic variants.	3 groups	Include
	58b	Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating multiple outcomes.	3 groups	Include
	58c	Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating multiple assumptions regarding mode of inheritance.	2 groups	Include
	59	Describe any methods used to adjust for extent of adherence in the analyses.	2 groups	Include
Results				
Participants	60a	Report the numbers of individuals at each stage of the study – e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.	3 groups	Include
	60b	Give reasons for non-participation at each stage.	0 groups	Cover by adding to E+E text of #60a
	60c	Consider use of a flow diagram.	0 groups	Cover by adding to E+E text of #60a

Category	#	Criteria	Consensus at Delphi	Decision at consensus meeting
	61	For each genetic variant, report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful.	1 group	Exclude after vote
SNPs	62	Report any SNPs that were excluded from analysis, and provide reasons for these exclusions.	2 groups	Include after vote
Descriptive data	63	Give characteristics of study participants (e.g., demographic, clinical, social) and information on potential confounders.	3 groups	Include
	64	Indicate the number of participants with missing data for each variable of interest.	3 groups	Include by adding to E+E text for #63
	65	For a cohort study, consider giving information listed in (63) and (64) by genotype.	1 group	Exclude after vote
	66	For a case-control study, give the information listed in (63) and (64) for cases and controls separately.	0 groups	Combine with #63 (add to E+E text)
	67	Report reasons for missing genotype data.	0 groups	Exclude
	68	Cohort study – Summarize follow-up time, e.g. average and total amount.	2 groups	Include after vote
	69	Where HWE tests have been undertaken, highlight SNPs that deviate from HWE.	3 groups	Include
	70	Where population stratification is assessed, report the results.	3 groups	Include
Outcome data	71a	For a cohort study, report all outcomes (phenotypes) investigated for each genotype category over time.	3 groups	Include
	71b	For a case-control study, report numbers in each genotype category for all outcomes investigated.	3 groups	Include
	71c	For a cross sectional study, report all outcomes (phenotypes) investigated for each genotype category.	3 groups	Include
	72	If a study includes more than one ethnic group, provide the summary data specified in (71) per ethnic group.	3 groups	Include
Main results	73	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.	3 groups	Include
	74	Report category boundaries when continuous variables were categorised.	3 groups	Include
	75	If relevant, consider translating effect estimates to number needed to test to illustrate potential clinical utility of any significant findings.	0 groups	Exclude after vote
	76	Report results of any adjustments for multiple comparisons.	3 groups	Cover by adding to E+E text of #73
	77	Report precise *p*-values for all associations.	3 groups	Cover by adding to E+E text of #73

Category	#	Criteria	Consensus at Delphi	Decision at consensus meeting
Other analyses	78	Report other analyses done – e.g., analyses of subgroups and interactions, and sensitivity analyses.	2 groups	Include
	79	If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken.	2 groups	Include
	80	If detailed results are available elsewhere, state how they can be accessed.	3 groups	Exclude after vote
Discussion				
Key results	81	Summarize key results with reference to study objectives.	3 groups	Include
Limitations	82	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	3 groups	Include
	83	Report on the risk of phenoconversion (genotype-phenotype mismatch) and its magnitude in the study population.	1 group	Exclude after vote
Interpretation	84	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	3 groups	Include
	85	Report genotype frequencies from other studies.	0 groups	Exclude
Generalisability	86	Discuss the generalisability (external validity) of the study results.	3 groups	Include
	87	Discuss, if pertinent, the health care relevance of the study results.	1 group	Exclude after vote
Other information				
Study registration/ protocol	88	State whether the protocol for the analysed data is publicly available and if so, how the protocol can be accessed.	0 groups	Exclude after vote
	89	State whether the study has been registered. If the study has been registered, provide details of the registry.	2 groups	Include after vote
Ethical approval	90a	Report whether ethical approval was obtained for the collection of genetic data.	3 groups	Include
	90b	If ethical approval was obtained, report the committee that gave ethical approval and a reference ID.	0 groups	Cover by adding to E+E text of #90a
Funding	91	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	3 groups	Include
Databases	92	State whether databases for the analysed data are or will become publicly available and if so, how they can be accessed.	1 group	Include after vote

E+E: explanation and elaboration document; HWE: Hardy-Weinberg equilibrium; SNP: single nucleotide polymorphism

Post-consensus meeting development

Following the consensus meeting, MC drafted the reporting guideline with guidance from other members of the steering committee. The following minor amendments were made to the reporting guideline:

- We excluded item 14 and item 49; while searching for examples for these items, we found very few pharmacogenetic studies that used a matched cohort design, or a cross-sectional design with a complex sampling strategy; these items would therefore be irrelevant to the vast majority of guideline users
- We removed 'Identify variables likely to be associated with population stratification (confounding by ethnic origin)' from item 22, as this is covered by item 54
- We added more terms ('major', 'reference', 'risk' and 'effect') that might be used to describe alleles to item 27
- Although we decided to cover item 55 by adding to the E+E text for item 34 at the consensus meeting, the steering committee subsequently agreed that relatedness of participants is a separate issue to genotype quality control. We decided to keep item 55 as a standalone item in the guideline
- We introduced a new subitem to item 42 to cover confounding, and made item 42 a generic introduction to the statistical methods subitems
- We modified item 68 to indicate that average and/or total follow-up time is sufficient
- Although we voted to exclude item 80 from the 'Other analyses' section of the reporting guideline at the consensus meeting, the intention was to consider this item under the 'Databases' section. However, time constraints meant that we did not discuss this item again. The steering committee subsequently agreed that this item relates to additional results, rather than individual patient data in databases. We decided to keep the item in its original position, and add 'i.e. in supplementary materials' so that the meaning of the item is clear

The resulting draft guideline was circulated to all members of the steering committee and consensus meeting attendees. All comments and revisions were taken into consideration and the checklist revised accordingly.

The STROPS guideline

In Table 31, we present the STROPS guideline. The accompanying E+E document is provided in Appendix 9.

Table 31 STROPS guideline

Category	#	Criteria
Abstract		
Abstract	1	Provide in the abstract an informative and balanced summary of what was done and what was found.
Introduction		
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported.
	3	Provide reasons for choosing the genes and SNPs genotyped.
Objectives	4	State specific objectives, including any pre-specified hypotheses.
	5	State if the study is the first report of a pharmacogenetic association, a replication effort, or both.
Methods		
Study design	6	Present key elements of study design early in the paper.
Setting	7	Describe the setting, locations and relevant dates, including periods of recruitment, follow-up, and data collection.
Participants	8	Give the eligibility criteria, and the sources and methods of selection of participants. For a cohort study, describe methods of follow-up. For a case-control study, state whether true controls or population controls were used. Give the rationale for the choice of cases and controls.
	9	Report the drug and regime participants were exposed to, and the length of exposure.
	10	For a matched case-control study, give matching criteria and the number of controls per case.
	11	Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.
	12	If other publications report results for the same patient cohort, or a subset of the patient cohort, provide information on this patient cohort overlap and references to the relevant publications.
	13	Report disease/clinical indication of patients using a standardised ontology when possible.
Variables	14	Clearly define all outcomes, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.
	15	Provide justification for choice of outcomes.
	16	Clearly define genetic exposures (genetic variants) using a widely used nomenclature system.
	17	Report the rs number of each genotyped SNP.
	18	Clearly state how haplotypes or star alleles were defined.
	19	If referring to the minor, major, wild-type, mutant, reference, risk or effect allele of a variant, state which allele this is and for which given population/cohort.
Data sources/ measurement	20	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.
	21	Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.
	22	Describe genotype quality control methods and findings.
	23	For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.

Category	#	Criteria
	24	Report how adherence to treatment was assessed, and report the results of the assessment.
Study size	25	Explain how the study size was arrived at, or provide details of the *a priori* power to detect effect sizes of varying degrees.
Quantitative variables	26	Explain how quantitative variables (confounders and effect modifiers) were handled in the analyses. If applicable, describe which groupings were chosen, and why.
Statistical methods	27	Address the following:
	a	Describe methods used to control for confounding.
	b	Describe any methods used to examine subgroups and interactions.
	c	Explain how missing data were addressed.
	d	Cohort study – If applicable, explain how loss to follow-up was addressed.
	e	Case-control study – If applicable, explain how matching of cases and controls was addressed.
	f	Describe any sensitivity analyses.
	28	State whether Hardy-Weinberg equilibrium (HWE) was considered and, if so, how.
	29	Describe any methods used for inferring genotypes or haplotypes.
	30	Describe any methods used to assess or address population stratification.
	31	Describe any methods used to assess and correct for relatedness among subjects. Report results of assessments for relatedness.
	32	Describe any methods used to address multiple comparisons or to control risk of false positive results due to a) multiple genetic variants b) multiple outcomes c) multiple assumptions regarding mode of inheritance.
	33	Describe any methods used to adjust for extent of adherence in the analyses.
Results		
Participants	34	Report the numbers of individuals at each stage of the study – e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.
SNPs	35	Report any SNPs that were excluded from analysis, and provide reasons for these exclusions.
Descriptive data	36	Give characteristics of study participants (e.g., demographic, clinical, social, ethnicity) and information on potential confounders.
	37	Cohort study – Summarize follow-up time, e.g. average and/or total amount.
	38	Where HWE tests have been undertaken, highlight SNPs that deviate from HWE.
	39	Where population stratification is assessed, report the results.
Outcome data	40a	For a cohort study, report all outcomes (phenotypes) investigated for each genotype category over time.
	40b	For a case-control study, report numbers in each genotype category for all outcomes investigated.
	40c	For a cross sectional study, report all outcomes (phenotypes) investigated for each genotype category.
	41	If a study includes more than one ethnic group, provide the summary data specified in (40) per ethnic group.
Main results	42	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.
	43	Report category boundaries when continuous variables were categorised.
Other analyses	44	Report other analyses done – e.g., analyses of subgroups and interactions, and sensitivity analyses.
	45	If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken.

Category	#	Criteria
	46	If detailed results are available elsewhere, i.e. in supplementary materials, state how they can be accessed.
Discussion		
Key results	47	Summarize key results with reference to study objectives.
Limitations	48	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.
Interpretation	49	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.
Generalisability	50	Discuss the generalisability (external validity) of the study results.
Other information		
Study registration	51	State whether the study has been registered. If the study has been registered, provide details of the registry.
Ethical approval	52	Report whether ethical approval was obtained for the collection of genetic data.
Funding	53	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.
Databases	54	State whether databases for the analysed data are or will become publicly available and if so, how they can be accessed.

HWE: Hardy-Weinberg equilibrium; SNP: single nucleotide polymorphism

6.4 Discussion

We conducted our project using robust methodology for developing reporting guidelines proposed by the EQUATOR network,[211] in order to ensure that the resulting reporting guideline would be useful and widely disseminated. The EQUATOR approach includes a face-to-face consensus meeting, which follows the Delphi survey. This meeting usually involves the steering group and a selection of stakeholders who took part in the Delphi survey. Due to a lack of funding for this project to cover travel and accommodation costs, we were unable to arrange a face-to-face consensus meeting. Our meeting was conducted via conference call, and only involved steering committee members and one or two representatives of each stakeholder group to ensure easy communication between attendees. The majority of meeting attendees were therefore UK-based. However, we invited a large, international and multidisciplinary cohort to participate in the Delphi survey, and meeting attendees were able to base their decisions on the opinions of this wider cohort.

We prioritised items for inclusion in the guideline if all stakeholder groups reached consensus; i.e. at least 70% of participants in each stakeholder group scored the item as 'critical'. We chose this threshold as consensus that an item ought to be included in the guideline requires agreement by the majority regarding critical importance of the outcome. This threshold has also been used previously in the development of the COS-STAR reporting guideline.[223] Although the choice of this threshold is somewhat subjective, pre-specification

of the threshold in our protocol[212] ought to provide assurance that we did not define consensus in a post-hoc way, and therefore our own opinions did not bias the results of the Delphi survey.[224]

The final phase of activities described by Moher et al.[211] in their guidance for guideline development relates to post-publication activities. To ensure that the guideline is widely disseminated and implemented, we circulated the published STROPS guideline to individuals who completed both Delphi rounds, and asked co-ordinators of pharmacogenetic networks to notify their members of the publication of the guideline. We also registered the guideline on the EQUATOR website.[213] We also plan to present the guideline at conferences relevant to pharmacogenetic research, and seek guideline endorsement from relevant journals. In the long term, there are further actions that could be taken to maximise the impact and usefulness of the guideline for many years to come, and these will be discussed further in Section 7.4.

It is important to note that the STROPS guideline is applicable only to studies designed to identify whether associations exist between genetic variants and treatment-related outcomes (i.e. study designs described in Table 1 of Chapter 1). Studies may be conducted to address other pharmacogenetic-related questions, as discussed in Section 1.5. For example, an RCT may be conducted to determine whether implementing genotype-guided prescribing of a particular drug leads to fewer toxicity events than standard prescribing practice. To the best of our knowledge, there are no reporting guidelines that specifically relate to these alternative types of pharmacogenetic study. Future research may be useful to determine whether specific reporting guidance for these types of studies would be beneficial to the field of pharmacogenetic research.

6.5 Conclusion

With the increasing number of meta-analyses of pharmacogenetic studies that are being undertaken, it is important that reporting of key data in study reports is improved in order to allow robust synthesis of the studies. There is currently no guideline for the reporting of pharmacogenetic studies that has been developed using a widely accepted and rigorous methodology. The STROPS guideline will first and foremost set a robust standard of reporting for pharmacogenetic studies, helping research authors to improve the completeness and transparency of their study reports. In turn, improvements in the reporting of pharmacogenetic studies will enable researchers to conduct high-quality

systematic reviews and meta-analyses, and thus improve the strength of the evidence base for pharmacogenetic associations.

7 Conclusions and further work

7.1 Introduction

The ultimate goal of conducting pharmacogenetic research is to identify associations between genetic variants and drug response, and knowledge of these associations can subsequently be used to inform clinical practice when making drug prescribing decisions, by way of genetic tests. However, before pharmacogenetic tests are introduced into clinical practice the evidence on their clinical utility needs to be strong, with the gold-standard approach to demonstrating clinical utility of genotype-guided in comparison to standard prescribing practices being the RCT. For researchers to consider spending valuable time and resources conducting RCTs of genotype-guided prescribing, the evidence base for the pharmacogenetic association of interest must be robust.

Meta-analysis can be used to improve sample size and therefore increase the power to detect pharmacogenetic associations, whilst also allowing researchers to investigate the possibility that associations observed in individual studies may be spurious. It can also be used to explore heterogeneity in observed effects between studies. The aim of this work was to identify and resolve challenges that might prevent researchers from performing robust synthesis of pharmacogenetic evidence by way of a systematic review and meta-analysis.

7.2 Summary of main findings

Methodology for meta-analysis of genetic association studies is well-established, and detailed guidance is available for reviewers in the HuGENet HuGE Review Handbook.[24] The existing methodology and guidance can be applied when synthesising pharmacogenetic evidence, along with Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[18]

We applied the existing methodology and guidance to our systematic review and meta-analysis of associations between genetic variants and anti-TB drug-related toxicity. Our review identified that *NAT2* slow/intermediate acetylators are significantly more likely to experience hepatotoxicity than *NAT2* rapid acetylators. We also observed possible associations between the *CYP2E1 Rsa*I (rs2031920) and *GSTM1* null polymorphisms and hepatotoxicity. To the best of our knowledge, no systematic reviews or meta-analyses have been conducted to investigate the association between anti-TB drug-related toxicity and genetic variants other than variants of the *NAT2*, *CYP2E1*, *GSTM1* and *GSTT1* genes. Our

review summarises the entire body of literature on genetic variants and anti-TB drug-related toxicity outcomes, and therefore adds to the existing understanding of the factors that may influence an individual's response to treatment with anti-TB drugs.

We observed that although the evidence base for this topic is extensive, with 69 articles contributing data to our review, synthesising data from the included pharmacogenetic studies was very challenging for a variety of reasons. In particular, reporting of key details, such as aspects of methodological quality, patient cohort overlap, and the ethnicity of included patients, was generally poor across the included studies.

A key feature of our systematic review was a thorough quality assessment using the checklist developed by Jorgensen and Williamson.[18] The quality of included studies was variable, with several areas of concern. However, we also observed that it was often difficult to assess whether a study satisfied quality assessment criteria due to the fact that relevant information was not sufficiently reported in the study publication. We concluded that it is possible that improvements in the reporting of methodological aspects of pharmacogenetic studies would improve our confidence in the findings of these studies, and consequently, the pooled results from meta-analyses.

We often found it difficult to determine whether patient cohorts from different articles were distinct, overlapping or identical. If we failed to identify multiple articles reporting data for overlapping or identical patient cohorts, some patients may have been double-counted in the meta-analyses; this is a form of unit-of-analysis error. Alternatively, if we incorrectly assumed that multiple articles reported data for overlapping or identical patient cohorts, our decision to exclude all but one of these articles from meta-analyses would have led to an unnecessary loss of data.

Further, only 24% of studies included in our systematic review reported ethnicity of the included patients; we were therefore unable to stratify our meta-analyses by ethnicity, as recommended in the HuGENet HuGE Review Handbook.[24] Consequently, we were unable to explore heterogeneity due to differences in ethnicity, and it is possible that our results were confounded by population stratification.

To synthesise data from the studies included in our systematic review, we initially applied standard methods of meta-analysis, which are subject to important limitations. We therefore decided it was important to also explore more complex methods of meta-analysis, namely the genetic model-free and bivariate analysis approaches, which are

potentially able to overcome the limitations of the standard methods of meta-analysis. We applied these methods of meta-analysis for variants of the *NAT2* gene only; these data were sufficient to allow us to investigate the impact of meta-analysis approach on results.

For the majority of SNPs, results from the genetic model-free and bivariate methods were very similar to those obtained from simple pairwise comparisons. However, for two SNPs, we observed considerable differences between results obtained from these different methods, either in terms of magnitude or statistical significance of the observed effect sizes. Based on these observations, we concluded that it is plausible that for some systematic reviews, the choice of meta-analysis method may impact the overall conclusions. We also identified that improvements in the reporting of pharmacogenetic studies, i.e. reporting of outcomes for each genotype group separately, would enable systematic reviewers greater freedom in terms of their analysis approach, and ensure that relevant data are not excluded from meta-analyses unnecessarily.

Due to the observations made while conducting the systematic review and meta-analyses (using both simple and more complex analysis methods), we identified the need to develop a reporting guideline for pharmacogenetic studies. We hoped that such a guideline would improve the transparency of reporting of pharmacogenetic studies and also facilitate future evidence synthesis. We used rigorous methodology to develop the guideline, including a two round Delphi survey and consensus meeting, both of which involved representatives from three key stakeholder groups.

A total of 52 individuals completed both rounds of the Delphi survey, scoring potential reporting items with regards to their importance for inclusion in the final reporting guideline. At the consensus meeting, the six members of the steering committee and four representatives of the key stakeholder groups discussed each reporting item in turn and decided whether to include each item in the final guideline. Subsequently, the 54 item STROPS guideline was finalised and published.[222] Initial dissemination activities included circulating the guideline to Delphi participants, asking co-ordinators of pharmacogenetic networks to notify network members of the guideline, and registering the guideline on the EQUATOR website.[213]

The final guideline includes items that address each of the specific reporting issues identified while conducting our meta-analyses of pharmacogenetic data; it is reassuring to note that Delphi survey participants and consensus meeting attendees shared our opinions on the importance of these items. Adherence to the guideline by pharmacogenetic authors

will therefore ensure that future systematic reviewers do not face the same challenges that we encountered when synthesising pharmacogenetic evidence.

7.3 Implications for practice and research

The work presented in this thesis has both clinical and methodological implications. The findings of the systematic review and meta-analysis presented in Chapter 3 and Chapter 4 have implications for the clinical management of TB, while our exploration of complex methods of analysis for variants of the *NAT2* gene and ATDH has methodological implications. Finally, the development of the STROPS reporting guideline has implications for members of the pharmacogenetic research community.

Clinical implications

Isoniazid is the anti-TB drug for which mechanisms of the genetic contribution to ATDH have been most widely studied. This drug is also part of the four drug combination regimen recommended by the WHO for treatment of drug-susceptible TB; isoniazid use is therefore routine and widespread. Indeed, of the studies included in our systematic review of genetic influence on anti-TB drug-related toxicity, only two[90, 173] reported using a regimen that did not contain isoniazid. Isoniazid is particularly important in the treatment of TB, as it is the mainstay of chemoprophylaxis in latent TB infection (LTBI). LTBI chemoprophylaxis is being rapidly expanded in recent strategies to eliminate TB as a public health problem, and it is therefore likely that global use of isoniazid will greatly increase over the coming years.

Our systematic review provides evidence to support a relationship between genetic variants of the *NAT2*, *CYP2E1* and *GSTM1* genes and ATDH. Perhaps the most clinically meaningful finding came from the comparison of slow acetylators *versus* intermediate/rapid acetylators, for which we calculated an OR of 3.12 (95% CI: 2.45 to 3.97). As the interpretation of ORs is not intuitive, and is also dependent on the underlying baseline risk of hepatotoxicity, it is useful to consider how the risk of hepatotoxicity would vary according to acetylator status in a hypothetical patient population. Specifically, we performed a random-effects meta-analysis of proportions in order to obtain a pooled estimate of baseline ATDH risk using data from studies included in the original meta-analysis. The estimate of baseline ATDH risk among intermediate/rapid acetylators was 13%; this value has been calculated solely to inform this discussion of clinical implications and is not referred to previously in this thesis. We then applied the OR of 3.12 obtained in our original meta-analysis of slow *versus* intermediate/rapid acetylators to the baseline risk of 13% to calculate the corresponding relative risk. Assuming a baseline risk of 13%, the risk

of hepatotoxicity would be approximately 2.5 times higher in slow acetylators than in rapid/intermediate acetylators. Clearly, this figure represents a clinically meaningful difference in hepatotoxicity risk between different acetylator groups.

In current clinical practice, transaminase testing is a readily available biomarker of possible ATDH; however, baseline values have modest predictive value and routine monitoring is not generally recommended. The findings of our systematic review and meta-analysis suggest that pharmacogenetic testing could make a clinically useful contribution to risk stratification for ATDH, particularly in populations where slow acetylator status is common. However, the need for testing of a relatively large panel of SNPs and the current lack of a clear substitute to isoniazid for LTBI chemoprophylaxis mean that such a strategy may not be cost effective or feasible. Studies investigating the cost-effectiveness and/or feasibility of such a strategy would be beneficial. High-quality studies that do not have the methodological limitations identified in Chapter 3 would also strengthen the evidence base.

A particularly concerning finding from our systematic review was that very few studies were conducted in Africa, where populations are most likely to be affected by TB. The lack of pharmacogenetic evidence on anti-TB drug related toxicity from African countries is, unfortunately, not surprising, considering that in general, the majority of pharmacogenetic studies have been performed in Asian and Caucasian populations.[225] A lack of funding may explain the limited nature of pharmacogenetic research in Africa; local researchers often have to approach foreign funding bodies to support their studies.[226] Due to the high level of disease burden across the African continent,[227] and the impact that adverse events and poor drug efficacy can have on already overstrained health care systems, African populations are likely to benefit from stratified medicine strategies. It is important that steps are taken to reduce the possibility that the pharmacogenetic research community may contribute to already existing health disparities between Africa and the rest of the world. It is reassuring that in September 2018, the African Pharmacogenomics Consortium was formally launched, which aims to address challenges of conducting and applying pharmacogenetics research in Africa.[228]

Despite the limitations of the identified evidence base, considering the approximately 2.5-fold increased risk of ATDH in slow acetylators estimated in our systematic review, pharmacogenetic epidemiology should certainly be a factor in national policymaking on the need for transaminase monitoring during treatment of active TB and LTBI locally.

Methodological implications: Further analyses of *NAT2* genetic variants and hepatotoxicity

Based on the findings of Chapter 5 of this thesis, systematic reviewers of pharmacogenetic studies ought to be aware that choice of meta-analysis method may have an important impact on observed estimates of association, and subsequently on overall conclusions. It is therefore important for the method to be used for meta-analysis to be pre-specified in a protocol for the systematic review, so that readers are assured that results from meta-analyses have not been selectively reported. It may also be informative to perform sensitivity analyses using alternative approaches to meta-analysis, as a way of investigating the robustness of results from the primary analysis.

Ideally, primary researchers should consider the fact that results from their study may be pooled with other similar studies, and endeavour to make future evidence synthesis as easy as possible. In our experience, if all studies had reported outcome data for each genotype group separately, this would have given us greater freedom in terms of our analysis approach, and ensured that all relevant data could be included in each of our meta-analyses. We are hopeful that item 40 of the STROPS guideline will ensure that future studies report data for each genotype group.

Methodological implications: STROPS guideline

The primary intention of the STROPS guideline is to help researchers transparently and completely report their pharmacogenetic studies. However, the STROPS guideline could also be used by peer reviewers and journal editors to strengthen manuscript review. As we continue with our dissemination and implementation activities (which will be discussed further in Section 7.4), we hope that the vast majority of pharmacogenetic researchers will be made aware of the STROPS guideline, and will choose to use the checklist in their future work. We recommend that pharmacogenetic journals endorse the STROPS guideline in order to optimise use of the guideline by study authors, peer reviewers and editors.

In addition, funders of pharmacogenetic research may consider introducing the STROPS guideline into their process for submission of applications.[211] Encouraging grant recipients to adhere to the STROPS guideline when reporting pharmacogenetic research will ensure that findings are clear, detailed, and consequently have a greater impact, whether that is by means of stimulating further research, or by impacting clinical practice and/or policy. Furthermore, it has previously been hypothesised that the availability of reporting guidelines may have a positive impact on how researchers design and conduct their research.[229] We would concur that familiarity with the STROPS guideline at early stages of

study design (such as when seeking funding) ought to ensure that researchers consider key methodological issues relating to pharmacogenetic studies, and plan appropriate strategies for handling these issues.

However, it is important to note that the STROPS guideline is not a quality assessment tool and should not be used as such. Reporting each of the items in the checklist does not ensure that the study is methodologically sound. For example, authors could provide full details about the extent of missing data and how this was handled in the analyses, satisfying item 27c of the STROPS guideline, however the approach outlined may be inappropriate. If missing data were extensive and likely not to be missing at random, simply analysing the available data would introduce bias to the study. The STROPS guideline will ensure that sufficient information is available from study reports for readers to make judgements about methodological quality, but the actual judgements should be made using a tool designed specifically for quality assessment, such as Jorgensen and Williamson's quality assessment tool.[18]

7.4 Future work

The work undertaken in this thesis has lead to the production of the STROPS reporting guideline. Initial dissemination activities included the notification of Delphi survey participants and co-ordinators of pharmacogenetic networks of the publication of the guideline, and the registration of the guideline on the EQUATOR website.[213] In the immediate future, we intend to present the guideline at key pharmacogenetic conferences and to contact editors of pharmacogenetic journals and co-ordinators of pharmacogenetic networks to discuss formal endorsement. To ensure that the STROPS guideline continues to be a useful resource over many years to come, further actions can be taken in the long-term.[211]

The development of a website for this project would have several benefits. Firstly, a website provides a single location where all resources relating to the project can be held, including: the checklist itself (in both PDF and DOC file formats), the E+E document, the list of Delphi participants and consensus meeting attendees, and relevant methodological publications. Furthermore, it is possible that researchers may wish to translate the guideline into other languages; we would welcome such efforts as availability of the guideline in more languages would undoubtedly improve the global reach of our guideline. If the guideline were to be translated, we would seek to be as actively involved as possible, to ensure that translations would be completed using robust methods (including back

translation). Translated versions of the guideline could also be made available on the website. Finally, a website could be a platform for stakeholders to submit feedback and criticism of the published guideline. Engaging with users of the guideline is particularly important if the guideline is to be updated in the future.

With regards to updating the guideline, it will be necessary to monitor the field closely for new methodological publications, and to regularly consider whether the time is right for an update of the guideline. Due to the time and effort involved in updating and implementing an existing guideline, we would aim to make infrequent, but major, updates to the guideline, rather than regularly making minor changes.

Another area for future work would be the evaluation of impact of the reporting guideline; if poor impact is observed by ourselves or another set of researchers, we would consider strategies that could improve the impact of our guideline. In particular, we would consider encouraging journals to not only endorse the guideline, but also to maximise adherence by asking authors of pharmacogenetic studies to submit completed copies of the guideline checklist along with their manuscript, and/or asking peer reviewers to refer to the checklist when reviewing pharmacogenetic studies.

In Chapter 2, we discussed how the focus of this work is aggregate data meta-analysis, whereby aggregate data available from study reports or communication with study authors are synthesised. An alternative approach is individual participant data (IPD) meta-analysis, which involves the collection of original research data, including detailed information for each individual participant. These data can then be re-analysed and included in meta-analyses. IPD meta-analysis should ideally be performed within the context of a systematic review, whereby authors seek to synthesise evidence from all relevant studies. However, some authors may use IPD meta-analysis to synthesise evidence from two or more studies for which IPD is available, without systematically searching for other relevant studies. Researchers have previously used an IPD approach to synthesise data from multiple pharmacogenetic studies.[230-233] In this work we did not specifically seek to identify and resolve challenges that may limit the potential for researchers to perform high-quality IPD meta-analyses of pharmacogenetic studies.

Nevitt et al.[234] investigated data retrieval success rates across IPD meta-analyses of all types of study design, for which studies were identified using systematic methods. Only 188 (25%) of 760 IPD meta-analyses published between 1987 and 2015 retrieved 100% of the eligible IPD for analysis; 324 (43%) of these IPD meta-analyses retrieved 80% or more of

relevant IPD. There was insufficient evidence to suggest that IPD retrieval rates have improved over time. These findings suggest that data retrieval is likely to be an obstacle for researchers conducting IPD meta-analyses of pharmacogenetic evidence within systematic reviews, although no formal review has been conducted in this area.

Raza and Hall[235] assert that effective sharing of genomic data does not occur in the UK, attributing this lack of sharing to 'limited technical and resource capacity to curate and upload data; concerns around the longevity and sustainability of third-party managed databases; and uncertainty around the legitimacy of sharing potentially identifiable patient data –especially into publically accessible databases'. An investigation of IPD retrieval rates across IPD meta-analyses of pharmacogenetic studies would be beneficial; such work would highlight the need for initiatives to overcome barriers to data sharing specifically within the pharmacogenetic research community.

Additionally, future work could also investigate whether challenges exist in the synthesis of pharmacogenetic evidence once IPD has been obtained; this work would apply to both IPD meta-analyses conducted in the context of a systematic review, and IPD meta-analyses for which systematic searches were not performed.

7.5 Concluding remarks

Stratified medicine is an evolving field that has the potential to transform mainstream health care by tailoring treatments to individuals. Using pharmacogenetic tests to guide prescribing practices can greatly improve outcomes for patients and ensure efficient use of health care providers' budgets. However, uptake of pharmacogenetic tests in clinical practice globally has been poor, despite the wealth of pharmacogenetic research that has been conducted in recent years. It has become evident that action must be taken to harness the efforts of researchers conducting pharmacogenetic studies, ensuring that the valuable data obtained in individual studies have a meaningful impact on health care policy and practice. In the UK, there has recently been a government-backed drive to integrate genomic medicine into routine NHS practice;[236] it is therefore now more important than ever to ensure that there is convincing evidence for actionable gene-drug associations.

Systematic reviews and meta-analyses can be used to combine results from individual pharmacogenetic studies, and are generally considered to be the "gold standard" of evidence-based research. If a systematic review and meta-analysis provides strong evidence for a pharmacogenetic association, this may prompt researchers to conduct RCTs

of genotype-guided prescribing practices, and in turn convince public bodies to introduce pharmacogenetic testing into clinical practice.

However, systematic reviews and meta-analyses rely heavily on information provided in the reports of included studies. This thesis has provided a detailed insight into the conduct of a systematic review and meta-analysis of associations between genetic variants and anti-TB drug-related tocixity, and has highlighted many inadequacies in the reporting of pharmacogenetic studies. We believe that the introduction of the STROPS guideline to the pharmacogenetic research community will have a valuable impact on the clarity and completeness of future study reports, and will allow the evidence base for pharmacogenetic associations to be strengthened by facilitating evidence synthesis.

References

1. Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000;405(6788):857-65.
2. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006;163(1):28-40.
3. Garrod AE. *Inborn errors of metabolism*. London: Frowde, Hodder & Stoughton; 1909.
4. Vogel F. Moderne probleme der humangenetik. In: Heilmeyer L, Schoen R, De Rudder B, editors. *Ergebnisse der inneren medizin und kinderheilkunde, vol 12*. Berlin: Springer; 1959:52-125.
5. Pirmohamed M. Pharmacogenetics and pharmacogenomics. *Br J Clin Pharmacol* 2001;52(4):345-7.
6. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 2012;92(4):414-7.
7. PharmGKB. *PharmGKB FAQs* [cited 2020 10 July]. Available from: https://www.pharmgkb.org/page/faqs.
8. Yadav ML, Mohapatra B. Intergenic. In: Vonk J, Shackelford T, editors. *Encyclopedia of Animal Cognition and Behavior*. Cham: Springer; 2017. Available from: https://doi.org/10.1007/978-3-319-47829-6_64-1.
9. Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C. Silent (synonymous) SNPs: should we care about them? In: Komar AA, editor. *Single nucleotide polymorphisms: methods and protocols.* 2nd ed. Totowa, NJ: Humana Press; 2009:23-39.
10. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet* 2015;24(R1):R102-R10.
11. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147(8):573-7.
12. Clinical Pharmacogenetics Implementation Consortium. *CPIC guidelines* 2019 [cited 2020 10 July]. Available from: https://cpicpgx.org/guidelines/.
13. Luzum JA, Pakyz RE, Elsey AR, Haidar CE, Peterson JF, Whirl-Carrillo M, et al. The Pharmacogenomics Research Network Translational Pharmacogenetics Program: outcomes and metrics of pharmacogenetic implementations across diverse healthcare systems. *Clin Pharmacol Ther* 2017;102(3):502-10.
14. Daly AK. Pharmacogenetics: a general review on progress to date. *Br Med Bull* 2017;124(1):65-79.
15. Verhoef TI, Redekop WK, Langenskiold S, Kamali F, Wadelius M, Burnside G, et al. Cost-effectiveness of pharmacogenetic-guided dosing of warfarin in the United Kingdom and Sweden. *Pharmacogenomics J* 2016;16(5):478-84.
16. Hughes DA, Vilar FJ, Ward CC, Alfirevic A, Park BK, Pirmohamed MJP, et al. Cost-effectiveness analysis of HLA B* 5701 genotyping in preventing abacavir hypersensitivity. 2004;14(6):335-42.
17. Zhang G, Nebert DW. Personalized medicine: genetic risk prediction of drug response. *Pharmacol Ther* 2017;175:75-90.
18. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. *Stat Med* 2008;27(30):6547-69.

19. Ioannidis JPA, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003;361(9357):567-71.
20. Smeraldi E, Zanardi R, Benedetti F, Di Bella D, Perez J, Catalano M. Polymorphism within the promoter of the serotonin transporter gene and antidepressant efficacy of fluvoxamine. *Mol Psychiatry* 1998;3(6):508-11.
21. Taylor MJ, Sen S, Bhagwagar Z. Antidepressant response and the serotonin transporter gene-linked polymorphic region. *Biol Psychiatry* 2010;68(6):536-43.
22. Keers R, Uher R, Huezo-Diaz P, Smith R, Jaffee S, Rietschel M, et al. Interaction between serotonin transporter gene variants and life events predicts response to antidepressants in the GENDEP project. *Pharmacogenomics J* 2011;11(2):138-45.
23. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21(11):1559-73.
24. Little J, Higgins J, editors. *The HuGENet™ HuGE review handbook, version 1.0*. Ottawa, Ontario, Canada: HuGENet Canada Coordinating Centre; 2006. Available from: http://www.medicine.uottawa.ca/public-health-genomics/web/assets/documents/HuGE_Review_Handbook_V1_0.pdf.
25. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane handbook for systematic reviews of interventions*. 2nd ed. Chichester (UK): John Wiley & Sons; 2019.
26. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263(10):1385-9.
27. Bouaziz M, Jeanmougin M, Guedj M. Multiple testing in large-scale genetic studies. In: Pompanon F, Bonin A, editors. *Data production and analysis in population genomics: methods and protocols*. Totowa, NJ: Humana Press; 2012:213-233.
28. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19(1):149-50.
29. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;38(2):209-13.
30. Qiu W, Lazarus R. *Power calculation for testing if disease is associated with marker in a case-control study using the GeneticsDesign package* 2019 [cited 2020 12 August]. Available from: https://www.bioconductor.org/packages/release/bioc/vignettes/GeneticsDesign/inst/doc/GPC.pdf.
31. Syed H, Jorgensen AL, Morris AP. SurvivalGWAS_Power: a user friendly tool for power calculations in pharmacogenetic studies with “time to event” outcomes. *BMC Bioinformatics* 2016;17:523. Doi:10.1186/s12859-016-1407-9.
32. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol* 1990;43(1):87-91.
33. Cha E-Y, Jeong H-E, Kim W-Y, Shin HJ, Kim H-S, Shin J-G. Brief introduction to current pharmacogenomics research tools. *Transl Clin Pharmacol* 2016;24(1):13-21.
34. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 2005;6(11):847-59.
35. Gordon D, Finch SJ, Nothnagel M, Ott J. Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum Hered* 2002;54(1):22-33.
36. Moskvina V, Craddock N, Holmans P, Owen MJ, O’Donovan MC. Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered* 2006;61(1):55-64.
37. Shi S, Yuan N, Yang M, Du Z, Wang J, Sheng X, et al. Comprehensive assessment of genotype imputation performance. *Hum Hered* 2018;83(3):107-16.

38. Little RJA, Rubin DB. *Statistical analysis with missing data*. 3rd ed. Chichester (UK): John Wiley & Sons; 2019.
39. Leschziner G, Jorgensen AL, Pirmohamed M, Williamson PR, Marson AG, Coffey AJ, et al. Clinical factors and ABCB1 polymorphisms in prediction of antiepileptic drug response: a prospective cohort study. *Lancet Neurol* 2006;5(8):668-76.
40. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437(7063):1299-320.
41. Hao K, Cawley S. Differential dropout among SNP genotypes and impacts on association tests. *Hum Hered* 2007;63(3-4):219-28.
42. Rogers AJ, Weiss ST. Epidemiologic and population genetic studies. In: Roberston D, Williams GH, editors. *Clinical and translational science: principles of human research.* 2nd ed. Academic Press; 2017:313-326.
43. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population stratification in genetic association studies. *Curr Protoc Hum Genet* 2017;95:1.22.1-1.22.23. Doi:10.1002/cphg.48.
44. Salanti G, Amountza G, Ntzani EE, Ioannidis JPA. Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet* 2005;13(7):840-8.
45. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *Eur J Hum Genet* 2004;12(5):395-9.
46. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol* 2003;56(4):297-303.
47. Bagos PG. Genetic model selection in genome-wide association studies: robust methods and the use of meta-analysis. *Stat Appl Genet Mol Biol* 2013;12(3):285-308.
48. Jorgensen AL, FitzGerald RJ, Oyee J, Pirmohamed M, Williamson PR. Influence of CYP2C9 and VKORC1 on patient response to warfarin: a systematic review and meta-analysis. *PLoS One* 2012;7(8):e44064. Doi:10.1371/journal.pone.0044064.
49. Contopoulos-Ioannidis DG, Alexiou GA, Gouvias TC, Ioannidis JPA. An empirical evaluation of multifarious outcomes in pharmacogenetics: beta-2 adrenoceptor gene polymorphisms in asthma treatment. *Pharmacogenet Genomics* 2006;16(10):705-11.
50. Sabaté E. *Adherence to long-term therapies: evidence for action*. Geneva: World Health Organization; 2003.
51. Horne R, Weinman J, Barber N, Elliott R, Morgan M, Cribb A, et al. *Concordance, adherence and compliance in medicine taking.* Report for the National Co-ordinating Centre for NHS Service Delivery and Organisation (NCCSDO), December 2005 [cited 2020 14 July]. Available from: http://www.netscc.ac.uk/hsdr/files/project/SDO_FR_08-1412-076_V01.pdf.
52. Mefford J, Witte JS. The covariate's dilemma. *PLoS Genet* 2012;8(11):e1003096. Doi:10.1371/journal.pgen.1003096.
53. Pirinen M, Donnelly P, Spencer CCA. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat Genet* 2012;44(8):848-51.
54. Zaitlen N, Lindström S, Pasaniuc B, Cornelis M, Genovese G, Pollack S, et al. Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet* 2012;8(11):e1003032. Doi:10.1371/journal.pgen.1003032.
55. Zaitlen N, Paşaniuc B, Patterson N, Pollack S, Voight B, Groop L, et al. Analysis of case–control association studies with known risk variants. *Bioinformatics* 2012;28(13):1729-37.
56. Zhang H, Chatterjee N, Rader D, Chen J. Adjustment of nonconfounding covariates in case-control genetic association studies. *Ann Appl Stat* 2018;12(1):200-21.

57. Centre for Reviews and Dissemination (CRD). *Systematic reviews.* 3rd ed. York: CRD, University of York; 2009. Available from: https://www.york.ac.uk/crd/guidance/.
58. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med* 2009;6(7):e1000097. Doi:10.1371/journal.pmed.1000097.
59. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;283(15):2008-12.
60. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004;33(1):30-42.
61. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol* 2005;34(6):1319-28.
62. Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JPA. Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. *Am J Epidemiol* 2006;163(4):300-9.
63. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. How should we use information about HWE in the meta-analyses of genetic association studies? *Int J Epidemiol* 2008;37(1):136-46.
64. Ioannidis JPA, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005;58(6):543-9.
65. Ioannidis JPA. Genetic associations: false or true? *Trends Mol Med* 2003;9(4):135-8.
66. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002;3(4):299-309.
67. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34.
68. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25(20):3443-57.
69. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006;295(6):676-80.
70. Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. *Stat Med* 2008;27(5):746-63.
71. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 2005;58(9):894-901.
72. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002. Doi:10.1136/bmj.d4002.
73. Salanti G, Higgins JPT. Meta-analysis of genetic association studies under different inheritance models using data reported as merged genotypes. *Stat Med* 2008;27(5):764-77.
74. Verzilli C, Shah T, Casas JP, Chapman J, Sandhu M, Debenham SL, et al. Bayesian meta-analysis of genetic association studies with different sets of markers. *Am J Hum Genet* 2008;82(4):859-72.
75. Thompson JR, Minelli C, Abrams KR, Thakkinstian A, Attia J. Combining information from related meta-analyses of genetic association studies. *J R Stat Soc Ser C Appl Stat* 2008;57(1):103-15.
76. World Health Organization. *Global tuberculosis report* 2019 [cited 2020 10 July]. Available from: https://www.who.int/tb/publications/global_report/en/.
77. Tostmann A, Boeree MJ, Aarnoutse RE, De Lange WCM, Van Der Ven AJAM, Dekhuijzen R. Antituberculosis drug-induced hepatotoxicity: concise up-to-date review. *J Gastroenterol Hepatol* 2008;23(2):192-202.

78. Dash LA, Comstock GW, Flynn JPG. Isoniazid preventive therapy: retrospect and prospect. *Am Rev Respir Dis* 1980;121(6):1039-44.
79. Roy PD, Majumder M, Roy B. Pharmacogenomics of anti-TB drugs-related hepatotoxicity. *Pharmacogenomics* 2008;9(3):311-21.
80. Fukino K, Sasaki Y, Hirai S, Nakamura T, Hashimoto M, Yamagishi F, et al. Effects of N-acetyltransferase 2 (NAT2), CYP2E1 and glutathione-S-transferase (GST) genotypes on the serum concentrations of isoniazid and metabolites in tuberculosis patients. *J Toxicol Sci* 2008;33(2):187-95.
81. Azuma J, Ohno M, Kubota R, Yokota S, Nagai T, Tsuyuguchi K, et al. NAT2 genotype guided regimen reduces isoniazid-induced liver injury and early treatment failure in the 6-month four-drug standard treatment of tuberculosis: a randomized controlled trial for pharmacogenetics-based therapy. *Eur J Clin Pharmacol* 2013;69(5):1091-101.
82. Ho H-T, Wang T-H, Hsiong C-H, Perng W-C, Wang N-C, Huang T-Y, et al. The NAT2 tag SNP rs1495741 correlates with the susceptibility of antituberculosis drug-induced hepatotoxicity. *Pharmacogenet Genomics* 2013;23(4):200-7.
83. Watanabe J, Hayashi S-I, Nakachi K, Imai K, Suda Y, Sekine T, et al. PstI and RsaI RFLPs in complete linkage disequilibrium at the CYP2E gene. *Nucleic Acids Res* 1990;18(23):7194.
84. Hayashi S-i, Watanabe J, Kawajiri K. Genetic polymorphisms in the 5'-flanking region change transcriptional regulation of the human cytochrome P45011E1 gene. *J Biochem* 1991;110(4):559-65.
85. Wu X, Amos CI, Kemp BL, Shi H, Jiang H, Wan Y, et al. Cytochrome P450 2E1 DraI polymorphisms in lung cancer in minority populations. *Cancer Epidemiol Biomarkers Prev* 1998;7(1):13-8.
86. Brito TC, Possuelo LG, Valim ARM, Todendi PF, Ribeiro AW, Gregianini TS, et al. Polymorphisms in CYP2E1, GSTM1 and GSTT1 and anti-tuberculosis drug-induced hepatotoxicity. *An Acad Bras Cienc* 2014;86(2):855-65.
87. Wang P, Pradhan K, Zhong X-B, Ma X. Isoniazid metabolism and hepatotoxicity. *Acta Pharm Sin B* 2016;6(5):384-92.
88. Pandit A, Sachdeva T, Bafna P. Drug-induced hepatotoxicity: a review. *J Appl Pharm Sci* 2012;2(5):233-43.
89. Kaplowitz N, DeLeve LD. *Drug-induced liver disease*. 3rd ed. Cambridge, MA: Academic Press; 2013.
90. Li LM, Chen L, Deng GH, Tan WT, Dan YJ, Wang RQ, et al. SLCO1B1 *15 haplotype is associated with rifampin-induced liver injury. *Mol Med Rep* 2012;6(1):75-82.
91. Ramachandran G, Swaminathan S. Role of pharmacogenomics in the treatment of tuberculosis: a review. *Pharmacogenomics Pers Med* 2012;5:89-98.
92. Cai Y, Yi J, Zhou C, Shen X. Pharmacogenetic study of drug-metabolising enzyme polymorphisms on the risk of anti-tuberculosis drug-induced liver injury: a meta-analysis. *PLoS One* 2012;7(10):e47769. Doi:10.1371/journal.pone.0047769.
93. Cai L, Cai MH, Wang MY, Xu YF, Chen WZ, Qin SY, et al. Meta-analysis-based preliminary exploration of the connection between ATDILI and schizophrenia by GSTM1/T1 gene polymorphisms. *PLoS One* 2015;10(6):e0128643. Doi:10.1371/journal.pone.0128643.
94. Deng R, Yang T, Wang Y, Tang N. CYP2E1 RsaI/PstI polymorphism and risk of anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis* 2012;16(12):1574-81.
95. Du H, Chen X, Fang Y, Yan O, Xu H, Li L, et al. Slow N-acetyltransferase 2 genotype contributes to anti-tuberculosis drug-induced hepatotoxicity: a meta-analysis. *Mol Biol Rep* 2013;40(5):3591-6.

96. Li C, Long J, Hu X, Zhou Y. GSTM1 and GSTT1 genetic polymorphisms and risk of anti-tuberculosis drug-induced hepatotoxicity: an updated meta-analysis. *Eur J Clin Microbiol Infect Dis* 2013;32(7):859-68.
97. Sheng Y-J, Wu G, He H-Y, Chen W, Zou Y-S, Li Q, et al. The association between CYP2E1 polymorphisms and hepatotoxicity due to anti-tuberculosis drugs: a meta-analysis. *Infect Genet Evol* 2014;24:34-40.
98. Shi J, Xie M, Wang J, Xu Y, Liu X. Susceptibility of N-acetyltransferase 2 slow acetylators to antituberculosis drug-induced liver injury: a meta-analysis. *Pharmacogenomics* 2015;16(18):2083-97.
99. Sun F, Chen Y, Xiang Y, Zhan S. Drug-metabolising enzyme polymorphisms and predisposition to anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis* 2008;12(9):994-1002.
100. Tang N, Deng R, Wang Y, Lin M, Li H, Qiu Y, et al. GSTM1 and GSTT1 null polymorphisms and susceptibility to anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis* 2013;17(1):17-25.
101. Wang PY, Xie SY, Hao Q, Zhang C, Jiang BF. NAT2 polymorphisms and susceptibility to anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis* 2012;16(5):589-95.
102. Wang FJ, Wang Y, Niu T, Lu WX, Sandford A, He JQ. Update meta-analysis of the CYP2E1 RsaI/PstI and DraI polymorphisms and risk of antituberculosis drug-induced hepatotoxicity: evidence from 26 studies. *J Clin Pharm Ther* 2016;41(3):334-40.
103. Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002;156(4):300-10.
104. Richardson M, Kirkham J, Dwan K, Sloan D, Davies G, Jorgensen A. Influence of genetic variants on toxicity to anti-tubercular agents: a systematic review and meta-analysis (protocol). *Syst Rev* 2017;6:142. Doi:10.1186/s13643-017-0533-4.
105. Veritas Health Innovation. *Covidence systematic review software* [cited 2020 10 July]. Available from: www.covidence.org.
106. U.S. National Library of Medicine. *Help me understand genetics* 2020 [cited 2020 10 July]. Available from: https://ghr.nlm.nih.gov/primer.
107. Huang YS, Chern HD, Su WJ, Wu JC, Lai SL, Yang SY, et al. Polymorphism of the N-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug–induced hepatitis. *Hepatology* 2002;35(4):883-9.
108. Rana SV, Kamboj JK, Sharma SK, Ola RP, Sinha SK, Singh K. Antioxidant status and GST gene polymorphisms in antitubercular treatment-induced hepatotoxicity patients. *Hepatol Int* 2013;7(3):876-82.
109. Rana SV, Ola RP, Sharma SK, Arora SK, Sinha SK, Pandhi P, et al. Comparison between acetylator phenotype and genotype polymorphism of N-acetyltransferase-2 in tuberculosis patients. *Hepatol Int* 2012;6(1):397-402.
110. Singh M, Gupta VH, Amarapurkar DN, Joshi JM, Baijal R, Ramegowda PH, et al. Association of genetic variants with anti-tuberculosis drug induced hepatotoxicity: a high resolution melting analysis. *Infect Genet Evol* 2014;23:42-8.
111. An HR, Wu XQ, Wang ZY, Zhang JX, Liang Y. NAT2 and CYP2E1 polymorphisms associated with antituberculosis drug-induced hepatotoxicity in Chinese patients. *Clin Exp Pharmacol Physiol* 2012;39(6):535-43.
112. Bose PD, Sarma MP, Medhi S, Das BC, Husain SA, Kar P. Role of polymorphic N-acetyl transferase2 and cytochrome P4502E1 gene in antituberculosis treatment-induced hepatitis. *J Gastroenterol Hepatol* 2011;26(2):312-8.
113. Çetintaş VB, Erer OF, Kosova B, Özdemir I, Topçuoğlu N, Aktoğu S, et al. Determining the relation between N-acetyltransferase-2 acetylator phenotype and

antituberculosis drug induced hepatitis by molecular biologic tests. *Tuberk Toraks* 2008;56(1):81-6.
114. Chamorro JG, Castagnino JP, Musella RM, Nogueras M, Aranda FM, Frías A, et al. Sex, ethnicity, and slow acetylator profile are the major causes of hepatotoxicity induced by antituberculosis drugs. *J Gastroenterol Hepatol* 2013;28(2):323-8.
115. Chang J-C, Liu EH, Lee C-N, Lin Y-C, Yu M-C, Bai K-J, et al. UGT1A1 polymorphisms associated with risk of induced liver disorders by anti-tuberculosis medications. *Int J Tuberc Lung Dis* 2012;16(3):376-8.
116. Chatterjee S, Lyle N, Mandal A, Kundu S. GSTT1 and GSTM1 gene deletions are not associated with hepatotoxicity caused by antitubercular drugs. *J Clin Pharm Ther* 2010;35(4):465-70.
117. Chen R, Wang J, Tang S, Zhang Y, Lv X, Wu S, et al. Association of polymorphisms in drug transporter genes (SLCO1B1 and SLC10A1) and anti-tuberculosis drug-induced hepatotoxicity in a Chinese cohort. *Tuberculosis (Edinb)* 2015;95(1):68-74.
118. Cho HJ, Koh WJ, Ryu YJ, Ki CS, Nam MH, Kim JW, et al. Genetic polymorphisms of NAT2 and CYP2E1 associated with antituberculosis drug-induced hepatotoxicity in Korean patients with pulmonary tuberculosis. *Tuberculosis (Edinb)* 2007;87(6):551-6.
119. Costa GNO, Magno LAV, Santana CVN, Konstantinovas C, Saito ST, Machado M, et al. Genetic interaction between NAT2, GSTM1, GSTT1, CYP2E1, and environmental factors is associated with adverse reactions to anti-tuberculosis drugs. *Mol Diagn Ther* 2012;16(4):241-50.
120. Dhoro M, Ngara B, Kadzirange G, Nhachi C, Masimirembwa C. Genetic variants of drug metabolizing enzymes and drug transporter (ABCB1) as possible biomarkers for adverse drug reactions in an HIV/AIDS cohort in Zimbabwe. *Curr HIV Res* 2013;11(6):481-90.
121. Feng FM, Guo M, Chen Y, Li S, Zhang P, Sun SF, et al. Genetic polymorphisms in metabolic enzymes and susceptibility to anti-tuberculosis drug-induced hepatic injury. *Genet Mol Res* 2014;13(4):9463-71.
122. Fernandes DC, Santos NPC, Moraes MR, Braga AC, Silva CA, Ribeiro-dos-Santos A, et al. Association of the CYP2B6 gene with anti-tuberculosis drug-induced hepatotoxicity in a Brazilian Amazon population. *Int J Infect Dis* 2015;33:28-31.
123. Fredj NB, Gam R, Kerkni E, Chaabane A, Chadly Z, Boughattas N, et al. Risk factors of isoniazid-induced hepatotoxicity in Tunisian tuberculosis patients. *Pharmacogenomics J* 2016;17(4):372-7.
124. Gogtay NJ, Kapileshwar SR, Shah SU, Bendkhale SR, Ramakrishna S, Sridharan K, et al. Evaluation of cytochrome P4502E1 polymorphisms in healthy adult Western Indians and patients with antituberculous drug-induced hepatotoxicity. *Indian J Pharmacol* 2016;48(1):42-6.
125. Gupta VH, Singh M, Amarapurkar DN, Sasi P, Joshi JM, Baijal R, et al. Association of GST null genotypes with anti-tuberculosis drug induced hepatotoxicity in Western Indian population. *Ann Hepatol* 2013a;12(6):959-65.
126. Gupta VH, Amarapurkar DN, Singh M, Sasi P, Joshi JM, Baijal R, et al. Association of N-acetyltransferase 2 and cytochrome P450 2E1 gene polymorphisms with antituberculosis drug-induced hepatotoxicity in Western India. *J Gastroenterol Hepatol* 2013b;28(8):1368-74.
127. He L, Gao L, Shi Z, Li Y, Zhu L, Li S, et al. Involvement of cytochrome P450 1A1 and glutathione S-transferase P1 polymorphisms and promoter hypermethylation in the progression of anti-tuberculosis drug-induced liver injury: a case–control study. *PLoS One* 2015;10(3):e0119481. Doi:10.1371/journal.pone.0119481.
128. Higuchi N, Tahara N, Yanagihara K, Fukushima K, Suyama N, Inoue Y, et al. NAT2* 6A, a haplotype of the N-acetyltransferase 2 gene, is an important biomarker for risk of

anti-tuberculosis drug-induced hepatotoxicity in Japanese patients with tuberculosis. *World J Gastroenterol* 2007;13(45):6003-8.
129. Huang YS, Chern HD, Su WJ, Wu JC, Chang SC, Chiang CH, et al. Cytochrome P450 2E1 genotype and the susceptibility to antituberculosis drug-induced hepatitis. *Hepatology* 2003;37(4):924-30.
130. Huang YS, Su WJ, Huang YH, Chen CY, Chang FY, Lin HC, et al. Genetic polymorphisms of manganese superoxide dismutase, NAD(P)H: quinone oxidoreductase, glutathione S-transferase M1 and T1, and the susceptibility to drug-induced liver injury. *J Hepatol* 2007;47(1):128-34.
131. Jung JA, Kim T-e, Lee H, Jeong B-H, Park HY, Jeon K, et al. A proposal for an individualized pharmacogenetic-guided isoniazid dosage regimen for patients with tuberculosis. *Drug Des Devel Ther* 2015;9:5433-8.
132. Khalili H, Fouladdel S, Sistanizad M, Hajiabdolbaghi M, Azizi E. Association of N-acetyltransferase-2 genotypes and anti-tuberculosis induced liver injury: first case-controlled study from Iran. *Curr Drug Saf* 2011;6(1):17-22.
133. Kim S-H, Kim S-H, Bahn J-W, Kim Y-K, Chang Y-S, Shin E-S, et al. Genetic polymorphisms of drug-metabolizing enzymes and anti-TB drug-induced hepatitis. *Pharmacogenomics* 2009;10(11):1767-79.
134. Kim S-H, Kim S-H, Yoon HJ, Shin DH, Park SS, Kim Y-S, et al. GSTT1 and GSTM1 null mutations and adverse reactions induced by antituberculosis drugs in Koreans. *Tuberculosis* 2010;90(1):39-43.
135. Kim S-H, Kim S-H, Yoon HJ, Shin DH, Park SS, Kim Y-S, et al. NAT2, CYP2C9, CYP2C19, and CYP2E1 genetic polymorphisms in anti-TB drug-induced maculopapular eruption. *Eur J Clin Pharmacol* 2011;67(2):121-7.
136. Kim S-H, Jee Y-K, Lee J-H, Lee B-H, Kim Y-S, Park J-S, et al. ABCC2 haplotype is associated with antituberculosis drug-induced maculopapular eruption. *Allergy Asthma Immunol Res* 2012a;4(6):362-6.
137. Kim S-H, Kim S-H, Lee J-H, Lee B-H, Kim Y-S, Park J-S, et al. Polymorphisms in drug transporter genes (ABCB1, SLCO1B1 and ABCC2) and hepatitis induced by antituberculosis drugs. *Tuberculosis* 2012b;92(1):100-4.
138. Kim SH, Kim SH, Yoon HJ, Shin DH, Park SS, Kim YS, et al. TNF-α genetic polymorphism− 308G/A and antituberculosis drug-induced hepatitis. *Liver Int* 2012c;32(5):809-14.
139. Kim S-H, Kim S-H, Lee J-H, Lee B-H, Yoon HJ, Shin DH, et al. Superoxide dismutase gene (SOD1, SOD2, and SOD3) polymorphisms and antituberculosis drug-induced hepatitis. *Allergy Asthma Immunol Res* 2015;7(1):88-91.
140. Kwon J-W, Shin E-S, Lee J-E, Kim S-H, Kim S-H, Jee Y-K, et al. Genetic variations in TXNRD1 as potential predictors of drug-induced liver injury. *Allergy Asthma Immunol Res* 2012;4(3):132-6.
141. Lee SW, Chung L, Huang HH, Chuang TY, Liou YH, Wu L. NAT2 and CYP2E1 polymorphisms and susceptibility to first-line anti-tuberculosis drug-induced hepatitis. *Int J Tuberc Lung Dis* 2010;14(5):622-6.
142. Leiro V, Fernández-Villar A, Valverde D, Constenla L, Vázquez R, Piñeiro L, et al. Influence of glutathione S-transferase M1 and T1 homozygous null mutations on the risk of antituberculosis drug-induced hepatotoxicity in a Caucasian population. *Liver Int* 2008;28(6):835-9.
143. Leiro-Fernandez V, Valverde D, Vázquez-Gallardo R, Botana-Rial M, Constenla L, Agúndez JA, et al. N-acetyltransferase 2 polymorphisms and risk of anti-tuberculosis drug-induced hepatotoxicity in Caucasians. *Int J Tuberc Lung Dis* 2011;15(10):1403-8.

144. Liu F, Jiao AX, Wu XR, Zhao W, Yin QQ, Qi H, et al. Impact of glutathione S-transferase M1 and T1 on anti-tuberculosis drug–induced hepatotoxicity in Chinese pediatric patients. *PLoS One* 2014;9(12):e115410. Doi:10.1371/journal.pone.0115410.
145. Lv X, Tang S, Xia Y, Zhang Y, Wu S, Yang Z, et al. NAT2 genetic polymorphisms and anti-tuberculosis drug-induced hepatotoxicity in Chinese community population. *Ann Hepatol* 2012;11(5):700-7.
146. Mahmoud LB, Ghozzi H, Kamoun A, Hakim A, Hachicha H, Hammami S, et al. Polymorphism of the N-acetyltransferase 2 gene as a susceptibility risk factor for antituberculosis drug-induced hepatotoxicity in Tunisian patients with tuberculosis. *Pathol Biol (Paris)* 2012;60(5):324-30.
147. Monteiro TP, El-Jaick KB, Jeovanio-Silva AL, Brasil PEAA, Costa MJM, Rolla VC, et al. The roles of GSTM1 and GSTT1 null genotypes and other predictors in anti-tuberculosis drug-induced liver injury. *J Clin Pharm Ther* 2012;37(6):712-8.
148. Nanashima K, Mawatari T, Tahara N, Higuchi N, Nakaura A, Inamine T, et al. Genetic variants in antioxidant pathway: risk factors for hepatotoxicity in tuberculosis patients. *Tuberculosis* 2012;92(3):253-9.
149. Ng C-S, Hasnat A, Al Maruf A, Ahmed MU, Pirmohamed M, Day CP, et al. N-acetyltransferase 2 (NAT2) genotype as a risk factor for development of drug-induced liver injury relating to antituberculosis drug treatment in a mixed-ethnicity patient group. *Eur J Clin Pharmacol* 2014;70(9):1079-86.
150. Ohno M, Yamaguchi I, Yamamoto I, Fukuda T, Yokota S, Maekura R, et al. Slow N-acetyltransferase 2 genotype affects the incidence of isoniazid and rifampicin-induced hepatotoxicity. *Int J Tuberc Lung Dis* 2000;4(3):256-61.
151. Possuelo LG, Castelan JA, De Brito TC, Ribeiro AW, Cafrune PI, Picon PD, et al. Association of slow N-acetyltransferase 2 profile and anti-TB drug-induced hepatotoxicity in patients from Southern Brazil. *Eur J Clin Pharmacol* 2008;64(7):673-81.
152. Rana SV, Sharma SK, Ola RP, Kamboj JK, Malik A, Morya RK, et al. N-acetyltransferase 2, cytochrome P4502E1 and glutathione S-transferase genotypes in antitubercular treatment-induced hepatotoxicity in North Indians. *J Clin Pharm Ther* 2014;39(1):91-6.
153. Roy B, Chowdhury A, Kundu S, Santra A, Dey B, Chakraborty M, et al. Increased risk of antituberculosis drug-induced hepatotoxicity in individuals with glutathione S-transferase M1 'null'mutation. *J Gastroenterol Hepatol* 2001;16(9):1033-7.
154. Roy B, Ghosh SK, Sutradhar D, Sikdar N, Mazumder S, Barman S. Predisposition of antituberculosis drug induced hepatotoxicity by cytochrome P450 2E1 genotype and haplotype in pediatric patients. *J Gastroenterol Hepatol* 2006;21(4):784-6.
155. Santos NPC, Callegari-Jacques SM, Ribeiro Dos Santos AKC, Silva CA, Vallinoto ACR, Fernandes DCRO, et al. N-acetyl transferase 2 and cytochrome P450 2E1 genes and isoniazid-induced hepatotoxicity in Brazilian patients. *Int J Tuberc Lung Dis* 2013;17(4):499-504.
156. Sharma SK, Jha BK, Sharma A, Sreenivas V, Upadhyay V, Jaisinghani C, et al. Genetic polymorphisms of CYP2E1 and GSTM1 loci and susceptibility to anti-tuberculosis drug-induced hepatotoxicity. *Int J Tuberc Lung Dis* 2014;18(5):588-93.
157. Shimizu Y, Dobashi K, Mita Y, Endou K, Moriya S, Osano K, et al. DNA microarray genotyping of N-acetyltransferase 2 polymorphism using carbodiimide as the linker for assessment of isoniazid hepatotoxicity. *Tuberculosis* 2006;86(5):374-81.
158. Singla N, Gupta D, Birbian N, Singh J. Association of NAT2, GST and CYP2E1 polymorphisms and anti-tuberculosis drug-induced hepatotoxicity. *Tuberculosis* 2014;94(3):293-8.

159. Sotsuka T, Sasaki Y, Hirai S, Yamagishi F, Ueno K. Association of isoniazid-metabolizing enzyme genotypes and isoniazid-induced hepatotoxicity in tuberculosis patients. *In Vivo* 2011;25(5):803-12.
160. Tang SW, Lv XZ, Zhang Y, Wu SS, Yang ZR, Xia YY, et al. CYP2E1, GSTM1 and GSTT1 genetic polymorphisms and susceptibility to antituberculosis drug-induced hepatotoxicity: a nested case–control study. *J Clin Pharm Ther* 2012;37(5):588-93.
161. Tang S, Lv X, Zhang Y, Wu S, Yang Z, Xia Y, et al. Cytochrome P450 2E1 gene polymorphisms/haplotypes and anti-tuberculosis drug-induced hepatitis in a Chinese cohort. *PLoS One* 2013a;8(2):e57526. Doi:10.1371/journal.pone.0057526.
162. Tang SW, Lv XZ, Chen R, Wu SS, Yang ZR, Chen DF, et al. Lack of association between genetic polymorphisms of CYP3A4, CYP2C9 and CYP2C19 and antituberculosis drug-induced liver injury in a community-based Chinese population. *Clin Exp Pharmacol Physiol* 2013b;40(5):326-32.
163. Teixeira RL, Morato RG, Cabello PH, Muniz LM, Moreira A, Kritski AL, et al. Genetic polymorphisms of NAT2, CYP2E1 and GST enzymes and the occurrence of antituberculosis drug-induced hepatitis in Brazilian TB patients. *Mem Inst Oswaldo Cruz* 2011;106(6):716-24.
164. Vuilleumier N, Rossier MF, Chiappe A, Degoumois F, Dayer P, Mermillod B, et al. CYP2E1 genotype and isoniazid-induced hepatotoxicity in patients treated for latent tuberculosis. *Eur J Clin Pharmacol* 2006;62(6):423-9.
165. Wang T, Yu HT, Wang W, Pan YY, He LX, Wang ZY. Genetic polymorphisms of cytochrome P450 and glutathione S-transferase associated with antituberculosis drug-induced hepatotoxicity in Chinese tuberculosis patients. *J Int Med Res* 2010;38(3):977-86.
166. Wang J-Y, Liu C-H, Hu F-C, Chang H-C, Liu J-L, Chen J-M, et al. Risk factors of hepatitis during anti-tuberculous treatment and implications of hepatitis virus load. *J Infect* 2011;62(6):448-55.
167. Wang J, Chen R, Tang S, Lv X, Wu S, Zhang Y, et al. Analysis of IL-6, STAT3 and HSPA1L gene polymorphisms in anti-tuberculosis drug-induced hepatitis in a nested case-control study. *PLoS One* 2015a;10(3):e0118862. Doi:10.1371/journal.pone.0118862.
168. Wang JY, Tsai CH, Lee YL, Lee LN, Hsu CL, Chang HC, et al. Gender-dimorphic impact of PXR genotype and haplotype on hepatotoxicity during antituberculosis treatment. *Medicine (Baltimore)* 2015b;94(24):e982. Doi:10.1097/MD.0000000000000982.
169. Wang J, Chen R, Tang S, Lv X, Wu S, Zhang Y, et al. Interleukin-4 and interleukin-10 polymorphisms and antituberculosis drug-induced hepatotoxicity in Chinese population. *J Clin Pharm Ther* 2015c;40(2):186-91.
170. Xiang Y, Ma L, Wu W, Liu W, Li Y, Zhu X, et al. The incidence of liver injury in Uyghur patients treated for TB in Xinjiang Uyghur autonomous region, China, and its association with hepatic enzyme polymorphisms NAT2, CYP2E1, GSTM1 and GSTT1. *PLoS One* 2014;9(1):e85905. Doi:10.1371/journal.pone.0085905.
171. Yamada S, Tang M, Richardson K, Halaschek-Wiener J, Chan M, Cook VJ, et al. Genetic variations of NAT2 and CYP2E1 and isoniazid hepatotoxicity in a diverse population. *Pharmacogenomics* 2009;10(9):1433-45.
172. Yamada S, Richardson K, Tang M, Halaschek-Wiener J, Cook VJ, Fitzgerald JM, et al. Genetic variation in carboxylesterase genes and susceptibility to isoniazid-induced hepatotoxicity. *Pharmacogenomics J* 2010;10(6):524-36.
173. Yimer G, Ueda N, Habtewold A, Amogne W, Suda A, Riedel K-D, et al. Pharmacogenetic & pharmacokinetic biomarker for efavirenz based ARV and rifampicin based anti-TB drug induced liver injury in TB-HIV infected patients. *PLoS One* 2011;6(12):e27810. Doi:10.1371/journal.pone.0027810.
174. Yuliwulandari R, Susilowati RW, Wicaksono BD, Viyati K, Prayuni K, Razari I, et al. NAT2 variants are associated with drug-induced liver injury caused by anti-tuberculosis drugs in Indonesian patients with tuberculosis. *J Hum Genet* 2016;61(6):533-7.

175. Zaverucha-do-Valle C, Monteiro SP, El-Jaick KB, Rosadas LA, Costa MJM, Quintana MSB, et al. The role of cigarette smoking and liver enzymes polymorphisms in anti-tuberculosis drug-induced hepatotoxicity in Brazilian patients. *Tuberculosis* 2014;94(3):299-305.
176. Zazuli Z, Barliana MI, Mulyani UA, Perwitasari DA, Ng H, Abdulah R. Polymorphism of PXR gene associated with the increased risk of drug-induced liver injury in Indonesian pulmonary tuberculosis patients. *J Clin Pharm Ther* 2015;40(6):680-4.
177. Benichou C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. *J Hepatol* 1990;11(2):272-6.
178. Danan G, Benichou C. Causality assessment of adverse reactions to drugs—I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J Clin Epidemiol* 1993;46(11):1323-30.
179. Takikawa H, Onji M. A proposal of the diagnostic scale of drug-induced liver injury. *Hepatol Res* 2005;32(4):250-1.
180. World Health Organization: Uppsala Monitoring Centre. *The use of the WHO-UMC system for standardized case causality assessment* 2005 [cited 2020 13 July]. Available from: https://www.who-umc.org/media/164200/who-umc-causality-assessment_new-logo.pdf.
181. Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. I. Performance characteristics of laboratory tests. *Clin Chem* 2000;46(12):2027-49.
182. DILIN. *Drug-Induced Liver Injury Network* [cited 2020 10 July]. Available from: http://www.dilin.org/.
183. Saukkonen JJ, Cohn DL, Jasmer RM, Schenker S, Jereb JA, Nolan CM, et al. An official ATS statement: hepatotoxicity of antituberculosis therapy. *Am J Respir Crit Care Med* 2006;174(8):935-52.
184. Maria VA, Victorino RM. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology* 1997;26(3):664-9.
185. Teschke R, Schwarzenboeck A, Hennermann KH. Causality assessment in hepatotoxicity by drugs and dietary supplements. *Br J Clin Pharmacol* 2008;66(6):758-66.
186. Bégaud B, Evreux JC, Jouglard J, Lagier G. Imputabilité des effets inattendus ou toxiques des médicaments: actualisation de la méthode utilisée en France. *Therapie* 1985;40(2):111-8.
187. Aithal GP, Rawlins MD, Day CP. Clinical diagnostic scale: a useful tool in the evaluation of suspected hepatotoxic adverse drug reactions. *J Hepatol* 2000;33(6):949-52.
188. Aithal GP, Watkins PB, Andrade RJ, Larrey D, Molokhia M, Takikawa H, et al. Case definition and phenotype standardization in drug-induced liver injury. *Clin Pharmacol Ther* 2011;89(6):806-15.
189. COMET Initiative. *Core Outcome Measures in Effectiveness Trials* [cited 2020 10 July]. Available from: http://www.comet-initiative.org/.
190. Sarah AT, Floyd AR, Françoise RF, Christopher E, Alessia R, Alain F, et al. The genetic structure and history of Africans and African Americans. *Science* 2009;324(5930):1035-44.
191. Mavridis D, Salanti G. How to assess publication bias: funnel plot, trim-and-fill method and selection models. *Evid Based Ment Health* 2014;17(1):30.
192. StataCorp. *Stata statistical software: release 14*. College Station, TX: StataCorp LP; 2015.
193. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
194. Sabbagh A, Langaney A, Darlu P, Gérard N, Krishnamoorthy R, Poloni ES. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. *BMC Genet* 2008;9:21. Doi:10.1186/471-2156-9-21.

195. Kim J-H, Cheong HS, Park BL, Kim LH, Shin HJ, Na HS, et al. Direct sequencing and comprehensive screening of genetic polymorphisms on CYP2 family genes (CYP2A6, CYP2B6, CYP2C8, and CYP2E1) in five ethnic populations. *Arch Pharm Res* 2015;38(1):115-28.
196. Polimanti R, Carboni C, Baesso I, Piacentini S, Iorio A, De Stefano GF, et al. Genetic variability of glutathione S-transferase enzymes in human populations: functional inter-ethnic differences in detoxification systems. *Gene* 2013;512(1):102-7.
197. Suvichapanich S, Wattanapokayakit S, Mushiroda T, Yanai H, Chuchottawon C, Kantima T, et al. Genomewide association study confirming the association of NAT2 with susceptibility to antituberculosis drug-induced liver injury in Thai patients. *Antimicrob Agents Chemother* 2019;63(8):e02692-18. Doi:10.1128/AAC.02692-18.
198. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308-11.
199. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, Von Elm E, et al. STrengthening the REporting of Genetic Association studies (STREGA)—an extension of the STROBE statement. *Genet Epidemiol* 2009;33(7):581-98.
200. Antonarakis SE, Group NW. Recommendations for a nomenclature system for human gene mutations. *Hum Mutat* 1998;11(1):1-3.
201. Den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000;15(1):7-12.
202. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Stat Med* 1993;12(24):2273-84.
203. Minelli C, Thompson JR, Abrams KR, Lambert PC. Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Stat Med* 2005;24(24):3845-61.
204. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.3. 2020. Available from: https://CRAN.R-project.org/package=gplots.
205. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2020. Available from: https://www.R-project.org/.
206. Bagos PG. A unification of multivariate methods for meta-analysis of genetic association studies. *Stat Appl Genet Mol Biol* 2008;7(1):31. Doi:10.2202/1544-6115.1408.
207. Langan D, Higgins JP, Simmonds M. An empirical comparison of heterogeneity variance estimators in 12 894 meta-analyses. *Res Synth Methods* 2015;6(2):195-205.
208. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007;28(2):105-14.
209. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med* 2007;26(1):78-97.
210. Cobos A, Sanchez P, Aguado J, Carrasco JL. Methodological quality in pharmacogenetic studies with binary assessment of treatment response: a review. *Pharmacogenet Genomics* 2011;21(5):243-50.
211. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7(2):e1000217. Doi:10.1371/journal.pmed.1000217.
212. Richardson M, Kirkham JJ, Dwan KM, Sloan DJ, Davies G, Jorgensen A. Protocol for the development of the STrengthening the Reporting Of Pharmacogenetic Studies (STROPS) guideline: checklist of items for reporting pharmacogenetic studies. *BMJ Open* 2019;9(7):e030212. Doi:10.1136/bmjopen-2019-030212.

213. UK EQUATOR Centre. *The EQUATOR network* 2020 [cited 2020 10 July]. Available from: https://www.equator-network.org/.
214. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ. Strengthening the reporting of Genetic RIsk Prediction Studies: the GRIPS Statement. *PLoS Med* 2011;8(3):e1000420. Doi:10.1371/journal.pmed.1000420.
215. SCImago Lab. *SJR - SCImago journal & country rank* 2020 [cited 2020 10 July]. Available from: http://www.scimagojr.com.
216. COMET Initiative. *DelphiManager* [cited 2020 1 July]. Available from: http://www.comet-initiative.org/delphimanager/.
217. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64(4):395-400.
218. Zoom Video Communications Inc. *Zoom* [cited 2020 1 July]. Available from: https://zoom.us/.
219. Turning Technologies. *TurningPoint* [cited 2020 1 July]. Available from: https://www.turningtechnologies.com/turningpoint/.
220. Thorn CF, Whirl-Carrillo M, Hachad H, Johnson JA, McDonagh EM, Ratain MJ, et al. Essential characteristics of pharmacogenomics study publications. *Clin Pharmacol Ther* 2019;105(1):86-91.
221. McDonagh EM, Whirl-Carrillo M, Altman RB, Klein TE. Enabling the curation of your pharmacogenetic study. *Clin Pharmacol Ther* 2015;97(2):116-9.
222. Chaplin M, Kirkham JJ, Dwan K, Sloan DJ, Davies G, Jorgensen AL. STrengthening the Reporting Of Pharmacogenetic Studies: Development of the STROPS guideline. *PLoS Med* 2020;17(9):e1003344. Doi:10.1371/journal.pmed.1003344.
223. Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core outcome set–STAndards for reporting: the COS-STAR statement. *PLoS Med* 2016;13(10):e1002148. Doi:10.1371/journal.pmed.1002148.
224. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;13(1):132. Doi:10.1186/745-6215-13-132.
225. Rajman I, Knapp L, Morgan T, Masimirembwa C. African genetic diversity: implications for cytochrome P450-mediated drug metabolism and drug development. *EBioMedicine* 2017;17:67-74.
226. Mpye KL, Matimba A, Dzobo K, Chirikure S, Wonkam A, Dandara C. Disease burden and the role of pharmacogenomics in African populations. *Glob Health Epidemiol Genom* 2017;2:e1. Doi:10.1017/gheg.2016.21.
227. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2197-223.
228. Dandara C, Masimirembwa C, Haffani YZ, Ogutu B, Mabuka J, Aklillu E, et al. African Pharmacogenomics Consortium: consolidating pharmacogenomics knowledge, capacity development and translation in Africa. *AAS Open Res* 2019;2:19. Doi:10.12688/aasopenres.12965.1.
229. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012;10(1):28-55.
230. Hamzic S, Kummer D, Froehlich TK, Joerger M, Aebi S, Palles C, et al. Evaluating the role of ENOSF1 and TYMS variants as predictors in fluoropyrimidine-related toxicities: an IPD meta-analysis. *Pharmacol Res* 2020;152:104594. Doi:10.1016/j.phrs.2019.104594.

231. Danese E, Raimondi S, Montagnana M, Tagetti A, Langaee T, Borgiani P, et al. Effect of CYP4F2, VKORC1, and CYP2C9 in influencing coumarin dose: a single-patient data meta-analysis in more than 15,000 individuals. *Clin Pharmacol Ther* 2019;105(6):1477-91.
232. International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009;360(8):753-64.
233. Province MA, Goetz MP, Brauch H, Flockhart DA, Hebert JM, Whaley R, et al. CYP2D6 genotype and adjuvant tamoxifen: meta-analysis of heterogeneous study populations. *Clin Pharmacol Ther* 2014;95(2):216-27.
234. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ* 2017;357:j1390. Doi:10.1136/bmj.j1390
235. Raza S, Hall A. Genomic medicine and data sharing. *Br Med Bull* 2017;123:35-45.
236. Turner RM, Newman WG, Bramon E, McNamee CJ, Wong WL, Misbah S, et al. Pharmacogenomics in the UK National Health Service: opportunities and challenges. *Pharmacogenomics* 2020;21(17):1237-46.

Appendices

Appendix 1. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Search strategy

Databases searched

Databases	Date searched	No. records retrieved
MEDLINE (Ovid) and MEDLINE In-Process (Ovid)[a]	03/03/2016	3029
EMBASE (Ovid)	03/03/2016	4778
PubMed[a,b]	03/03/2016	379
Web of science	03/03/2016	421
Biosis	03/03/2016	328

[a] Although PubMed includes MEDLINE (and PubMed is more up to date than MEDLINE), both databases were searched separately. The information specialist searched MEDLINE via Ovid first (due to the to ease of conducting effective searches of this database), and then searched PubMed for results from the last year that may have not yet been available in MEDLINE via Ovid.
[b] A date limit was applied so that only records from 03/03/2015 to 03/03/2016 were searched.

Search Strategy: EMBASE

#	Searches	Results
1	antitubercular agents/ or aminosalicylic acid/ or diarylquinolines/ or ethambutol/ or ethionamide/ or isoniazid/ or prothionamide/ or pyrazinamide/ or thioacetazone/ or antibiotics, antitubercular/ or capreomycin/ or cycloserine/ or enviomycin/ or rifabutin/ or rifampin/ or viomycin/	151901
2	((Antitubercul* or tuberculos* or TB) adj4 (agent* or drug* or antibiotic* or medicine* or medication* or treatment*)).tw.	40664
3	(aminosalicylic acid or diarylquinoline* or ethambutol* or ethionamide* or isoniazid* or prothionamide* or pyrazinamide* or thioacetazone* or capreomycin* or cycloserine* or enviomycin* or rifabutin* or rifampin* or viomycin*).tw.	34743
4	1 or 2 or 3	172588
5	Polymorphism, Genetic/	102257
6	genetic predisposition to disease/ or anticipation, genetic/	97585
7	Pharmacogenetics/	17431
8	Genetic Association Studies/	876
9	((Genetic or gene*) adj2 associat* adj2 (studies or study or analys*)).tw.	7890
10	((genetic* or gene*) adj3 (suscept* or predisposit* or anticipat*)).tw.	61544
11	Polymorphism, Single Nucleotide/	98303
12	(single* adj2 nucleotid* adj2 polymorph*).tw.	75841
13	(SNP or Genotyp* or Phenotyp* or Allele* or Pharmacogenet* or Pharmacogenom* or Polymorph*).tw.	1171894
14	((gene* or genetic*) adj5 (mutat* or variant*)).tw.	294715
15	Genotype/ or Phenotype/ or Alleles/	777386
16	or/5-15	1548879
17	exp Tuberculosis/	197008
18	(TB or Tuberculosis*).tw.	187590
19	Antitubercul*.tw.	16330
20	or/17-19	253048
21	4 and 16 and 20	5380
22	animal/ not human/	1357016
23	21 not 22	5360
24	limit 23 to em=188300-201608	4778

Appendix 2. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Full quality assessment results

Table 32 Full quality assessment results: choosing which genes and SNPs to genotype

Group identifier (if applicable) and study name	Was a literature review undertaken and the findings summarised?	Are reasons given for choosing the genes and SNPs genotyped?	If reasons include previous association studies are key details from these provided?	If reasons include functional studies are supporting data provided?
ADACS - Chen 2015	Y	Y	N	N
ADACS - Lv 2012	Y	Y	N	N/A
ADACS - Tang 2012	Y	N	N	N/A
ADACS - Tang 2013a	Y	Y	N	N/A
ADACS - Tang 2013b	Y	Y	N	N
ADACS - Wang 2015a	Y	Y	N	N
ADACS - Wang 2015c	Y	Y	N	N
An 2012	Y	Y	N	N/A
Azuma 2013	Y	Y	N	N
Bose 2011	Y	Y	N	N
Çetintaş 2008	Y	N	N	N/A
Chamorro 2013	Y	Y	N	N
Chang 2012	Y	Y	N	N
Chatterjee 2010	Y	Y	N	N/A
Cho 2007	Y	N	N	N
Costa 2012	Y	Y	N	N
Dhoro 2013	Y	N	N	N/A
Feng 2014	Y	Y	N	N
Fredj 2016	Y	Y	N	N
Gogtay 2016	Y	Y	N	N/A
GUPTA - Gupta 2013a	Y	Y	N	N
GUPTA - Gupta 2013b	Y	Y	N	N
He 2015	Y	Y	N	N
HIGUCHI - Higuchi 2007	Y	N	N	N
HIGUCHI - Nanashima 2012	Y	Y	N	N
Ho 2013	Y	Y	N	N/A
HUANG - Huang 2003	Y	Y	N	N
HUANG - Huang 2007	Y	N	N	N
Jung 2015	Y	Y	N	N/A
Khalili 2011	Y	Y	N	N/A
KIM - Kim 2009	Y	Y	N/A	N/A
KIM - Kim 2010	Y	Y	N	N/A
KIM - Kim 2011	Y	Y	N	N/A
KIM - Kim 2012a	Y	Y	N	N
KIM - Kim 2012b	Y	Y	N	N/A

Group identifier (if applicable) and study name	Was a literature review undertaken and the findings summarised?	Are reasons given for choosing the genes and SNPs genotyped?	If reasons include previous association studies are key details from these provided?	If reasons include functional studies are supporting data provided?
KIM - Kim 2012c	Y	Y	N	N
KIM - Kim 2015	Y	Y	N	N/A
Kwon 2012	Y	Y	N/A	N
Lee 2010	Y	Y	N	N
LEIRO - Leiro 2008	Y	Y	N	N
LEIRO - Leiro-Fernandez 2011	Y	Y	N	N/A
Li 2012	Y	N	N	N
Liu 2014	Y	Y	N	N/A
Mahmoud 2012	Y	Y	N	N
Monteiro 2012	Y	Y	N	N
Ng 2014	Y	Y	N	N
NTUH - Wang 2011	Y	N	N	N
NTUH - Wang 2015b	Y	Y	N	N
Ohno 2000	Y	Y	N	N/A
POSSUELO - Brito 2014	Y	N	N	N/A
POSSUELO - Possuelo 2008	Y	Y	N	N
Rana 2014	Y	N	N	N
Roy 2001	Y	N	N	N
Roy 2006	Y	Y	N	N/A
SANTOS - Fernandes 2015	Y	N	N	N
SANTOS - Santos 2013	Y	N	N	N/A
Sharma 2014	Y	Y	N	N/A
Shimizu 2006	Y	Y	N	N/A
Singla 2014	Y	N	N	N
Sotsuka 2011	Y	Y	N	N
Teixeira 2011	Y	Y	N	N
Vuilleumier 2006	Y	N	N	N
Wang 2010	Y	N	N	N
Xiang 2014	Y	N	N	N
YAMADA - Yamada 2009	Y	Y	N	N
YAMADA - Yamada 2010	Y	Y	N/A	N
Yimer 2011	Y	N	N	N
Yuliwulandari 2016	Y	N	N	N
Zaverucha-do-Valle 2014	Y	N	N	N/A
Zazuli 2015	Y	N	N	N

N: no; N/A: not applicable; SNP: single nucleotide polymorphism; Y: yes

Table 33 Full quality assessment results: sample size

Group identifier (if applicable) and study name	What is the sample size?	Are details given of how the sample size was calculated?	Are details given of the *a priori* power to detect effect sizes of varying degrees, or a justified specific effect size?
ADACS - Chen 2015	445	N	N
ADACS - Lv 2012	445	N	N
ADACS - Tang 2012	445	N	N
ADACS - Tang 2013a	445	N	N
ADACS - Tang 2013b	445	N	N
ADACS - Wang 2015a	445	N	N
ADACS - Wang 2015c	445	N	N
An 2012	208	N	N
Azuma 2013	121	Y	Y
Bose 2011	218	N	N
Çetintaş 2008	100	N	N
Chamorro 2013	175	N	N
Chang 2012	98	N	N
Chatterjee 2010	151	N	N
Cho 2007	132	N	N
Costa 2012	129	N	N
Dhoro 2013	Unclear	N	N
Feng 2014	346	N	N
Fredj 2016	71	N	N
Gogtay 2016	75	N	N
GUPTA - Gupta 2013a	296	N	N
GUPTA - Gupta 2013b	215	N	N
He 2015	254	Y	N
HIGUCHI - Higuchi 2007	100	N	N
HIGUCHI - Nanashima 2012	100	N	N
Ho 2013	348	N	N
HUANG - Huang 2003	318	N	N
HUANG - Huang 2007	230	N	N
Jung 2015	153	N	N
Khalili 2011	100	N	N
KIM - Kim 2009	226	N	N
KIM - Kim 2010	341	N	N
KIM - Kim 2011	221	N	N
KIM - Kim 2012a	221	N	N
KIM - Kim 2012b	226	N	N
KIM - Kim 2012c	306	N	N
KIM - Kim 2015	321	N	N
Kwon 2012	238	N	N
Lee 2010	140	N	N
LEIRO - Leiro 2008	95	N	N

Group identifier (if applicable) and study name	What is the sample size?	Are details given of how the sample size was calculated?	Are details given of the *a priori* power to detect effect sizes of varying degrees, or a justified specific effect size?
LEIRO - Leiro-Fernandez 2011	117	Y	N
Li 2012	273	N	N
Liu 2014	163	Y	Y
Mahmoud 2012	66	N	N
Monteiro 2012	177	N	N
Ng 2014	127	N	N
NTUH - Wang 2011	360	N	N
NTUH - Wang 2015b	355	N	N
Ohno 2000	77	N	N
POSSUELO - Brito 2014	245	N	N
POSSUELO - Possuelo 2008	254	N	N
Rana 2014	300	N	N
Roy 2001	66	N	N
Roy 2006	109	N	N
SANTOS - Fernandes 2015	220	N	N
SANTOS - Santos 2013	270	N	N
Sharma 2014	314	N	N
Shimizu 2006	42	N	N
Singla 2014	408	N	N
Sotsuka 2011	144	N	N
Teixeira 2011	167	N	N
Vuilleumier 2006	89	N	N
Wang 2010	215	N	N
Xiang 2014	2244	N	N
YAMADA - Yamada 2009	170	N	N
YAMADA - Yamada 2010	170	N	N
Yimer 2011	353	N	N
Yuliwulandari 2016	241	N	N
Zaverucha-do-Valle 2014	131	Y	Y
Zazuli 2015	106	N	N

N: no; Y: yes

Table 34 Full quality assessment results: study design

Group identifier (if applicable) and study name	What is the study design?	If study is case-control are the two groups clearly defined?	If study is case-control were they genotyped in mixed batches?
ADACS - Chen 2015	Case-control	Y	Unclear
ADACS - Lv 2012	Case-control	Y	Unclear
ADACS - Tang 2012	Case-control	Y	Unclear
ADACS - Tang 2013a	Case-control	Y	Unclear
ADACS - Tang 2013b	Case-control	Y	Unclear
ADACS - Wang 2015a	Case-control	Y	Unclear
ADACS - Wang 2015c	Case-control	Y	Unclear
An 2012	Case-control	Y	Unclear
Azuma 2013	RCT	N/A	N/A
Bose 2011	Cohort	N/A	N/A
Çetintaş 2008	Cohort	N/A	N/A
Chamorro 2013	Cohort	N/A	N/A
Chang 2012	Cohort	N/A	N/A
Chatterjee 2010	Case-control	Y	Unclear
Cho 2007	Cohort	N/A	N/A
Costa 2012	Cohort	N/A	N/A
Dhoro 2013	Case-control	N	Unclear
Feng 2014	Case-control	Y	Unclear
Fredj 2016	Cohort	N/A	N/A
Gogtay 2016	Case-control	Y	Unclear
GUPTA - Gupta 2013a	Cohort	N/A	N/A
GUPTA - Gupta 2013b	Cohort	N/A	N/A
He 2015	Case-control	Y	Y
HIGUCHI - Higuchi 2007	Cohort	N/A	N/A
HIGUCHI - Nanashima 2012	Cohort	N/A	N/A
Ho 2013	Cohort	N/A	N/A
HUANG - Huang 2003	Cohort	N/A	N/A
HUANG - Huang 2007	Case-control	Y	Unclear
Jung 2015	Cohort	N/A	N/A
Khalili 2011	Case-control	Y	Unclear
KIM - Kim 2009	Case-control	Y	Unclear
KIM - Kim 2010	Case-control	Y	Unclear
KIM - Kim 2011	Case-control	Y	Unclear
KIM - Kim 2012a	Case-control	Y	Unclear
KIM - Kim 2012b	Case-control	Y	Unclear
KIM - Kim 2012c	Case-control	Y	Unclear
KIM - Kim 2015	Case-control	Y	Unclear
Kwon 2012	Case-control	Y	Unclear
Lee 2010	Cohort	N/A	N/A
LEIRO - Leiro 2008	Case-control	Y	Unclear

Group identifier (if applicable) and study name	What is the study design?	If study is case-control are the two groups clearly defined?	If study is case-control were they genotyped in mixed batches?
LEIRO - Leiro-Fernandez 2011	Case-control	Y	Unclear
Li 2012	Case-control	N	Unclear
Liu 2014	Case-control	Y	Unclear
Mahmoud 2012	Cohort	N/A	N/A
Monteiro 2012	Cohort	N/A	N/A
Ng 2014	Case-control	Y	Unclear
NTUH - Wang 2011	Cohort	N/A	N/A
NTUH - Wang 2015b	Cohort	N/A	N/A
Ohno 2000	Cohort	N/A	N/A
POSSUELO - Brito 2014	Cohort	N/A	N/A
POSSUELO - Possuelo 2008	Cohort	N/A	N/A
Rana 2014	Cohort	N/A	N/A
Roy 2001	Case-control	Y	Unclear
Roy 2006	Cohort	N/A	N/A
SANTOS - Fernandes 2015	Cohort	N/A	N/A
SANTOS - Santos 2013	Cohort	N/A	N/A
Sharma 2014	Case-control	Y	Unclear
Shimizu 2006	Cohort	N/A	N/A
Singla 2014	Cohort	N/A	N/A
Sotsuka 2011	Cohort	N/A	N/A
Teixeira 2011	Case-control	Y	Unclear
Vuilleumier 2006	Cohort	N/A	N/A
Wang 2010	Case-control	Y	Unclear
Xiang 2014	Cohort	N/A	N/A
YAMADA - Yamada 2009	Cohort	N/A	N/A
YAMADA - Yamada 2010	Cohort	N/A	N/A
Yimer 2011	Cohort	N/A	N/A
Yuliwulandari 2016	Case-control	Y	Unclear
Zaverucha-do-Valle 2014	Cohort	N/A	N/A
Zazuli 2015	Cohort	N/A	N/A

N: no; N/A: not applicable; RCT: randomised controlled trial; Y: yes

Table 35 Full quality assessment results: reliability of genotypes

Group identifier (if applicable) and study name	Is the genotyping procedure described?	Are the primers described?	Were quality control methods used and described?	Were findings from quality control methods, if used, described?	Are any genotype frequencies previously reported quoted?	Were genotyping personnel blinded to outcome status?	If human inference required was this independently undertaken by at least two people?
ADACS - Chen 2015	Y	N	Y	Y	N	Y	Not mentioned
ADACS - Lv 2012	Y	Y	Y	N	Y - for a subset of investigated genetic variants only	Y	Y
ADACS - Tang 2012	Y	Y	Y	Y	N	Y	Y
ADACS - Tang 2013a	Y	Y	Y	Y	N	Y	Not mentioned
ADACS - Tang 2013b	Y	Y	Y	Y	N	Y	Not mentioned
ADACS - Wang 2015a	Y	Y	Y	Y	N	Y	Y
ADACS - Wang 2015c	Y	Y	Y	Y	Y - for a subset of investigated genetic variants only	Y	Y
An 2012	Y	Y	Not mentioned	N/A	N	Not mentioned	NK if applicable
Azuma 2013	Y	N	Y	N	Y	Y	NK if applicable
Bose 2011	Y	Y	Not mentioned	N/A	N	Y	NK if applicable
Çetintaş 2008	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	N	Y	NK if applicable
Chamorro 2013	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Chang 2012	Y	N	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Chatterjee 2010	Y	Y	Not mentioned	N/A	Y	Not mentioned	NK if applicable

Group identifier (if applicable) and study name	Is the genotyping procedure described?	Are the primers described?	Were quality control methods used and described?	Were findings from quality control methods, if used, described?	Are any genotype frequencies previously reported quoted?	Were genotyping personnel blinded to outcome status?	If human inference required was this independently undertaken by at least two people?
Cho 2007	Y	Available upon request	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Costa 2012	Y	Y - for a subset of genes (primers may be inherent to the assay used for other genes)	Y	N	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Dhoro 2013	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	Y	Not mentioned	NK if applicable
Feng 2014	Y	N	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable
Fredj 2016	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable
Gogtay 2016	Y	Y	Not mentioned	N/A	Y	Not mentioned	NK if applicable
GUPTA - Gupta 2013a	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	N	Y	NK if applicable
GUPTA - Gupta 2013b	Y	Y - for a subset of genes	Not mentioned	N/A	N	Y	NK if applicable
He 2015	Y	Y	Y	N	N	Y	Not mentioned
HIGUCHI - Higuchi 2007	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable
HIGUCHI - Nanashima 2012	Y	Y	Not mentioned	N/A	N	Y	NK if applicable
Ho 2013	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable

Group identifier (if applicable) and study name	Is the genotyping procedure described?	Are the primers described?	Were quality control methods used and described?	Were findings from quality control methods, if used, described?	Are any genotype frequencies previously reported quoted?	Were genotyping personnel blinded to outcome status?	If human inference required was this independently undertaken by at least two people?
HUANG - Huang 2003	Y	Y - for a subset of genes	Not mentioned	N/A	N	Y	NK if applicable
HUANG - Huang 2007	Y	Y	Y - for a subset of investigated genetic variants only	N	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable
Jung 2015	Y	N	Not mentioned	N/A	N	Y	NK if applicable
Khalili 2011	Y	Y	Not mentioned	N/A	Y	Not mentioned	NK if applicable
KIM - Kim 2009	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable
KIM - Kim 2010	Y	Y	Not mentioned	N/A	Y	Not mentioned	NK if applicable
KIM - Kim 2011	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable
KIM - Kim 2012a	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	N	Not mentioned	NK if applicable
KIM - Kim 2012b	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	Y	Not mentioned	NK if applicable
KIM - Kim 2012c	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	N	Y	N/A
KIM - Kim 2015	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable

Group identifier (if applicable) and study name	Is the genotyping procedure described?	Are the primers described?	Were quality control methods used and described?	Were findings from quality control methods, if used, described?	Are any genotype frequencies previously reported quoted?	Were genotyping personnel blinded to outcome status?	If human inference required was this independently undertaken by at least two people?
Kwon 2012	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	N	Not mentioned	NK if applicable
Lee 2010	Y	Y	Not mentioned	N/A	N	Y	NK if applicable
LEIRO - Leiro 2008	Y	Y	Not mentioned	N/A	N	Not mentioned	NK if applicable
LEIRO - Leiro-Fernandez 2011	Y	Y	Not mentioned	N/A	Y	Y	Not mentioned
Li 2012	Y	N - primers may be inherent to the assay used)	Y	N	Y - but only for haplotypes (not individual SNPs)	Not mentioned	NK if applicable
Liu 2014	Y	Y	Not mentioned	N/A	N	Not mentioned	NK if applicable
Mahmoud 2012	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Monteiro 2012	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable
Ng 2014	Y	Y	Not mentioned	N/A	N	Not mentioned	NK if applicable
NTUH - Wang 2011	Y	Y	Not mentioned	N/A	N	Y	NK if applicable
NTUH - Wang 2015b	Y	N	Y - for a subset of investigated genetic variants only	Y	N	Y	NK if applicable
Ohno 2000	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable
POSSUELO - Brito 2014	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
POSSUELO - Possuelo 2008	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable

Group identifier (if applicable) and study name	Is the genotyping procedure described?	Are the primers described?	Were quality control methods used and described?	Were findings from quality control methods, if used, described?	Are any genotype frequencies previously reported quoted?	Were genotyping personnel blinded to outcome status?	If human inference required was this independently undertaken by at least two people?
Rana 2014	Y	N	Not mentioned	N/A	Y	Y	NK if applicable
Roy 2001	Y	N	Not mentioned	N/A	N	Not mentioned	NK if applicable
Roy 2006	N	N	Not mentioned	N/A	N	Y	NK if applicable
SANTOS - Fernandes 2015	Y	N - primers may be inherent to the assay used)	Not mentioned	N/A	N	Y	NK if applicable
SANTOS - Santos 2013	Y	Y	Not mentioned	N/A	Y	Y	NK if applicable
Sharma 2014	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable
Shimizu 2006	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Singla 2014	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Sotsuka 2011	Y	N	Not mentioned	N/A	Y	Y	NK if applicable
Teixeira 2011	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Not mentioned	NK if applicable
Vuilleumier 2006	Y	Y - for a subset of genes (primers may be inherent to the assay used for other genes)	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
Wang 2010	Y	Y	Not mentioned	N/A	N	Not mentioned	NK if applicable

Group identifier (if applicable) and study name	Is the genotyping procedure described?	Are the primers described?	Were quality control methods used and described?	Were findings from quality control methods, if used, described?	Are any genotype frequencies previously reported quoted?	Were genotyping personnel blinded to outcome status?	If human inference required was this independently undertaken by at least two people?
Xiang 2014	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	NK if applicable
YAMADA - Yamada 2009	Y	Y	Not mentioned	N/A	Y - for a subset of investigated genetic variants only	Y	Y
YAMADA - Yamada 2010	Y	Y	Y	Y	N	Y	Y
Yimer 2011	Y	Y - for a subset of genes (primers may be inherent to the assay used for other genes)	Not mentioned	N/A	N	Y	NK if applicable
Yuliwulandari 2016	Y	Y	Not mentioned	N/A	N	Not mentioned	NK if applicable
Zaverucha-do-Valle 2014	Y	Y	Not mentioned	N/A	N	Y	NK if applicable
Zazuli 2015	Y	Y	Not mentioned	N/A	N	Y	NK if applicable

N: no; N/A: not applicable; Y: yes

Table 36 Full quality assessment results: missing genotype data

Group identifier (if applicable) and study name	Is extent of missing data summarised?	If yes, are reasons for missing data given?	If yes, are checks for missingness at random reported?	Are missing genotype data imputed?	Does paper quote number of patients contributing to each analysis?	If paper does quote number of patients contributing to analyses, does this agree to sample size?
ADACS - Chen 2015	N	N/A	N/A	N	Y (we spotted errors in the reported numbers of patients contributing to each analysis, and emailed authors to obtain corrected data)	N
ADACS - Lv 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
ADACS - Tang 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
ADACS - Tang 2013a	N	N/A	N/A	N	Y	N
ADACS - Tang 2013b	N	N/A	N/A	N	Y	N
ADACS - Wang 2015a	N	N/A	N/A	N	Y	N
ADACS - Wang 2015c	N	N/A	N/A	N	Y	N
An 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Azuma 2013	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Bose 2011	N	N/A	N/A	N	N	N/A
Çetintaş 2008	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Chamorro 2013	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Chang 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Chatterjee 2010	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Cho 2007	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y

Group identifier (if applicable) and study name	Is extent of missing data summarised?	If yes, are reasons for missing data given?	If yes, are checks for missingness at random reported?	Are missing genotype data imputed?	Does paper quote number of patients contributing to each analysis?	If paper does quote number of patients contributing to analyses, does this agree to sample size?
Costa 2012	Y	N	N	N	Y	N
Dhoro 2013	N	N/A	N/A	N	Y	N
Feng 2014	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Fredj 2016	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Gogtay 2016	Y	Y	N	N	Y	N
GUPTA - Gupta 2013a	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
GUPTA - Gupta 2013b	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
He 2015	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
HIGUCHI - Higuchi 2007	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
HIGUCHI - Nanashima 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Ho 2013	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
HUANG - Huang 2003	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
HUANG - Huang 2007	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Jung 2015	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Khalili 2011	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
KIM - Kim 2009	N	N/A	N/A	N	Y	N
KIM - Kim 2010	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
KIM - Kim 2011	N	N/A	N/A	N	Y	N
KIM - Kim 2012a	N	N/A	N/A	N	Y	N
KIM - Kim 2012b	N	N/A	N/A	N	Y	N
KIM - Kim 2012c	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y

Group identifier (if applicable) and study name	Is extent of missing data summarised?	If yes, are reasons for missing data given?	If yes, are checks for missingness at random reported?	Are missing genotype data imputed?	Does paper quote number of patients contributing to each analysis?	If paper does quote number of patients contributing to analyses, does this agree to sample size?
KIM - Kim 2015	N	N/A	N/A	N	Y	N
Kwon 2012	N	N/A	N/A	N	N	N/A
Lee 2010	N	N/A	N/A	N	Y	N
LEIRO - Leiro 2008	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
LEIRO - Leiro-Fernandez 2011	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Li 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Liu 2014	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Mahmoud 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Monteiro 2012	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Ng 2014	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
NTUH - Wang 2011	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
NTUH - Wang 2015b	Y	Y	N	N	N	N/A
Ohno 2000	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
POSSUELO - Brito 2014	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
POSSUELO - Possuelo 2008	N	N/A	N/A	N	N	N/A
Rana 2014	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Roy 2001	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Roy 2006	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
SANTOS - Fernandes 2015	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
SANTOS - Santos 2013	N	N/A	N/A	N	Y	N
Sharma 2014	Y	Y	N	N	Y	N

Group identifier (if applicable) and study name	Is extent of missing data summarised?	If yes, are reasons for missing data given?	If yes, are checks for missingness at random reported?	Are missing genotype data imputed?	Does paper quote number of patients contributing to each analysis?	If paper does quote number of patients contributing to analyses, does this agree to sample size?
Shimizu 2006	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Singla 2014	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Sotsuka 2011	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Teixeira 2011	Y	Y	N	N	Y	N
Vuilleumier 2006	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Wang 2010	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Xiang 2014	Y - for a subset of investigated genetic variants only	N	N	N	Y	N
YAMADA - Yamada 2009	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
YAMADA - Yamada 2010	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Yimer 2011	Y	N	N	N	Y	N
Yuliwulandari 2016	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y
Zaverucha-do-Valle 2014	Y	N	N	N	Y	N
Zazuli 2015	Y (no missing data)	N/A (no missing data)	N/A (no missing data)	N/A (no missing data)	Y	Y

N: no; N/A: not applicable; Y: yes

Table 37 Full quality assessment results: population stratification

Group identifier (if applicable) and study name	Are tests undertaken for cryptic population stratification?	If so, are results quoted?
ADACS - Chen 2015	N	N/A
ADACS - Lv 2012	N	N/A
ADACS - Tang 2012	N	N/A
ADACS - Tang 2013a	N	N/A
ADACS - Tang 2013b	N	N/A
ADACS - Wang 2015a	N	N/A
ADACS - Wang 2015c	N	N/A
An 2012	N	N/A
Azuma 2013	N	N/A
Bose 2011	N	N/A
Çetintaş 2008	N	N/A
Chamorro 2013	N	N/A
Chang 2012	N	N/A
Chatterjee 2010	N	N/A
Cho 2007	N	N/A
Costa 2012	N	N/A
Dhoro 2013	N	N/A
Feng 2014	N	N/A
Fredj 2016	N	N/A
Gogtay 2016	N	N/A
GUPTA - Gupta 2013a	N	N/A
GUPTA - Gupta 2013b	N	N/A
He 2015	N	N/A
HIGUCHI - Higuchi 2007	N	N/A
HIGUCHI - Nanashima 2012	N	N/A
Ho 2013	N	N/A
HUANG - Huang 2003	N	N/A
HUANG - Huang 2007	N	N/A
Jung 2015	N	N/A
Khalili 2011	N	N/A
KIM - Kim 2009	N	N/A
KIM - Kim 2010	N	N/A
KIM - Kim 2011	N	N/A
KIM - Kim 2012a	N	N/A
KIM - Kim 2012b	N	N/A
KIM - Kim 2012c	N	N/A
KIM - Kim 2015	N	N/A
Kwon 2012	N	N/A
Lee 2010	N	N/A
LEIRO - Leiro 2008	N	N/A
LEIRO - Leiro-Fernandez 2011	N	N/A
Li 2012	N	N/A

Group identifier (if applicable) and study name	Are tests undertaken for cryptic population stratification?	If so, are results quoted?
Liu 2014	N	N/A
Mahmoud 2012	N	N/A
Monteiro 2012	N	N/A
Ng 2014	N	N/A
NTUH - Wang 2011	N	N/A
NTUH - Wang 2015b	N	N/A
Ohno 2000	N	N/A
POSSUELO - Brito 2014	N	N/A
POSSUELO - Possuelo 2008	N	N/A
Rana 2014	N	N/A
Roy 2001	N	N/A
Roy 2006	N	N/A
SANTOS - Fernandes 2015	Y	Y
SANTOS - Santos 2013	Y	Y
Sharma 2014	N	N/A
Shimizu 2006	N	N/A
Singla 2014	N	N
Sotsuka 2011	N	N/A
Teixeira 2011	N	N/A
Vuilleumier 2006	N	N/A
Wang 2010	N	N/A
Xiang 2014	N	N/A
YAMADA - Yamada 2009	N	N/A
YAMADA - Yamada 2010	N	N/A
Yimer 2011	N	N/A
Yuliwulandari 2016	N	N/A
Zaverucha-do-Valle 2014	N	N/A
Zazuli 2015	N	N/A

N: no; N/A: not applicable; Y: yes

Table 38 Full quality assessment results: Hardy-Weinberg equilibrium

Group identifier (if applicable) and study name	What test is undertaken to check for HWE?	Where test undertaken, is *p* value threshold applied to determine deviation from HWE quoted?	Where test undertaken, are SNPs deviating from HWE highlighted?	Where test undertaken, and some SNPs found to deviate, are steps taken to explore deviation from HWE reported?	Where test undertaken, and some SNPs found to deviate, are deviating SNPs excluded from further analysis?
ADACS - Chen 2015	Chi square test	Y	Y (none deviated)	N/A	N/A
ADACS - Lv 2012	Chi square test	Y	Y (none deviated)	N/A	N/A
ADACS - Tang 2012	Chi square test	Y	N	N/A	N/A
ADACS - Tang 2013a	Chi square test	Y	Y	N	N
ADACS - Tang 2013b	Chi square test	Y	Y (none deviated)	N/A	N/A
ADACS - Wang 2015a	Chi square test	Y	Y (none deviated)	N/A	N/A
ADACS - Wang 2015c	Chi square test	Y	Y (none deviated)	N/A	N/A
An 2012	Test not described, test only mentioned for subset of SNPs	Y	Y (none deviated)	N/A	N/A
Azuma 2013	No test mentioned	N/A	N/A	N/A	N/A
Bose 2011	No test mentioned	N/A	N/A	N/A	N/A
Çetintaş 2008	No test mentioned	N/A	N/A	N/A	N/A
Chamorro 2013	No test mentioned	N/A	N/A	N/A	N/A
Chang 2012	Test not described	N	Y (none deviated)	N/A	N/A
Chatterjee 2010	No test mentioned	N/A	N/A	N/A	N/A
Cho 2007	No test mentioned	N/A	N/A	N/A	N/A
Costa 2012	Test not described	Y	Y	N	Y
Dhoro 2013	No test mentioned	N/A	N/A	N/A	N/A
Feng 2014	Goodness of fit test	Y	Y (none deviated)	N/A	N/A
Fredj 2016	Test not described	N	N	N/A	N/A
Gogtay 2016	No test mentioned	N/A	N/A	N/A	N/A
GUPTA - Gupta 2013a	No test mentioned	N/A	N/A	N/A	N/A

Group identifier (if applicable) and study name	What test is undertaken to check for HWE?	Where test undertaken, is *p* value threshold applied to determine deviation from HWE quoted?	Where test undertaken, are SNPs deviating from HWE highlighted?	Where test undertaken, and some SNPs found to deviate, are steps taken to explore deviation from HWE reported?	Where test undertaken, and some SNPs found to deviate, are deviating SNPs excluded from further analysis?
GUPTA - Gupta 2013b	Chi square test	Y	N	N/A	N/A
He 2015	No test mentioned	N/A	N/A	N/A	N/A
HIGUCHI - Higuchi 2007	Chi square test	Y	N	N/A	N/A
HIGUCHI - Nanashima 2012	Chi square test	Y	Y (none deviated)	N/A	N/A
Ho 2013	Chi square test	N	N	N/A	N/A
HUANG - Huang 2003	Chi square test	Y	Y (none deviated)	N/A	N/A
HUANG - Huang 2007	Chi square test	Y	Y (none deviated)	N/A	N/A
Jung 2015	No test mentioned	N/A	N/A	N/A	N/A
Khalili 2011	Chi square test	Y	Y (none deviated)	N/A	N/A
KIM - Kim 2009	Chi square test	Y	Y	N	N
KIM - Kim 2010	No test mentioned	N/A	N/A	N/A	N/A
KIM - Kim 2011	Chi square test	Y	N	N/A	N/A
KIM - Kim 2012a	Chi square test	Y	N	N/A	N/A
KIM - Kim 2012b	Chi square test	Y	Y (none deviated)	N/A	N/A
KIM - Kim 2012c	No test mentioned	N/A	N/A	N/A	N/A
KIM - Kim 2015	Chi square test	Y	N	N/A	N/A
Kwon 2012	Chi square test	Y	N	N/A	N/A
Lee 2010	No test mentioned	N/A	N/A	N/A	N/A
LEIRO - Leiro 2008	No test mentioned	N/A	N/A	N/A	N/A
LEIRO - Leiro-Fernandez 2011	No test mentioned	N/A	N/A	N/A	N/A
Li 2012	HWE was tested using the random permutation procedure implemented in the Arlequin package	Y	Y (none deviated)	N/A	N/A
Liu 2014	No test mentioned	N/A	N/A	N/A	N/A

Group identifier (if applicable) and study name	What test is undertaken to check for HWE?	Where test undertaken, is *p* value threshold applied to determine deviation from HWE quoted?	Where test undertaken, are SNPs deviating from HWE highlighted?	Where test undertaken, and some SNPs found to deviate, are steps taken to explore deviation from HWE reported?	Where test undertaken, and some SNPs found to deviate, are deviating SNPs excluded from further analysis?
Mahmoud 2012	Test not described	Y	N	N/A	N/A
Monteiro 2012	No test mentioned	N/A	N/A	N/A	N/A
Ng 2014	Chi square test	N	Y (none deviated)	N/A	N/A
NTUH - Wang 2011	No test mentioned	N/A	N/A	N/A	N/A
NTUH - Wang 2015b	Test not described	Y	Y	N	N
Ohno 2000	Test not described	N	N	N/A	N/A
POSSUELO - Brito 2014	No test mentioned	N/A	N/A	N/A	N/A
POSSUELO - Possuelo 2008	No test mentioned	N/A	N/A	N/A	N/A
Rana 2014	Test not described, test only mentioned for subset of SNPs	N	N	N/A	N/A
Roy 2001	No test mentioned	N/A	N/A	N/A	N/A
Roy 2006	No test mentioned	N/A	N/A	N/A	N/A
SANTOS - Fernandes 2015	Test not described	Y	Y	N	N
SANTOS - Santos 2013	Exact test	Y	Y (none deviated)	N/A	N/A
Sharma 2014	No test mentioned	N/A	N/A	N/A	N/A
Shimizu 2006	No test mentioned	N/A	N/A	N/A	N/A
Singla 2014	No test mentioned	N/A	N/A	N/A	N/A
Sotsuka 2011	No test mentioned	N/A	N/A	N/A	N/A
Teixeira 2011	Chi square test (only mentioned for a subset of SNPs)	Y	N	N/A	N/A
Vuilleumier 2006	Test not described	N	N	N/A	N/A
Wang 2010	No test mentioned	N/A	N/A	N/A	N/A

Group identifier (if applicable) and study name	What test is undertaken to check for HWE?	Where test undertaken, is *p* value threshold applied to determine deviation from HWE quoted?	Where test undertaken, are SNPs deviating from HWE highlighted?	Where test undertaken, and some SNPs found to deviate, are steps taken to explore deviation from HWE reported?	Where test undertaken, and some SNPs found to deviate, are deviating SNPs excluded from further analysis?
Xiang 2014	Chi square test	N	Y (none deviated)	N/A	N/A
YAMADA - Yamada 2009	Fisher's exact test	N	Y (none deviated)	N/A	N/A
YAMADA - Yamada 2010	Fisher's exact test	Y	Y	N	Y
Yimer 2011	Chi square test	N	N	N/A	N/A
Yuliwulandari 2016	Chi square test	Y	Y (none deviated)	N/A	N/A
Zaverucha-do-Valle 2014	Chi square test	Y	N	N/A	N/A
Zazuli 2015	Chi square test	N	Y (none deviated)	N/A	N/A

HWE: Hardy-Weinberg equilibrium; N: no; N/A: not applicable; SNP: single nucleotide polymorphism; Y: yes

Table 39 Full quality assessment results: mode of inheritance

Group identifier (if applicable) and study name	Is a specific mode of inheritance assumed? If so which?	Is justification provided for assumptions made regarding mode of inheritance?
ADACS - Chen 2015	More than one mode assumed	N
ADACS - Lv 2012	None	N
ADACS - Tang 2012	More than one mode assumed	N
ADACS - Tang 2013a	More than one mode assumed	N
ADACS - Tang 2013b	More than one mode assumed	N
ADACS - Wang 2015a	More than one mode assumed	N
ADACS - Wang 2015c	More than one mode assumed	N
An 2012	Different modes for different SNPs	N
Azuma 2013	N/A - RCT not association study	N/A
Bose 2011	Unclear	N/A
Çetintaş 2008	Dominant	N
Chamorro 2013	Dominant	Y
Chang 2012	N/A - no homozygous MT pts	N/A
Chatterjee 2010	Dominant	N
Cho 2007	Unclear	N/A
Costa 2012	Different modes for different SNPs	N
Dhoro 2013	Dominant	N
Feng 2014	Dominant	N
Fredj 2016	Dominant	N
Gogtay 2016	Unclear	N/A
GUPTA - Gupta 2013a	Dominant	Y
GUPTA - Gupta 2013b	Unclear	N/A
He 2015	Dominant	N
HIGUCHI - Higuchi 2007	Unclear	N/A
HIGUCHI - Nanashima 2012	More than one mode assumed	N
Ho 2013	None	N
HUANG - Huang 2003	More than one mode assumed	N
HUANG - Huang 2007	Unclear	N/A
Jung 2015	Additive	N
Khalili 2011	None	N
KIM - Kim 2009	More than one mode assumed	N
KIM - Kim 2010	Dominant	N
KIM - Kim 2011	Dominant	N
KIM - Kim 2012a	More than one mode assumed	N
KIM - Kim 2012b	More than one mode assumed	N
KIM - Kim 2012c	Dominant	N
KIM - Kim 2015	Dominant	N
Kwon 2012	Unclear	N/A
Lee 2010	Dominant	N
LEIRO - Leiro 2008	Dominant	N

Group identifier (if applicable) and study name	Is a specific mode of inheritance assumed? If so which?	Is justification provided for assumptions made regarding mode of inheritance?
LEIRO - Leiro-Fernandez 2011	None	N
Li 2012	None	N
Liu 2014	None	N
Mahmoud 2012	More than one mode assumed	N
Monteiro 2012	Dominant	N
Ng 2014	Dominant	N
NTUH - Wang 2011	Dominant	N
NTUH - Wang 2015b	More than one mode assumed	N
Ohno 2000	None	N
POSSUELO - Brito 2014	Dominant	N
POSSUELO - Possuelo 2008	More than one mode assumed	N
Rana 2014	None	N
Roy 2001	Dominant	Y
Roy 2006	Dominant	N
SANTOS - Fernandes 2015	Unclear	N/A
SANTOS - Santos 2013	Different modes for different SNPs	N
Sharma 2014	Different modes for different SNPs	Y
Shimizu 2006	None	N
Singla 2014	More than one mode assumed	N
Sotsuka 2011	Different modes for different SNPs	N
Teixeira 2011	Different modes for different SNPs	N
Vuilleumier 2006	Different modes for different SNPs	N
Wang 2010	Different modes for different SNPs	N
Xiang 2014	Different modes for different SNPs	Y
YAMADA - Yamada 2009	Different modes for different SNPs	N
YAMADA - Yamada 2010	Different modes for different SNPs	N
Yimer 2011	Different modes for different SNPs	N
Yuliwulandari 2016	More than one mode assumed	N
Zaverucha-do-Valle 2014	Different modes for different SNPs	N
Zazuli 2015	Different modes for different SNPs	N

N: no; N/A: not applicable; RCT: randomised controlled trial; SNP: single nucleotide polymorphism; Y: yes

Table 40 Full quality assessment results: choice and definition of outcomes

Group identifier (if applicable) and study name	Does the paper clearly define all outcomes investigated?	Is justification provided for the choice of outcomes?	Are results shown for all outcomes mentioned for all variants investigated in the paper?
ADACS - Chen 2015	Y	N	Y
ADACS - Lv 2012	Y	Y	Y
ADACS - Tang 2012	Y	Y	Y
ADACS - Tang 2013a	Y	Y	Y
ADACS - Tang 2013b	Y	Y	Y
ADACS - Wang 2015a	Y	Y	Y
ADACS - Wang 2015c	Y	Y	Y
An 2012	Y	Y	Y
Azuma 2013	N	Y	N
Bose 2011	N	N	Y
Çetintaş 2008	Y	Y	Y
Chamorro 2013	Y	Y	Y
Chang 2012	Y	Y	Y
Chatterjee 2010	Y	Y	Y
Cho 2007	Y	Y	Y
Costa 2012	Y	N	N
Dhoro 2013	N	Y	N (but the authors provided reasons to measure outcomes only for some variants)
Feng 2014	Y	Y	Y
Fredj 2016	Y	Y	Y
Gogtay 2016	Y	Y	Y
GUPTA - Gupta 2013a	Y	Y	Y
GUPTA - Gupta 2013b	Y	Y	Y
He 2015	Y	Y	Y
HIGUCHI - Higuchi 2007	N	Y	Y
HIGUCHI - Nanashima 2012	Y	Y	Y
Ho 2013	Y	N	N
HUANG - Huang 2003	Y	Y	N (but the authors provided reasons to measure outcomes only for some variants)
HUANG - Huang 2007	Y	N	N (but the authors provided reasons to measure outcomes only for some variants)
Jung 2015	Y	Y	Y
Khalili 2011	Y	Y	Y
KIM - Kim 2009	Y	Y	N (but the authors provided reasons to measure outcomes only for some variants)

Group identifier (if applicable) and study name	Does the paper clearly define all outcomes investigated?	Is justification provided for the choice of outcomes?	Are results shown for all outcomes mentioned for all variants investigated in the paper?
KIM - Kim 2010	Y	Y	Y
KIM - Kim 2011	Y	Y	Y
KIM - Kim 2012a	Y	Y	Y
KIM - Kim 2012b	Y	Y	Y
KIM - Kim 2012c	Y	Y	Y
KIM - Kim 2015	Y	Y	Y
Kwon 2012	Y	Y	Y
Lee 2010	Y	Y	N
LEIRO - Leiro 2008	Y	Y	Y
LEIRO - Leiro-Fernandez 2011	Y	Y	Y
Li 2012	Y	Y	Y
Liu 2014	Y	Y	Y
Mahmoud 2012	Y	Y	Y
Monteiro 2012	Y	N	Y
Ng 2014	N	N	Y
NTUH - Wang 2011	Y	Y	Y
NTUH - Wang 2015b	Y	Y	Y
Ohno 2000	Y	N	Y
POSSUELO - Brito 2014	Y	Y	Y
POSSUELO - Possuelo 2008	Y	N	Y
Rana 2014	Y	Y	N
Roy 2001	Y	Y	Y
Roy 2006	N	Y	N
SANTOS - Fernandes 2015	Y	Y	Y
SANTOS - Santos 2013	Y	Y	Y
Sharma 2014	Y	Y	Y
Shimizu 2006	Y	Y	Y
Singla 2014	Y	Y	Y
Sotsuka 2011	Y	Y	N
Teixeira 2011	N	N	N
Vuilleumier 2006	Y	N	Y
Wang 2010	Y	Y	Y
Xiang 2014	Y	Y	Y
YAMADA - Yamada 2009	Y	Y	N
YAMADA - Yamada 2010	Y	Y	Y
Yimer 2011	Y	Y	Y
Yuliwulandari 2016	N	Y	Y
Zaverucha-do-Valle 2014	Y	N	N
Zazuli 2015	Y	Y	Y

N: no; Y: yes

Table 41 Full quality assessment results: treatment adherence

Group identifier (if applicable) and study name	Is adherence to treatment measured?	If adherence is measured, are adjustments for non-adherence made in the analyses?
ADACS - Chen 2015	N	N/A
ADACS - Lv 2012	N	N/A
ADACS - Tang 2012	N	N/A
ADACS - Tang 2013a	N	N/A
ADACS - Tang 2013b	N	N/A
ADACS - Wang 2015a	N	N/A
ADACS - Wang 2015c	N	N/A
An 2012	N	N/A
Azuma 2013	N	N/A
Bose 2011	N	N/A
Çetintaş 2008	N	N/A
Chamorro 2013	N	N/A
Chang 2012	N	N/A
Chatterjee 2010	N/A (DOTS)	N/A
Cho 2007	Y	N/A
Costa 2012	Y	N/A
Dhoro 2013	N	N/A
Feng 2014	N	N/A
Fredj 2016	N	N/A
Gogtay 2016	N	N/A
GUPTA - Gupta 2013a	N	N/A
GUPTA - Gupta 2013b	N	N/A
He 2015	N	N/A
HIGUCHI - Higuchi 2007	N	N/A
HIGUCHI - Nanashima 2012	N	N/A
Ho 2013	N	N/A
HUANG - Huang 2003	N	N/A
HUANG - Huang 2007	N	N/A
Jung 2015	N	N/A
Khalili 2011	N	N/A
KIM - Kim 2009	N	N/A
KIM - Kim 2010	N	N/A
KIM - Kim 2011	Y	Y
KIM - Kim 2012a	Y	Y
KIM - Kim 2012b	N	N/A
KIM - Kim 2012c	Y	Y
KIM - Kim 2015	Y	Y
Kwon 2012	N	N/A
Lee 2010	N	N/A
LEIRO - Leiro 2008	Y	Y

Group identifier (if applicable) and study name	Is adherence to treatment measured?	If adherence is measured, are adjustments for non-adherence made in the analyses?
LEIRO - Leiro-Fernandez 2011	Y	Y
Li 2012	N	N/A
Liu 2014	N	N/A
Mahmoud 2012	N	N/A
Monteiro 2012	N	N/A
Ng 2014	N	N/A
NTUH - Wang 2011	N	N/A
NTUH - Wang 2015b	N	N/A
Ohno 2000	N	N/A
POSSUELO - Brito 2014	N	N/A
POSSUELO - Possuelo 2008	Y	N
Rana 2014	N	N/A
Roy 2001	N	N/A
Roy 2006	N	N/A
SANTOS - Fernandes 2015	N	N/A
SANTOS - Santos 2013	N	N/A
Sharma 2014	N	N/A
Shimizu 2006	N	N/A
Singla 2014	N/A (DOTS)	N/A
Sotsuka 2011	N	N/A
Teixeira 2011	N	N/A
Vuilleumier 2006	Y	Y
Wang 2010	N	N/A
Xiang 2014	N	N/A
YAMADA - Yamada 2009	N	N/A
YAMADA - Yamada 2010	Y	N
Yimer 2011	N	N/A
Yuliwulandari 2016	N	N/A
Zaverucha-do-Valle 2014	N	N/A
Zazuli 2015	N/A (DOTS)	N/A

DOTS: directly observed treatment, short-course; N: no; N/A: not applicable; Y: yes

Appendix 3. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Results of the sensitivity analyses

NAT2 acetylator status

Sensitivity analysis 1: Pairwise comparisons

Slow *versus* rapid acetylator status

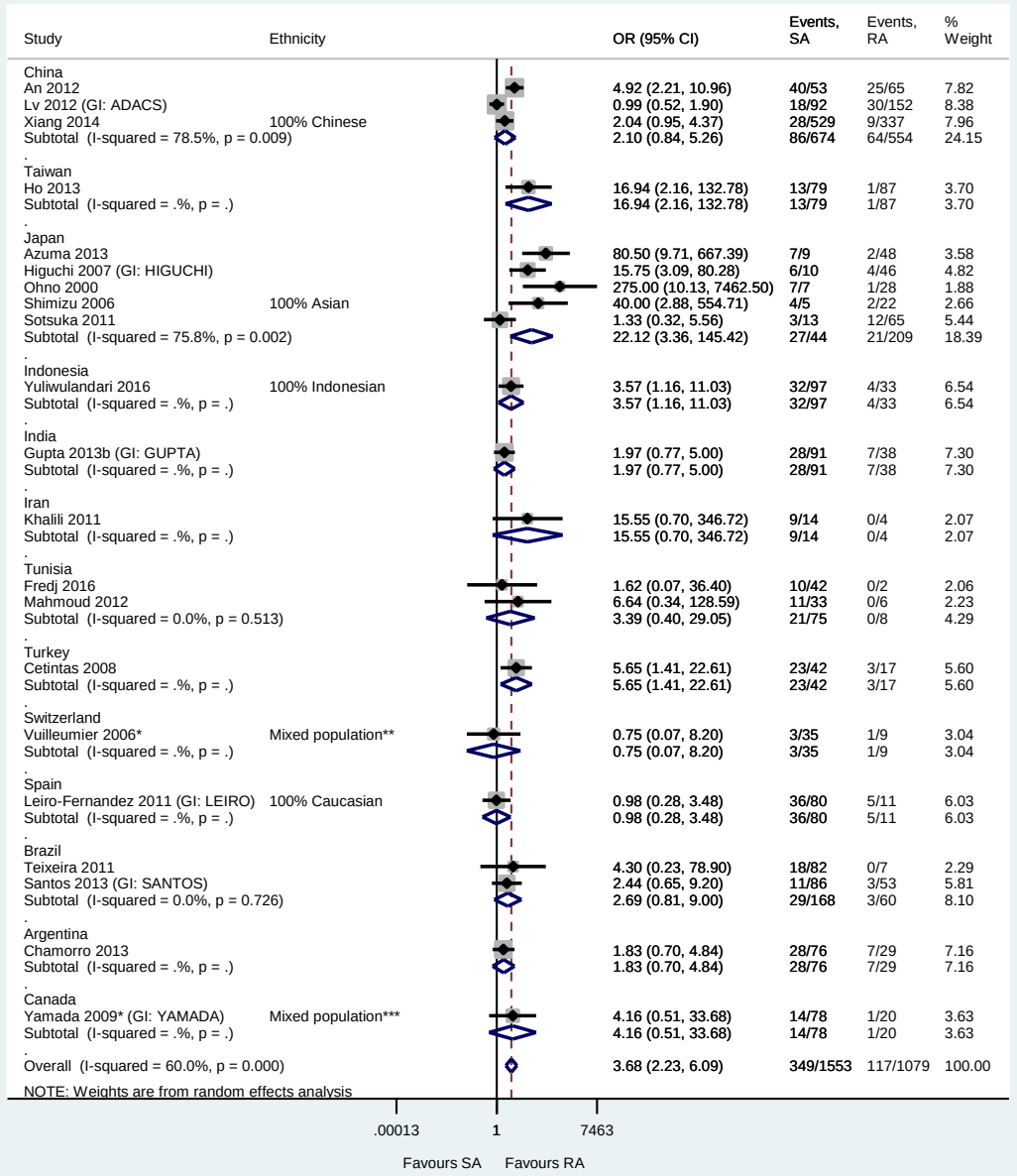

Figure 26 NAT2 *acetylator status and hepatotoxicity: slow* versus *rapid acetylator status*

*Vuilleumier 2006 and Yamada 2009 were both conducted in the latent TB population.
**Caucasian: 38 (43%), Hispanic: 8 (9%), African: 22 (25%), South American: 15 (17%), Asian: 5 (6%), Middle Eastern: 1 (1%)
*** Asian: 72 (42%), Caucasian: 49 (29%), South Asian: 22 (13%), Hispanic: 7 (4%), Middle Eastern: 8 (5%), First nations: 5 (3%), other/mixed/unknown: 7 (4%)
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio; RA: rapid acetylators; SA: slow acetylators; TB: tuberculosis

Intermediate *versus* rapid acetylator status

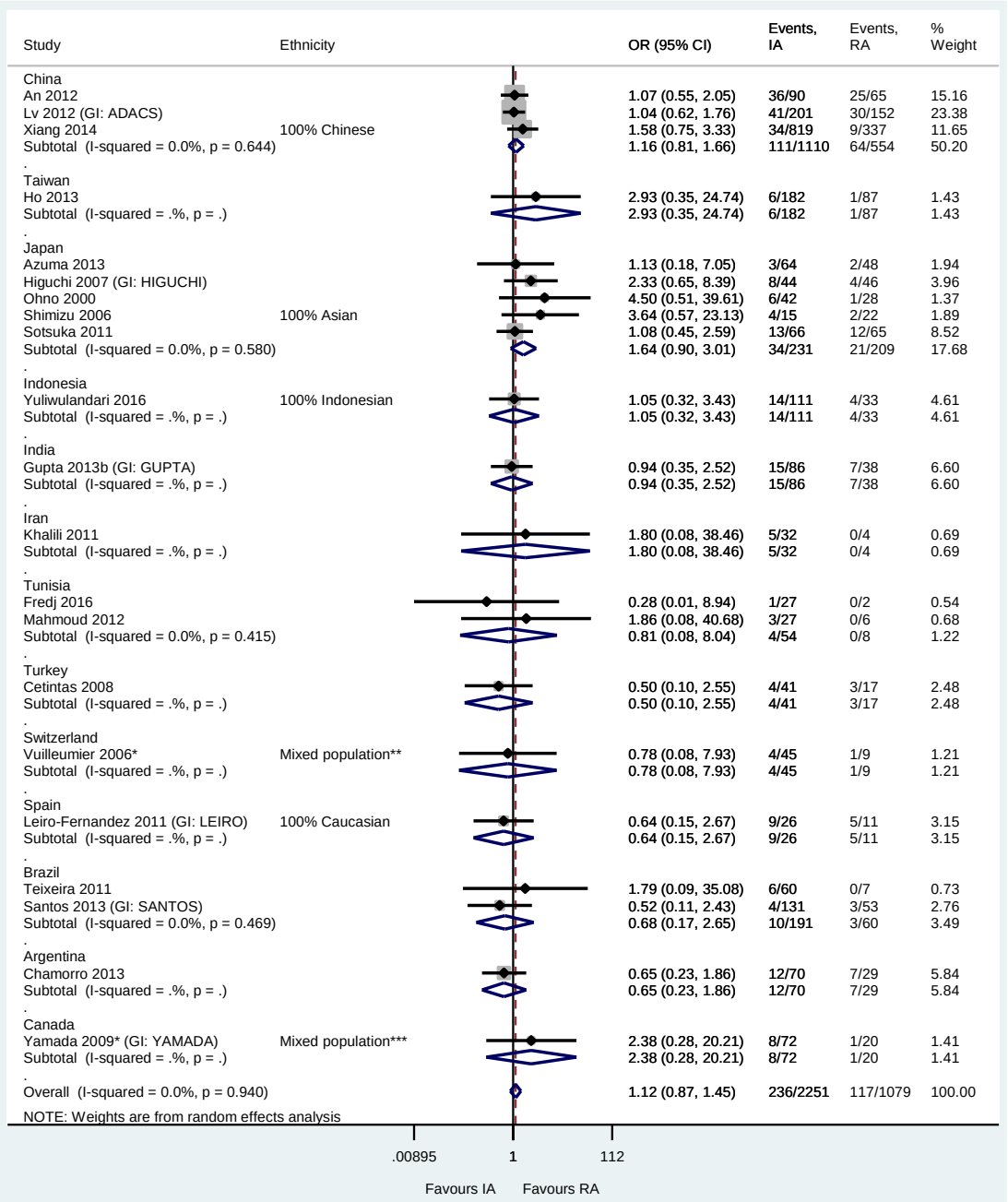

Figure 27 NAT2 *acetylator status and hepatotoxicity: intermediate* versus *rapid acetylator status*

*Vuilleumier 2006 and Yamada 2009 were both conducted in the latent TB population.
**Caucasian: 38 (43%), Hispanic: 8 (9%), African: 22 (25%), South American: 15 (17%), Asian: 5 (6%), Middle Eastern: 1 (1%)
*** Asian: 72 (42%), Caucasian: 49 (29%), South Asian: 22 (13%), Hispanic: 7 (4%), Middle Eastern: 8 (5%), First nations: 5 (3%), other/mixed/unknown: 7 (4%)
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; IA: intermediate acetylators; OR: odds ratio; RA: rapid acetylators; TB: tuberculosis

Sensitivity analysis 2: Slow versus *rapid/intermediate acetylator status*

Results from this sensitivity analysis are presented in the main thesis (Section 4.2).

*CYP2E1 Rsa*I polymorphism

Sensitivity analysis 1: Excluding studies where genotypes deviated from HWE

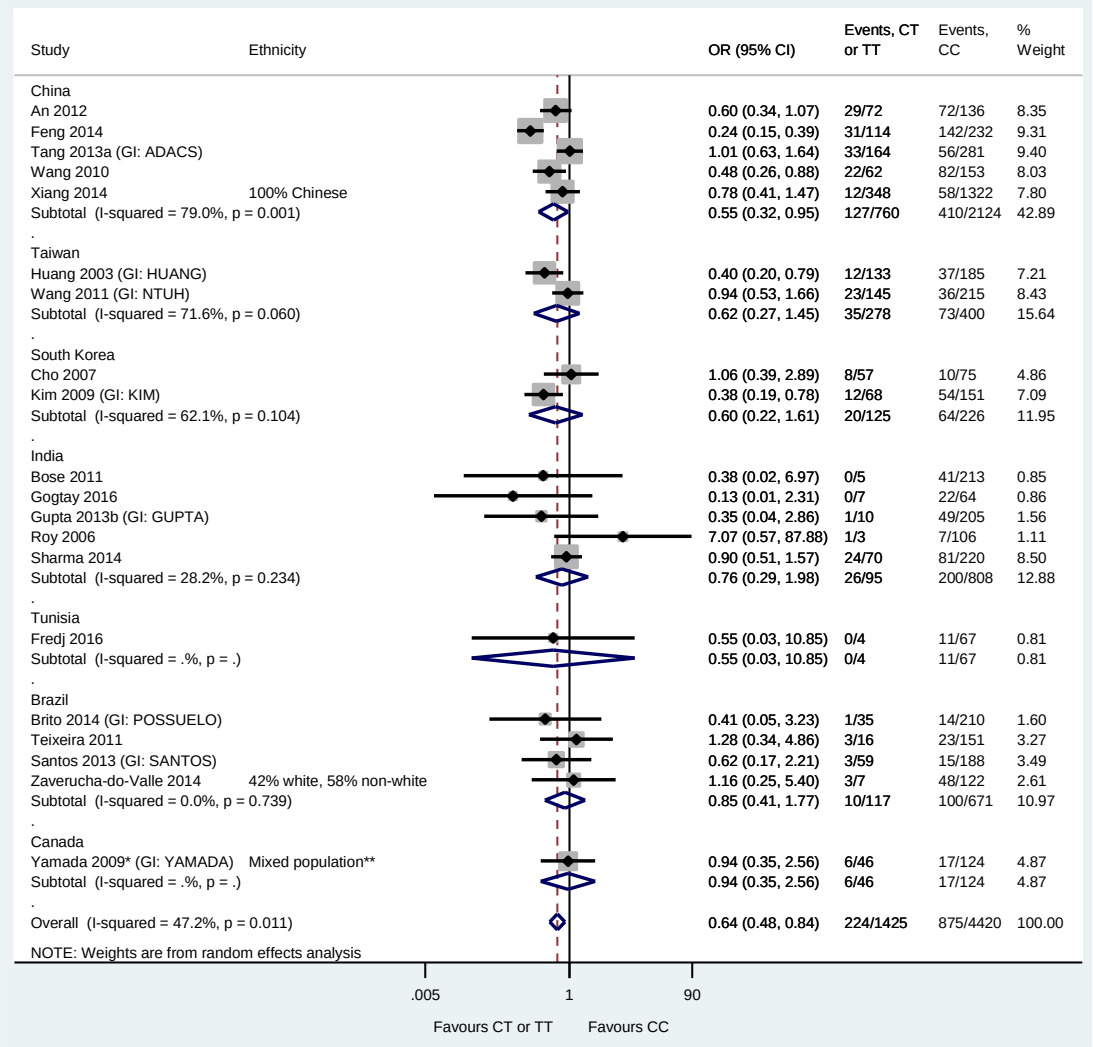

Figure 28 CYP2E1 Rsa*I polymorphism and hepatotoxicity: homozygous mutant-type (TT) or heterozygous (CT)* versus *homozygous wild-type (CC), sensitivity analysis excluding studies where genotypes deviated from HWE (Rana 2014, Singla 2014 and Sotsuka 2011)*

* Yamada 2009 was conducted in the latent TB population.
**Asian: 72 (42%), Caucasian: 49 (29%), South Asian: 22 (13%), Hispanic: 7 (4%), Middle Eastern: 8 (5%), First Nations: 5 (3%), other/mixed/unknown: 7 (4%)
It was not possible to test HWE for five studies that did not report data for each genotype group separately (Feng 2014, Wang 2011, Cho 2007, Brito 2014, Yamada 2009).
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; HWE: Hardy-Weinberg equilibrium; OR: odds ratio; TB: tuberculosis

Sensitivity analysis 2: Pairwise comparisons

Heterozygous (CT) *versus* homozygous wild-type (CC)

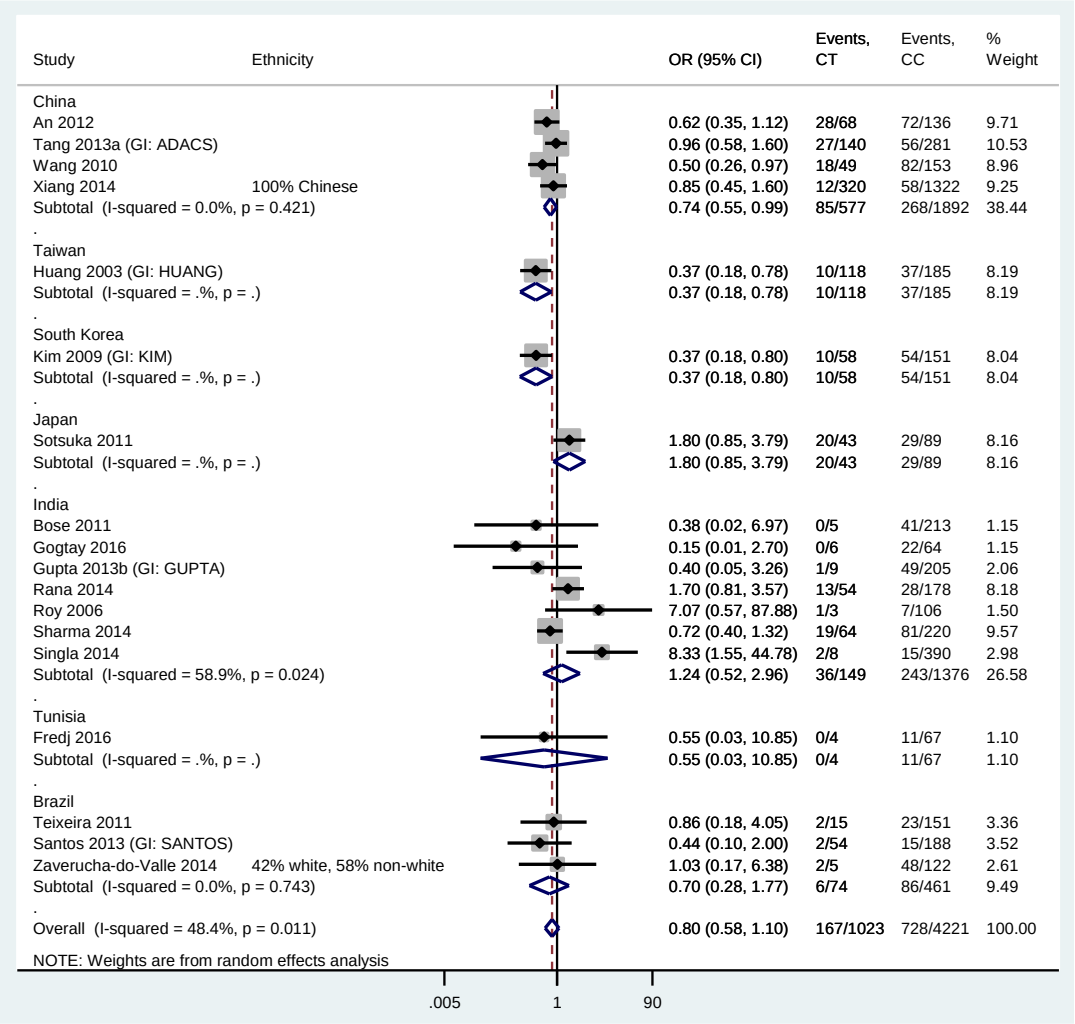

Figure 29 CYP2E1 RsaI *polymorphism and hepatotoxicity: heterozygous (CT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) *versus* homozygous wild-type (CC)

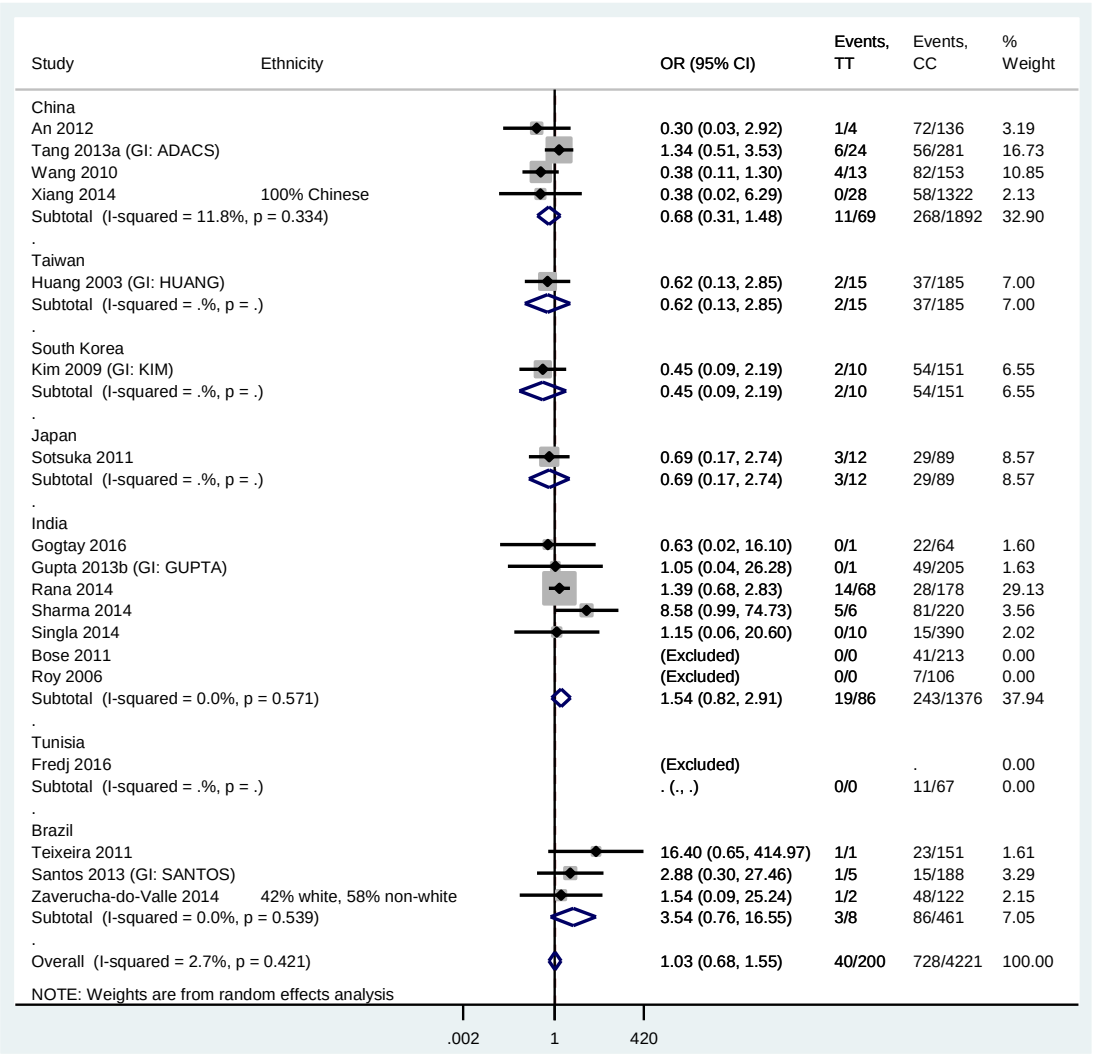

Figure 30 CYP2E1 RsaI *polymorphism and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

*CYP2E1 Dra*I polymorphism

Sensitivity analysis 1: Excluding studies where genotypes deviated from HWE

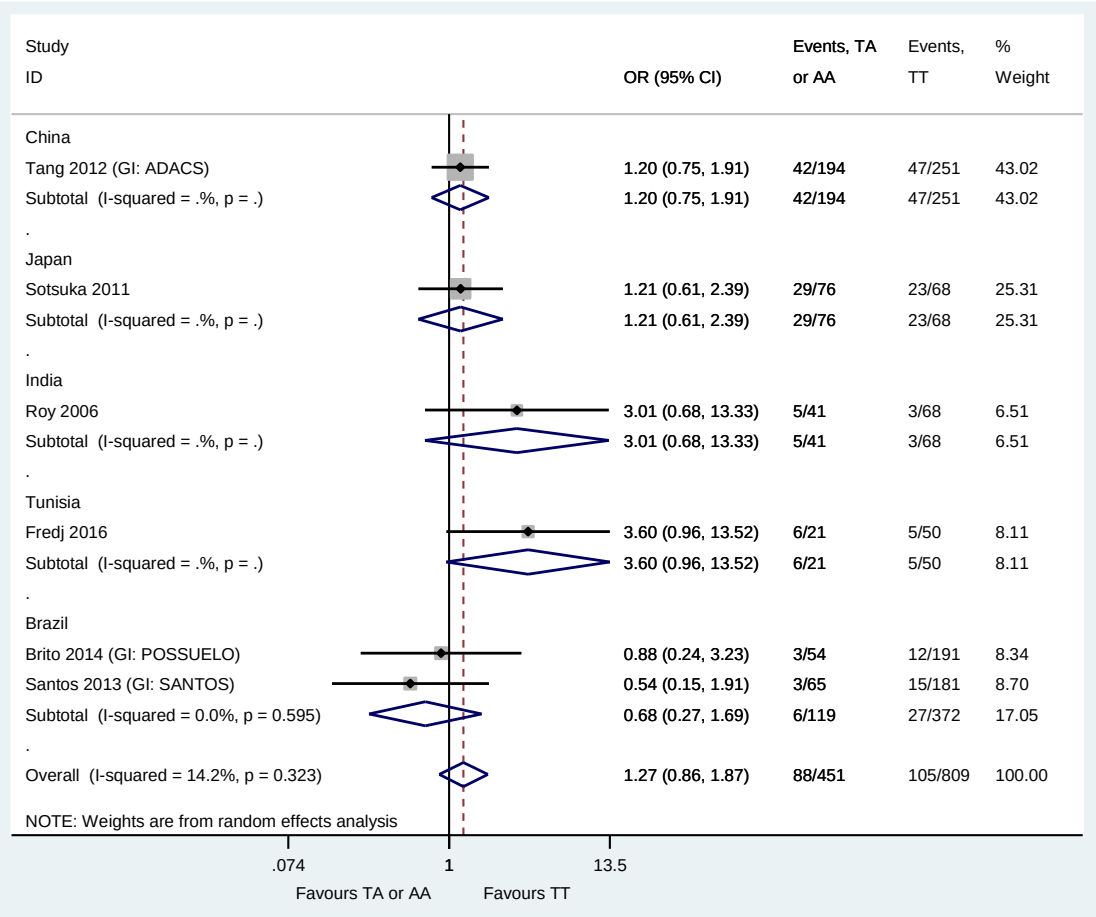

Figure 31 CYP2E1 Dra*I polymorphism and hepatotoxicity: homozygous mutant-type (AA) or heterozygous (TA)* versus *homozygous wild-type (TT), sensitivity analysis excluding two studies where genotypes deviated from HWE (Bose 2011 and Gupta 2013b)*

None of the included studies reported ethnicity so this information is not provided on the forest plot.
It was not possible to test HWE for one study that did not report data for each genotype group separately (Brito 2014).
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; HWE: Hardy-Weinberg equilibrium; OR: odds ratio

Sensitivity analysis 2: Pairwise comparisons

Heterozygous (TA) versus *homozygous wild-type (TT)*

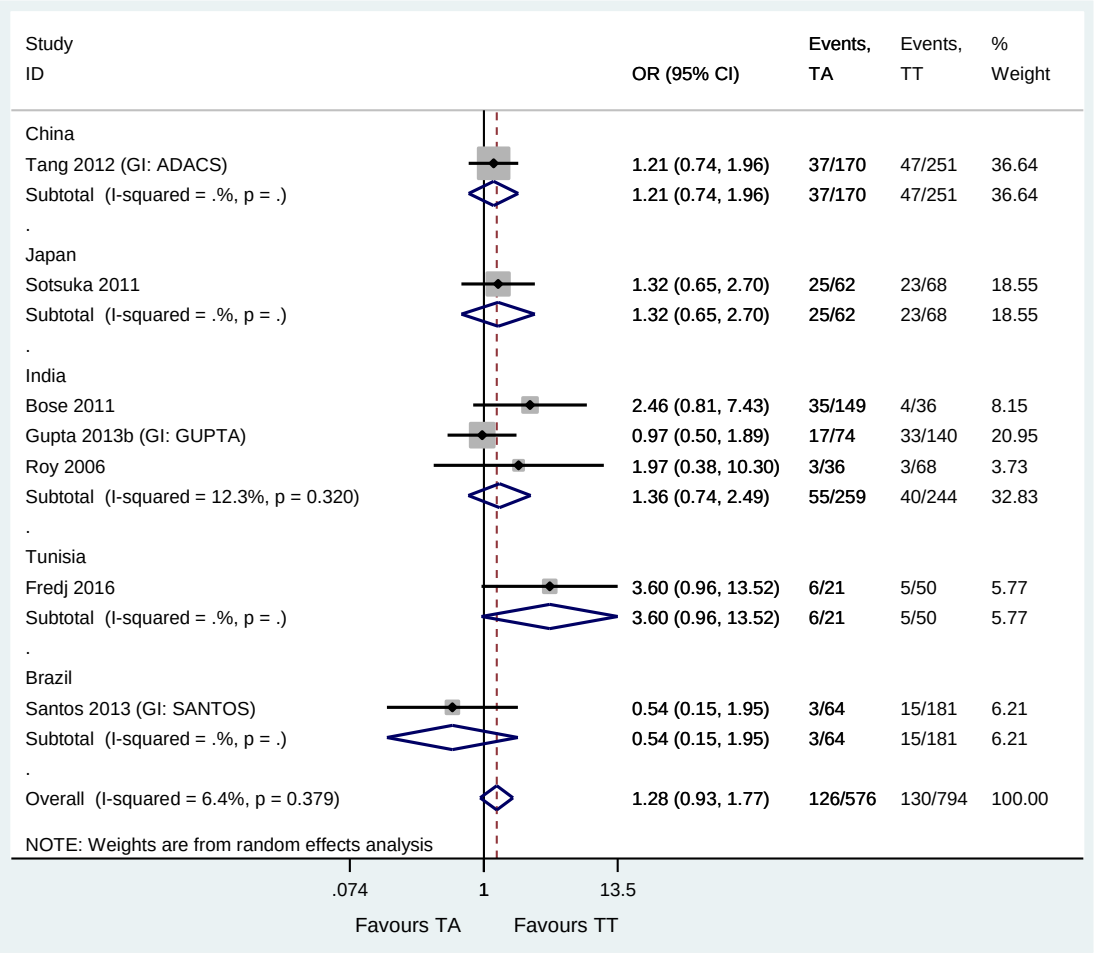

Figure 32 CYP2E1 DraI *polymorphism and hepatotoxicity: heterozygous (TA)* versus *homozygous wild-type (TT)*

None of the included studies reported ethnicity so this information is not provided on the forest plot.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (AA) versus *homozygous wild-type (TT)*

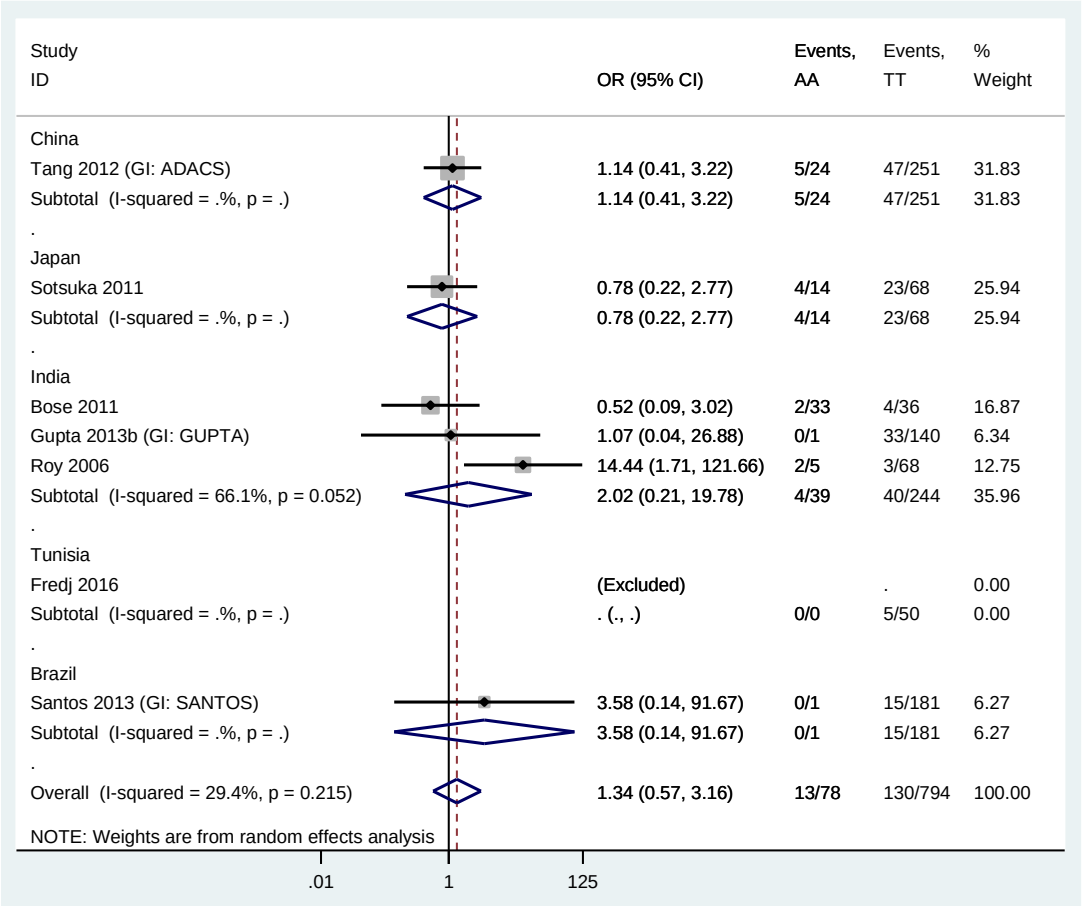

Figure 33 CYP2E1 *DraI polymorphism and hepatotoxicity: homozygous mutant-type (AA)* versus *homozygous wild-type (TT)*

None of the included studies reported ethnicity so this information is not provided on the forest plot.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

*CYP2E1 Pst*I polymorphism

Sensitivity analysis 1: Excluding studies where genotypes deviated from HWE

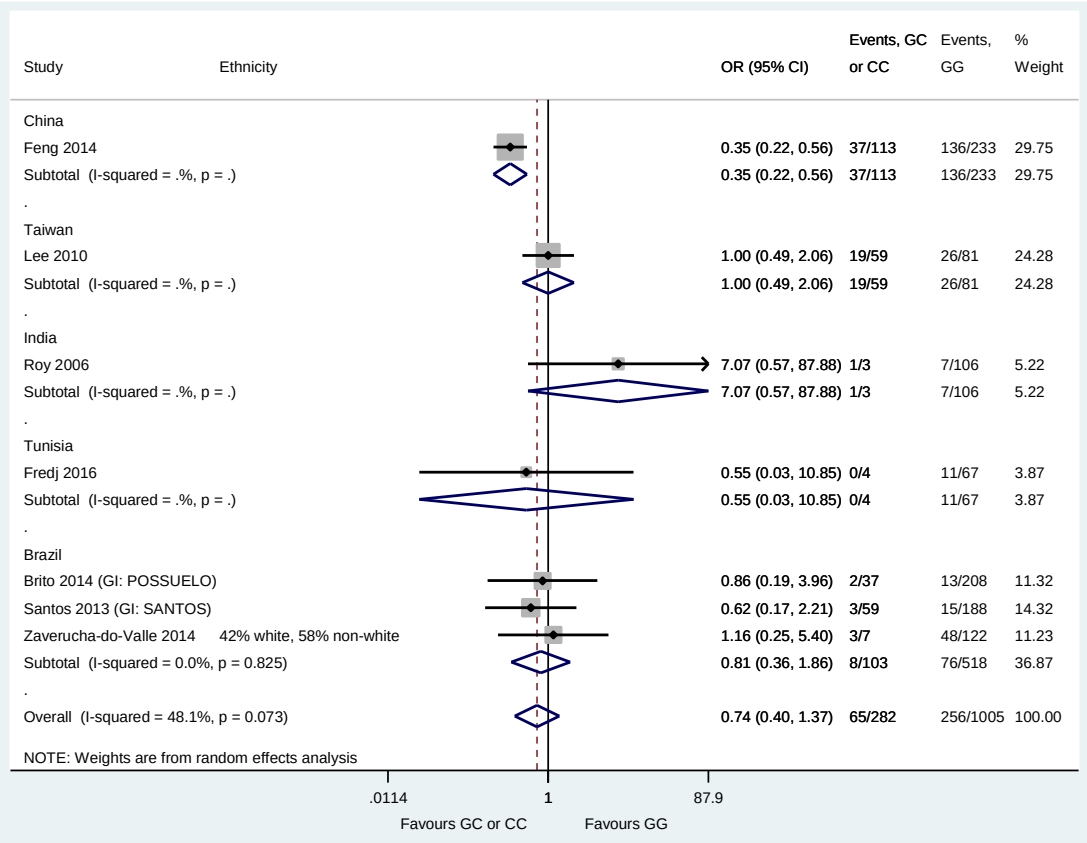

Figure 34 CYP2E1 Pst*I polymorphism and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (GC)* versus *homozygous wild-type (GG), sensitivity analysis excluding one study where genotypes deviated from HWE (Chamorro 2013)*

It was not possible to test HWE for two studies that did not report data for each genotype group separately (Brito 2014 and Feng 2014).
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; HWE: Hardy-Weinberg equilibrium; OR: odds ratio

Sensitivity analysis 2: Pairwise comparisons

Heterozygous (GC) *versus* homozygous wild-type (GG)

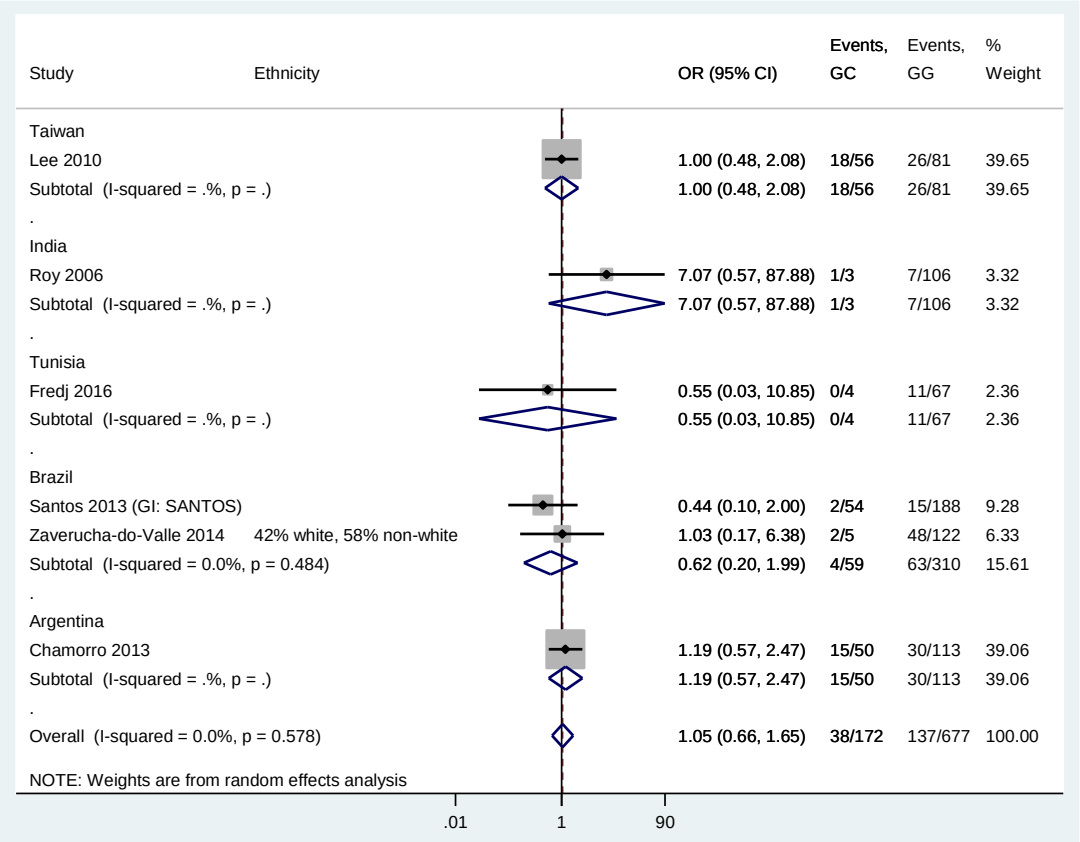

Figure 35 CYP2E1 PstI *polymorphism and hepatotoxicity: heterozygous (GC)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio;

Homozygous mutant-type (CC) *versus* homozygous wild-type (GG)

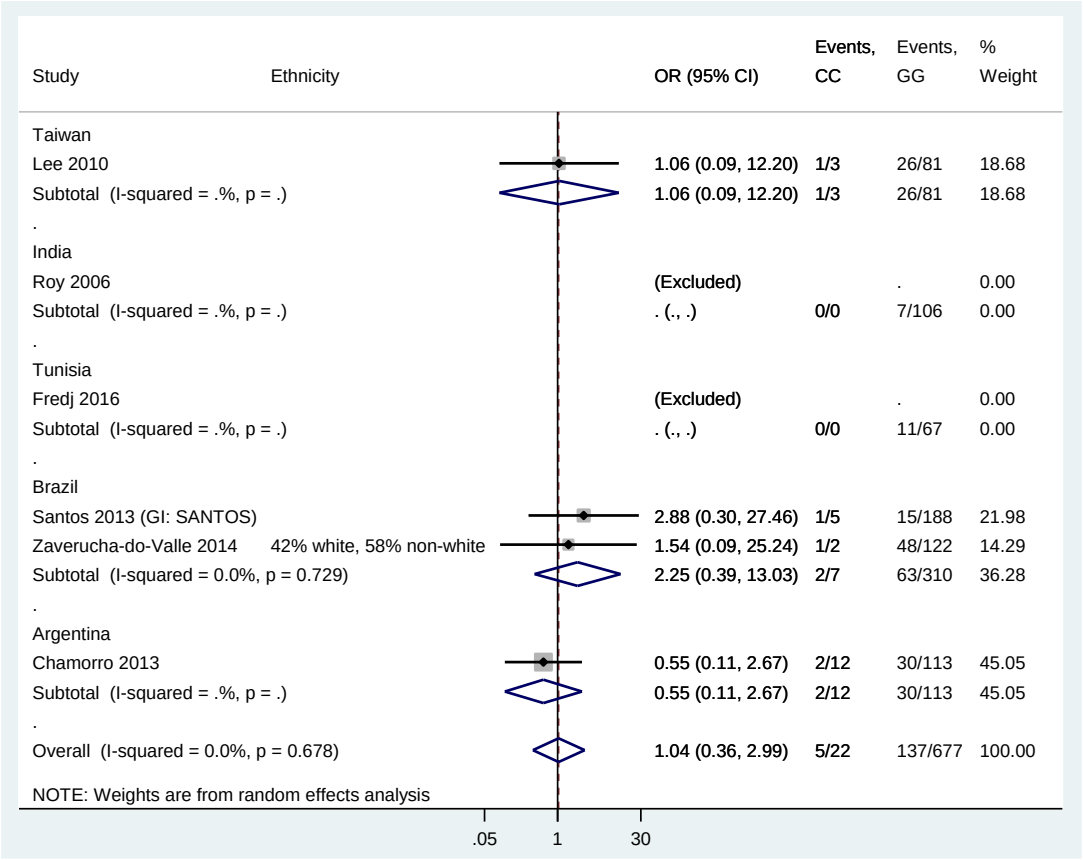

Figure 36 CYP2E1 PstI *polymorphism and hepatotoxicity: homozygous mutant-type (CC)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

GSTM1 null polymorphism

Sensitivity analysis 1: Excluding studies where genotypes deviated from HWE

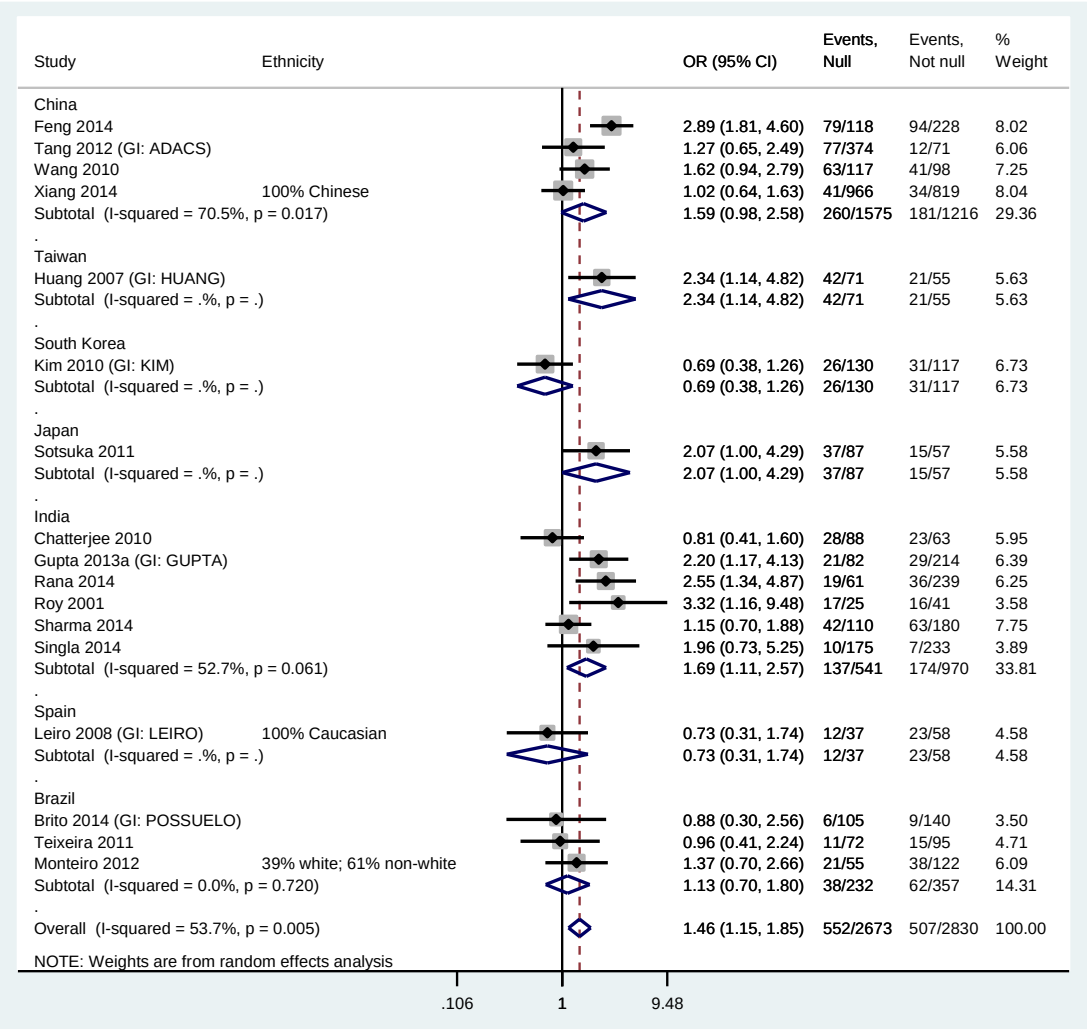

Figure 37 GSTM1 *null polymorphism and hepatotoxicity: homozygous null* versus *heterozygous or homozygous present, sensitivity analysis excluding one study where genotypes deviated from HWE (Liu 2014)*

It was only possible to test deviation from HWE for one study (Liu 2014), which provided data for each genotype group separately.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; HWE: Hardy-Weinberg equilibrium; OR: odds ratio

Sensitivity analysis 2: Pairwise comparisons

As only one study reported on each genotype group separately for the *GSTM1* gene, no meta-analysis was performed. Instead, we calculated ORs and corresponding 95% CIs, as shown in Table 42 below.

Table 42 GSTM1 *null polymorphism and hepatotoxicity: results of pairwise comparisons*

Study	Country	Ethnicity	Comparison	OR (95% CI)	# cases	# controls
Liu 2014	China	NR	Het *vs* Hom present	0.42 (0.02 to 8.18)	6	47
			Hom null *vs* Hom present	0.97 (0.35 to 2.71)	20	136

CI: confidence interval; Het: heterozygous; Hom: homozygous; NR: not reported; OR: odds ratio

GSTT1 null polymorphism

Sensitivity analysis: Pairwise comparisons

Heterozygous *versus* homozygous present

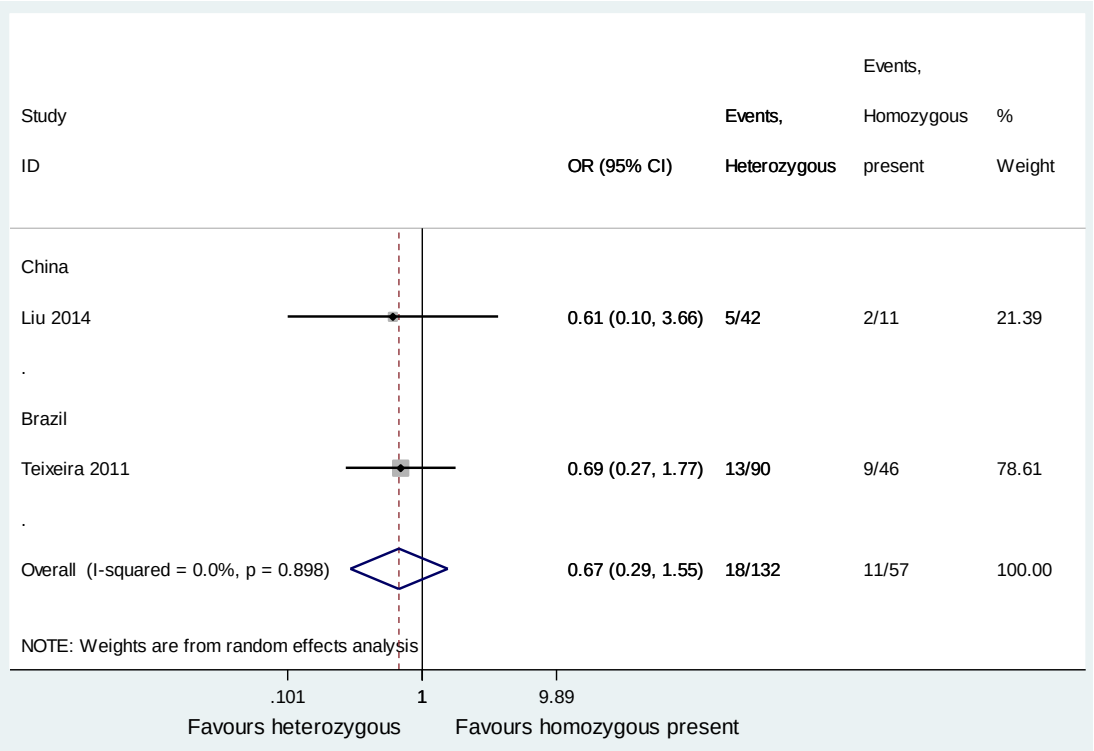

Figure 38 GSTT1 *null polymorphism and hepatotoxicity: heterozygous* versus *homozygous present*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; OR: odds ratio

Homozygous null *versus* homozygous present

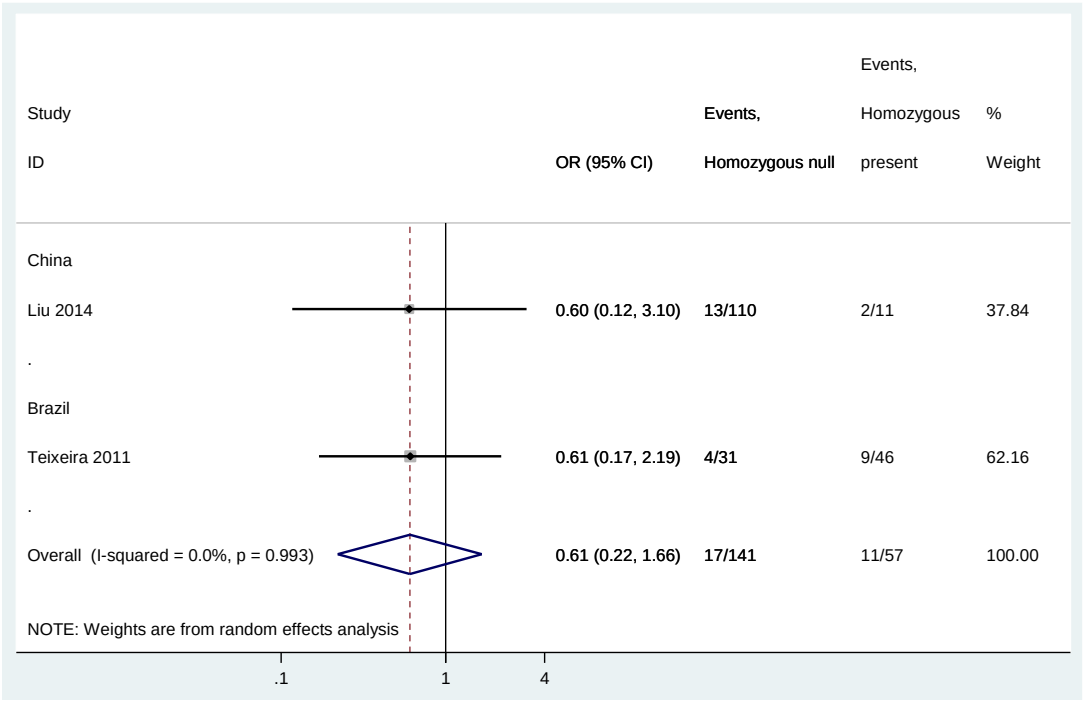

Figure 39 GSTT1 *null polymorphism and hepatotoxicity: homozygous null* versus *homozygous present*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; OR: odds ratio

Appendix 4. Systematic review and meta-analysis of genetic variants and toxicity related to anti-tuberculosis drugs: Forest plots for the secondary analyses

Pairwise comparisons for *NAT2* 282C-T

Heterozygous (CT) versus *homozygous wild-type (CC)*

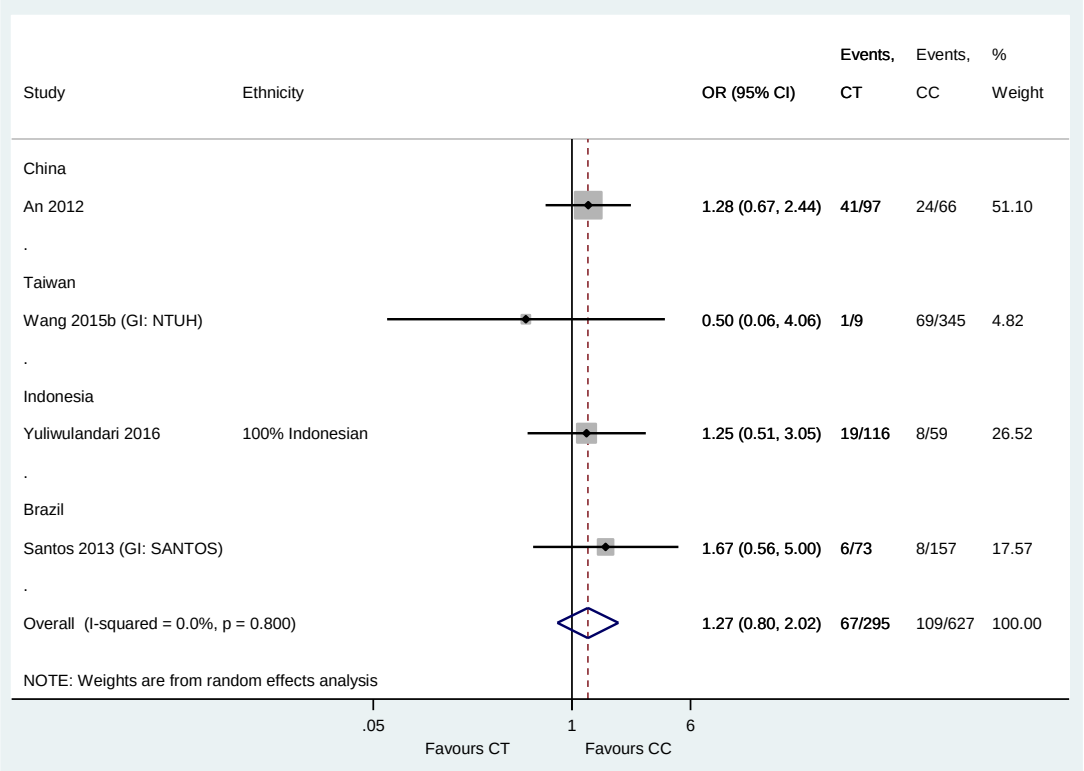

Figure 40 NAT2 *282C-T and hepatotoxicity: heterozygous (CT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

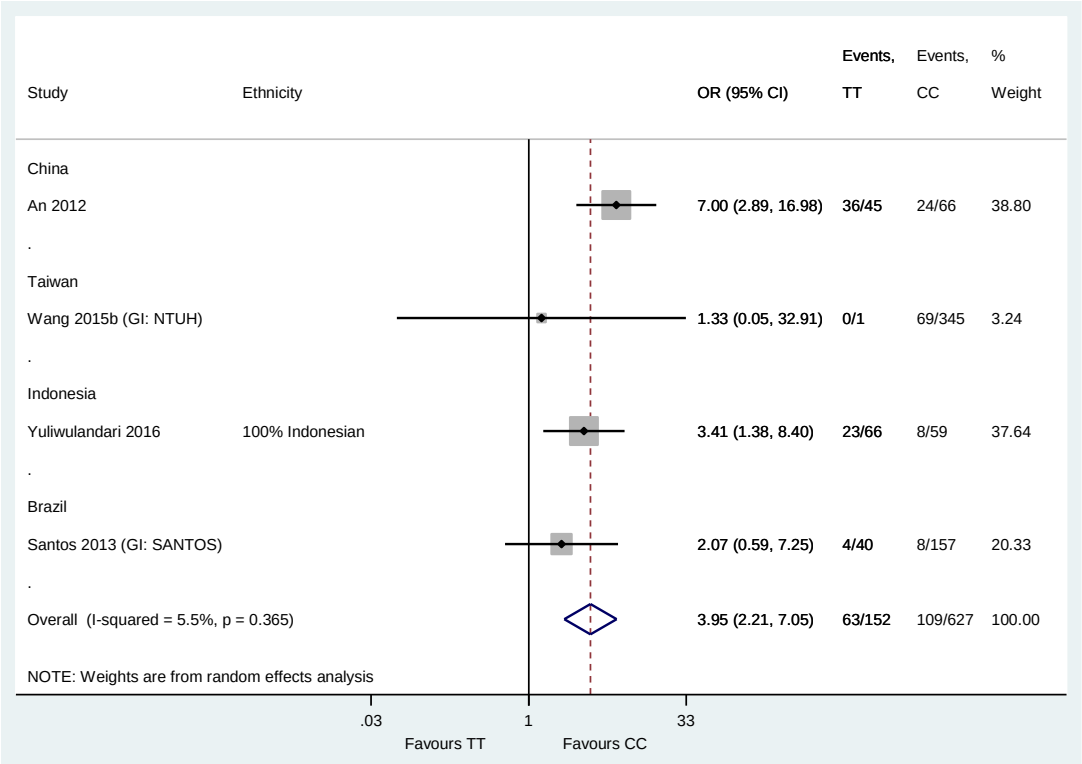

Figure 41 NAT2 *282C-T and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *NAT2* 341T-C

Heterozygous (TC) versus *homozygous wild-type (TT)*

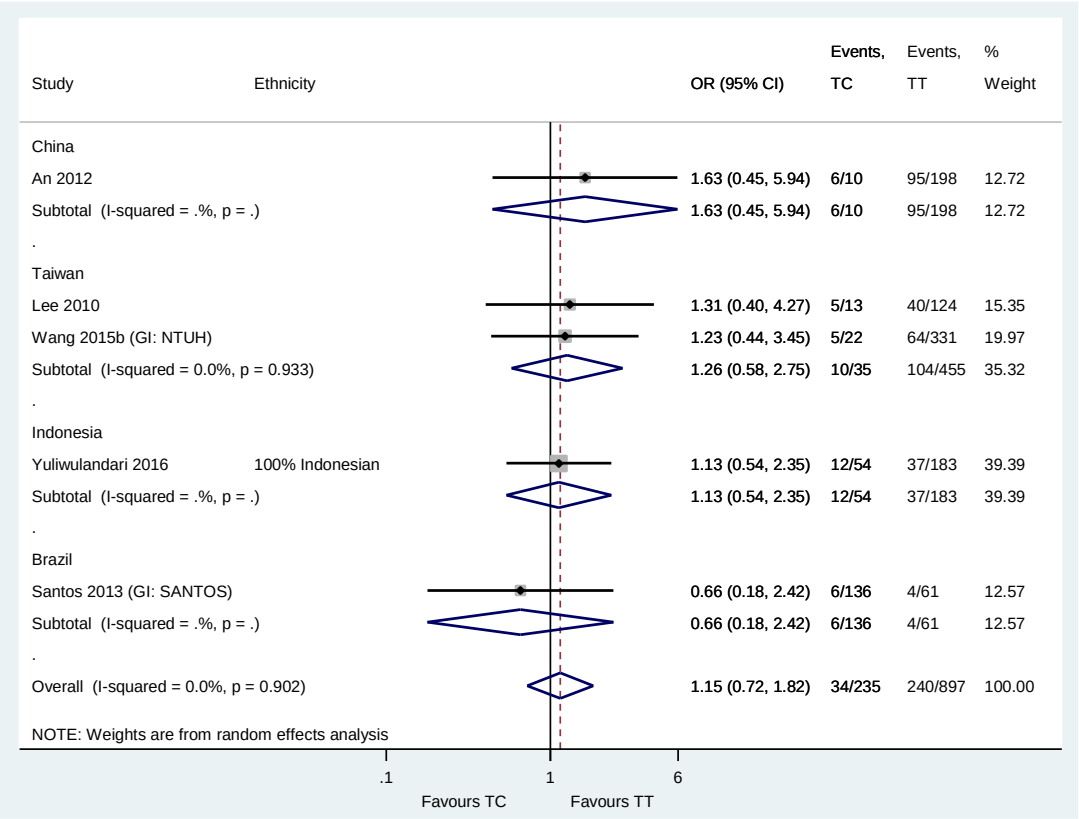

Figure 42 NAT2 *341T-C and hepatotoxicity: heterozygous (TC)* versus *homozygous wild-type (TT)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (CC) versus *homozygous wild-type (TT)*

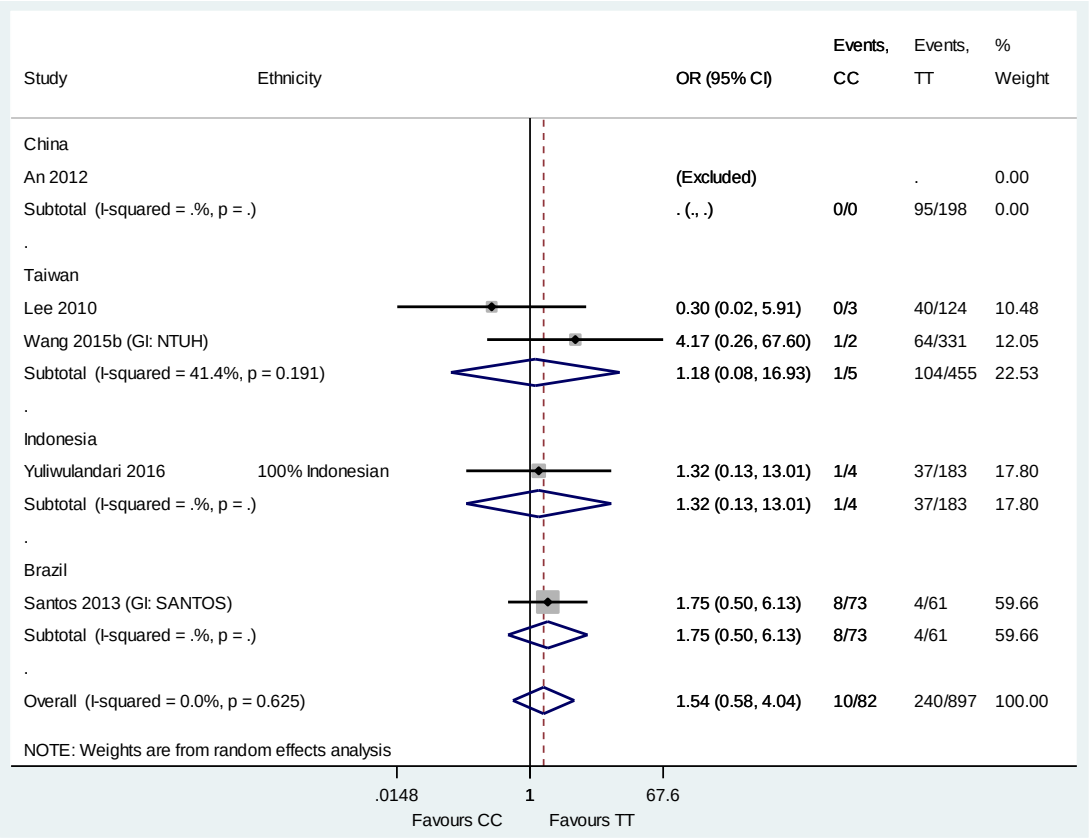

Figure 43 NAT2 *341T-C and hepatotoxicity: homozygous mutant-type (CC)* versus *homozygous wild-type (TT)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *NAT2* 481C-T

Heterozygous (CT) versus *homozygous wild-type (CC)*

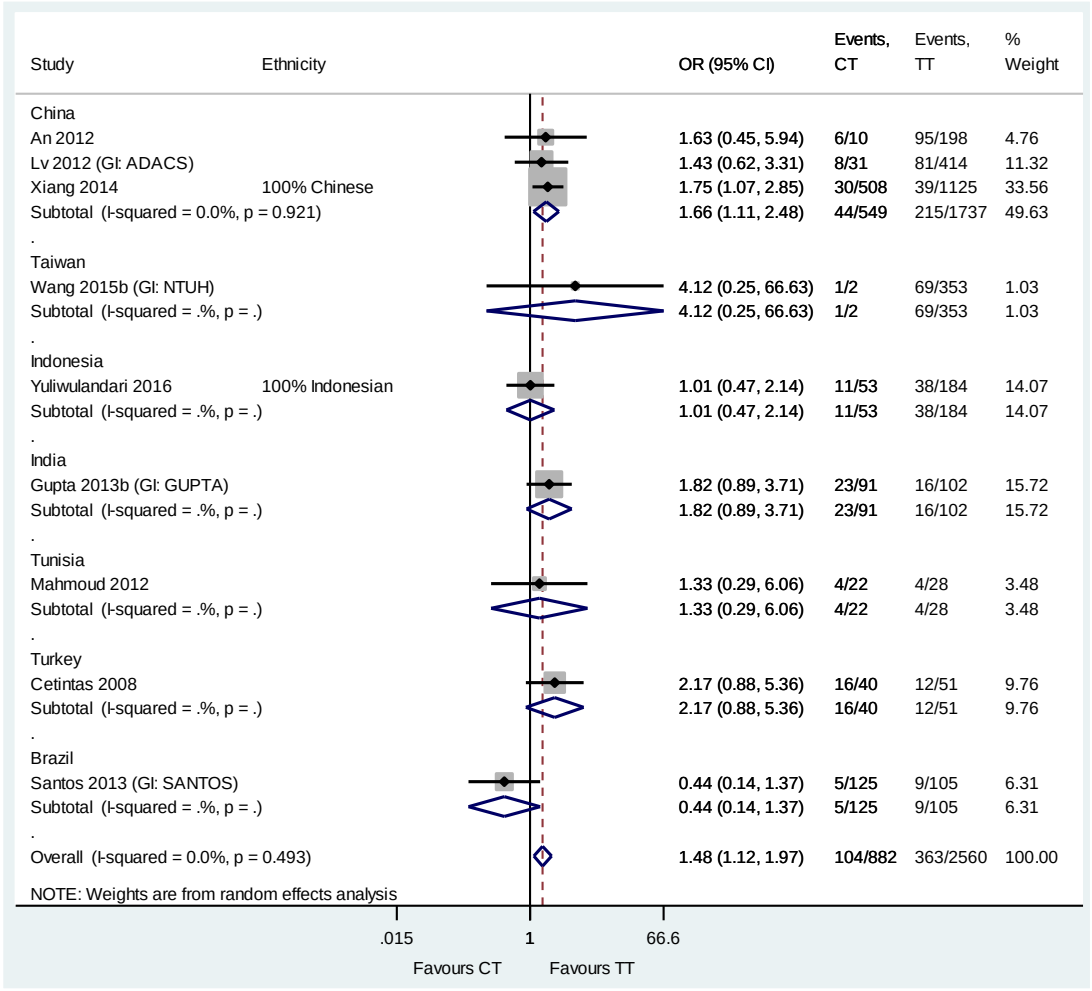

Figure 44 NAT2 *481C-T and hepatotoxicity: heterozygous (CT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

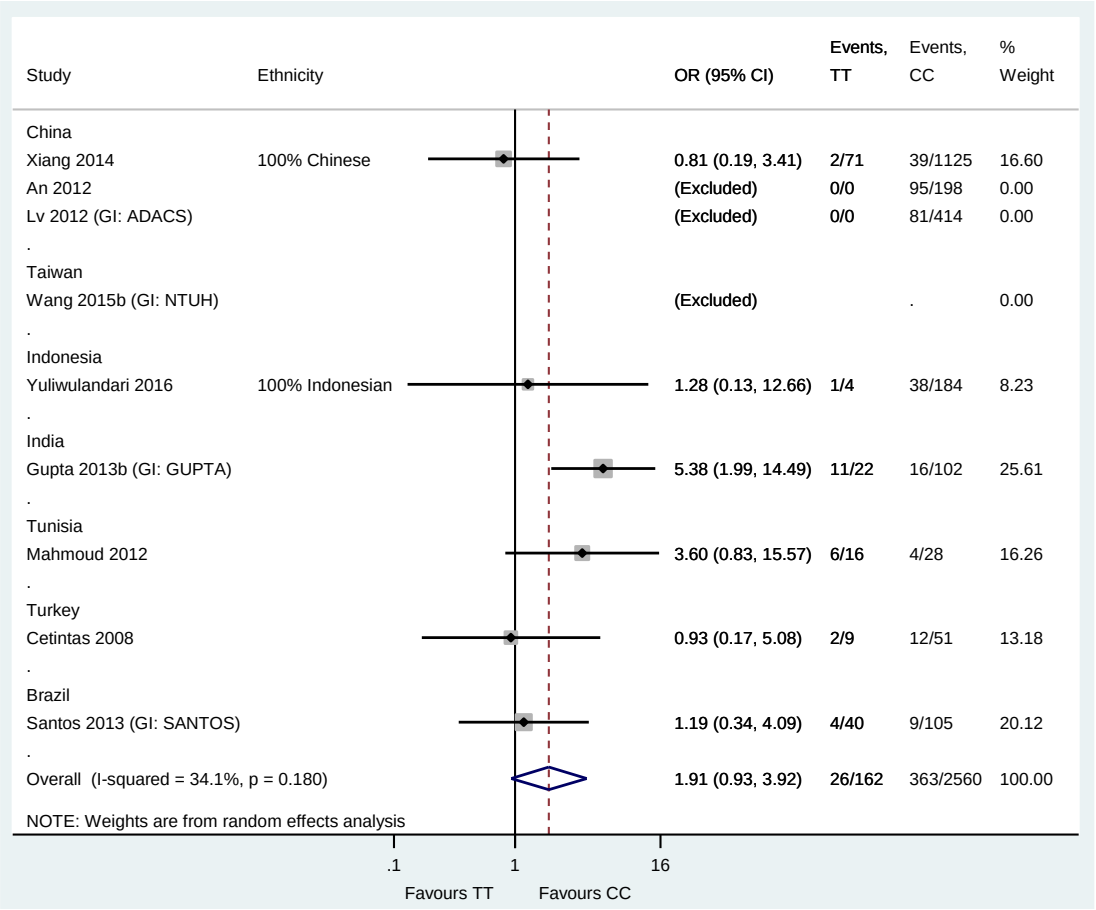

Figure 45 NAT2 *481C-T and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *NAT2* 590G-A

Heterozygous (GA) versus *homozygous wild-type (GG)*

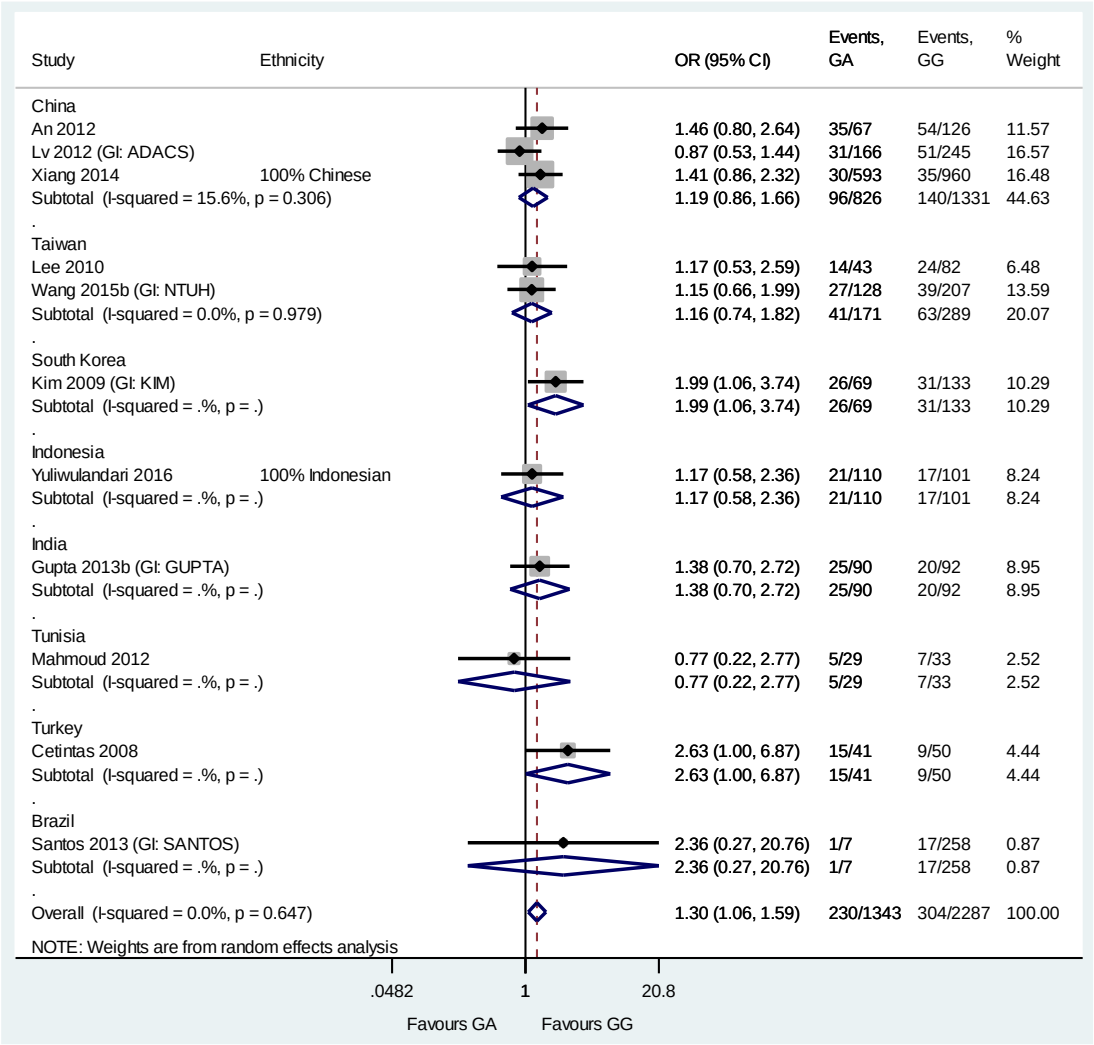

Figure 46 NAT2 *590G-A and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (AA) versus *homozygous wild-type (GG)*

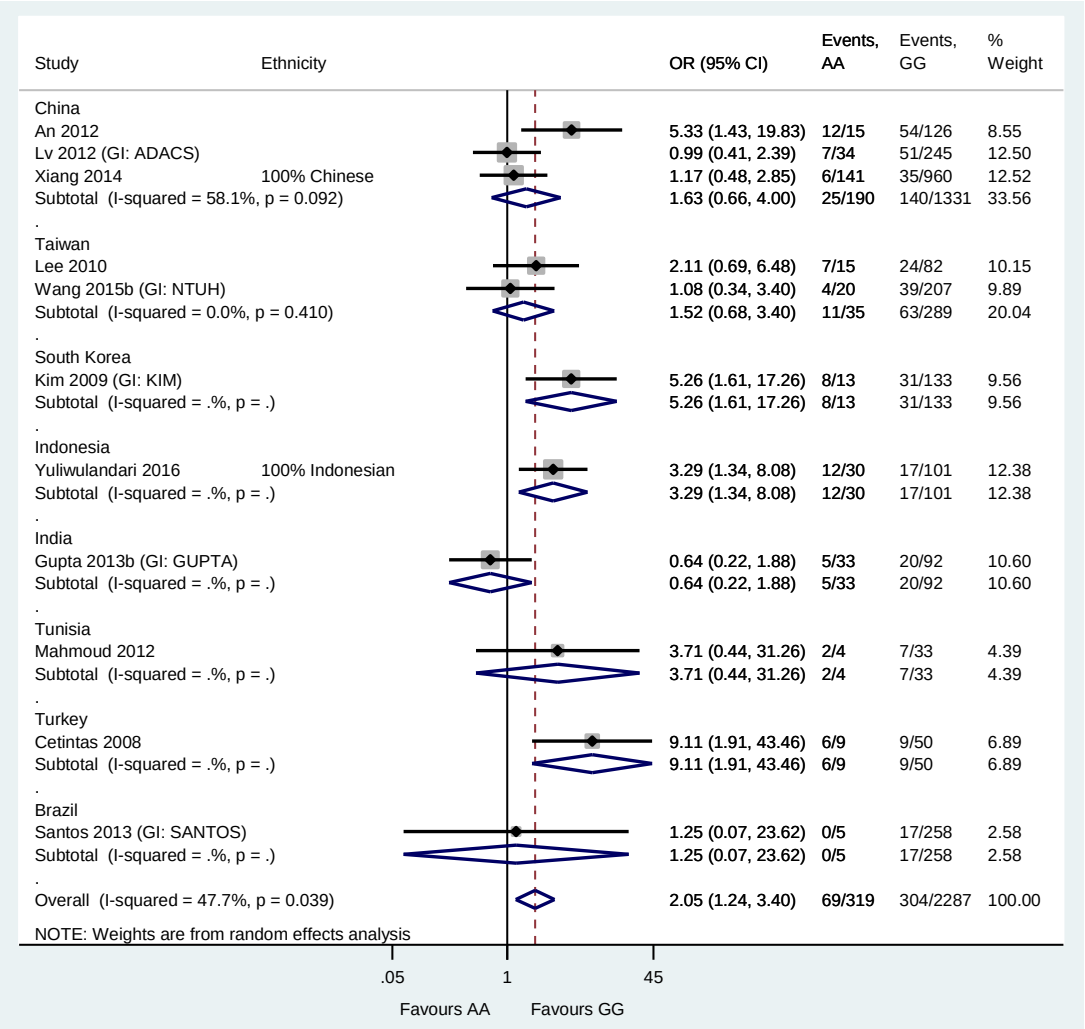

Figure 47 NAT2 *590G-A and hepatotoxicity: homozygous mutant-type (AA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *NAT2* 803A-G

Heterozygous (GA) versus *homozygous wild-type (AA)*

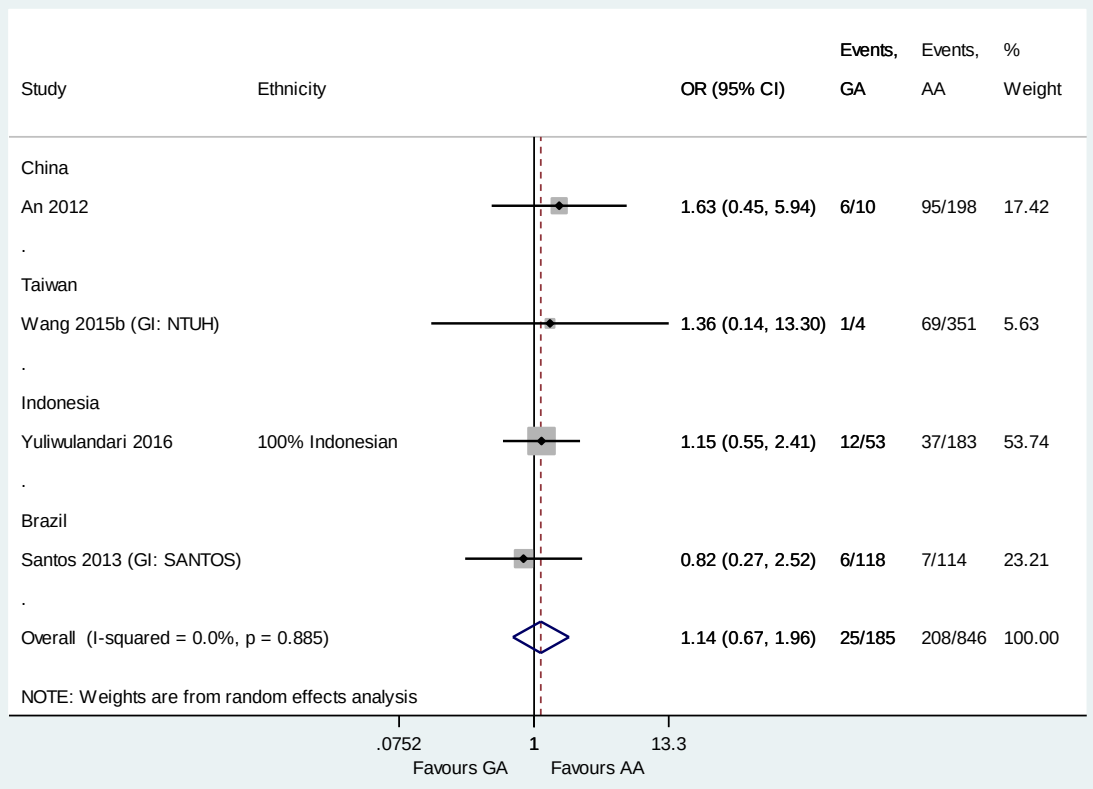

Figure 48 NAT2 *803A-G and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (GG) versus *homozygous wild-type (AA)*

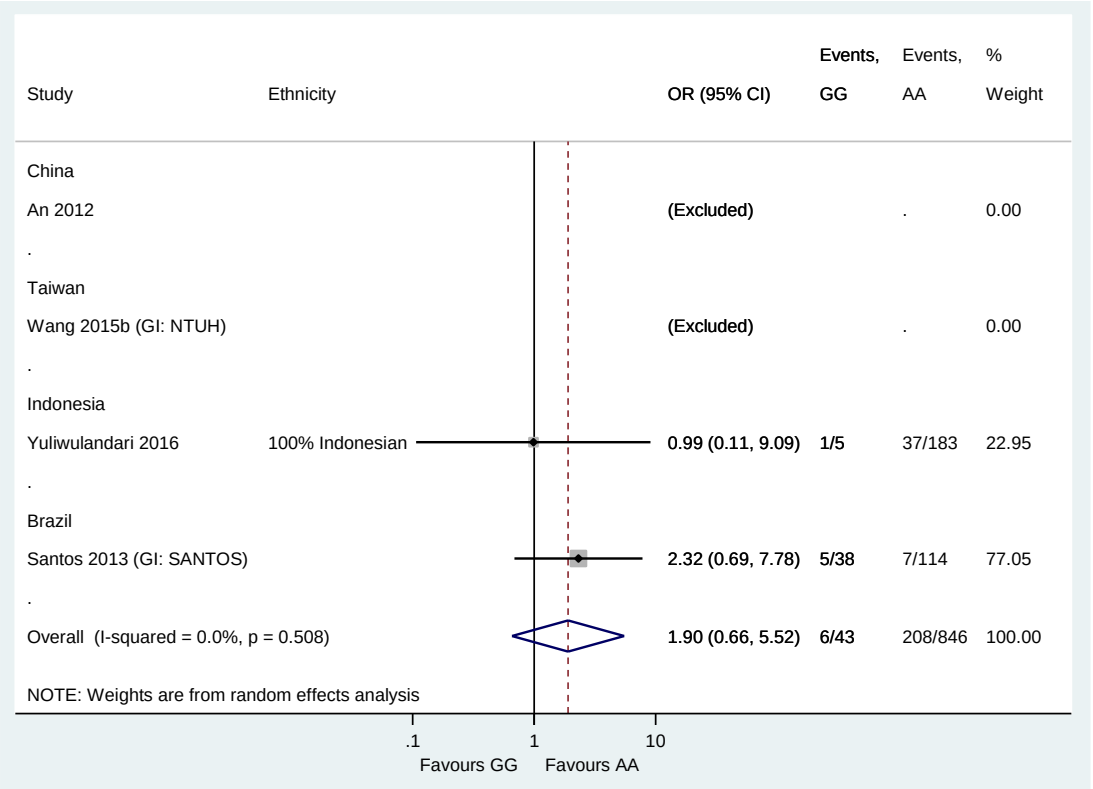

Figure 49 NAT2 *803A-G and hepatotoxicity: homozygous mutant-type (GG)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *NAT2* 857G-A

Heterozygous (GA) versus *homozygous wild-type (GG)*

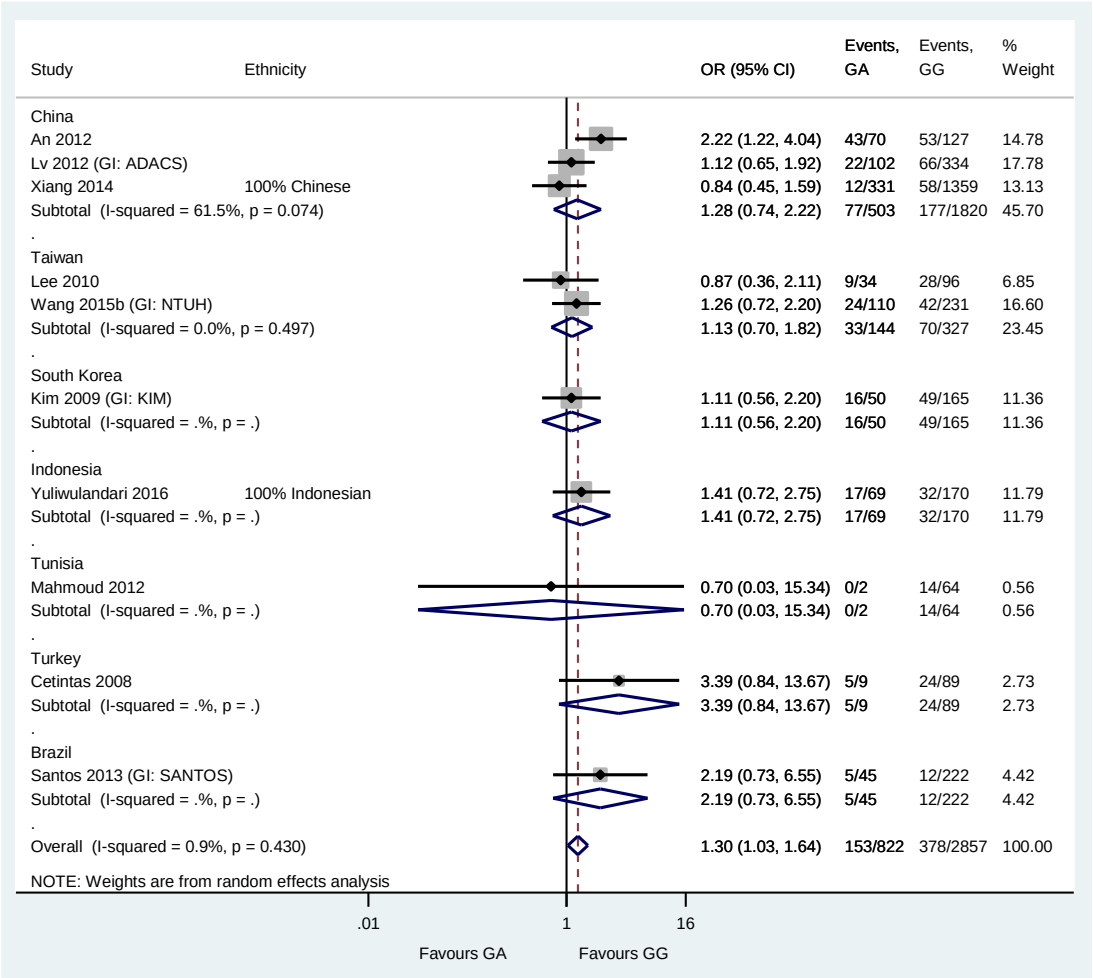

Figure 50 NAT2 *857G-A and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (AA) versus *homozygous wild-type (GG)*

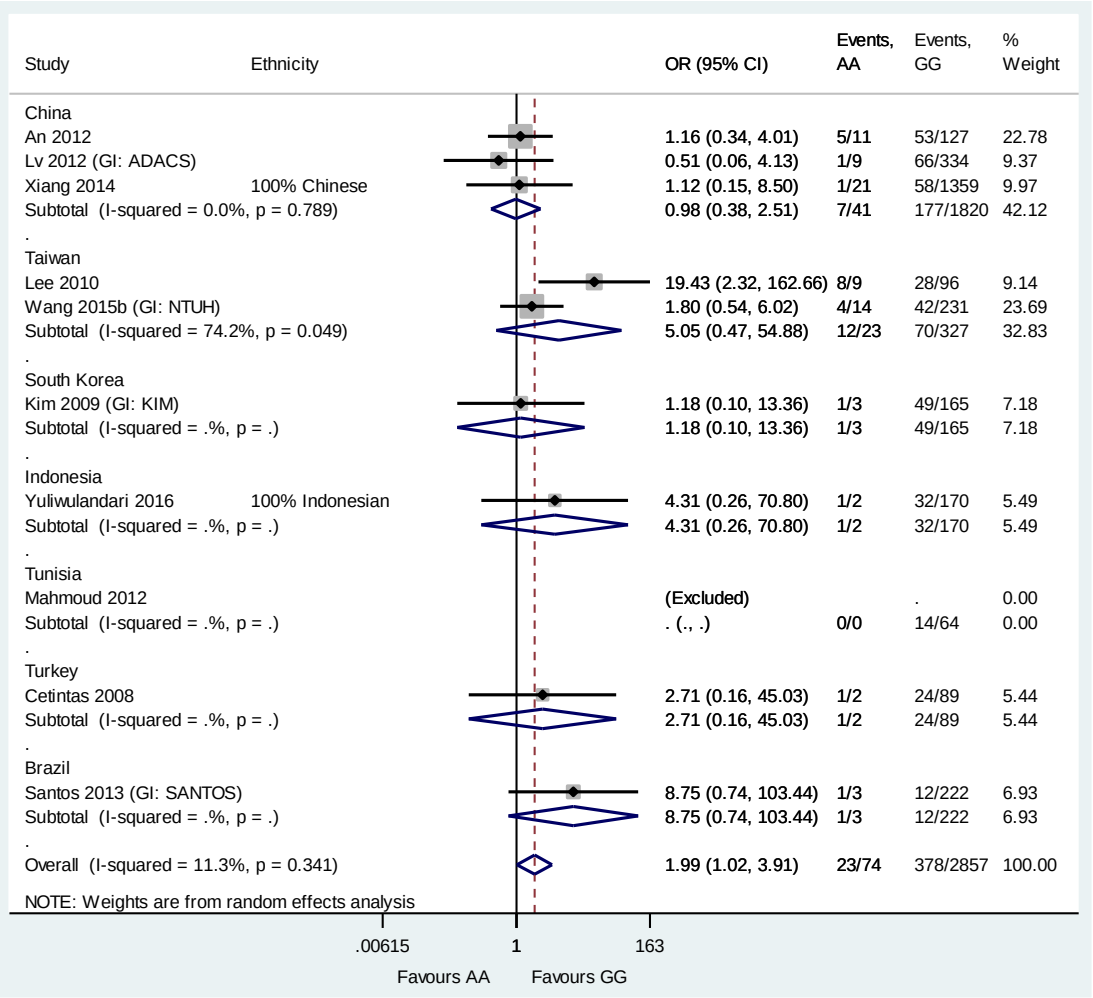

Figure 51 NAT2 *857G-A and hepatotoxicity: homozygous mutant-type (AA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *CYP2E1* 96-bp (deletion-insertion SNP)

Heterozygous (DI) versus *homozygous wild-type (DD)*

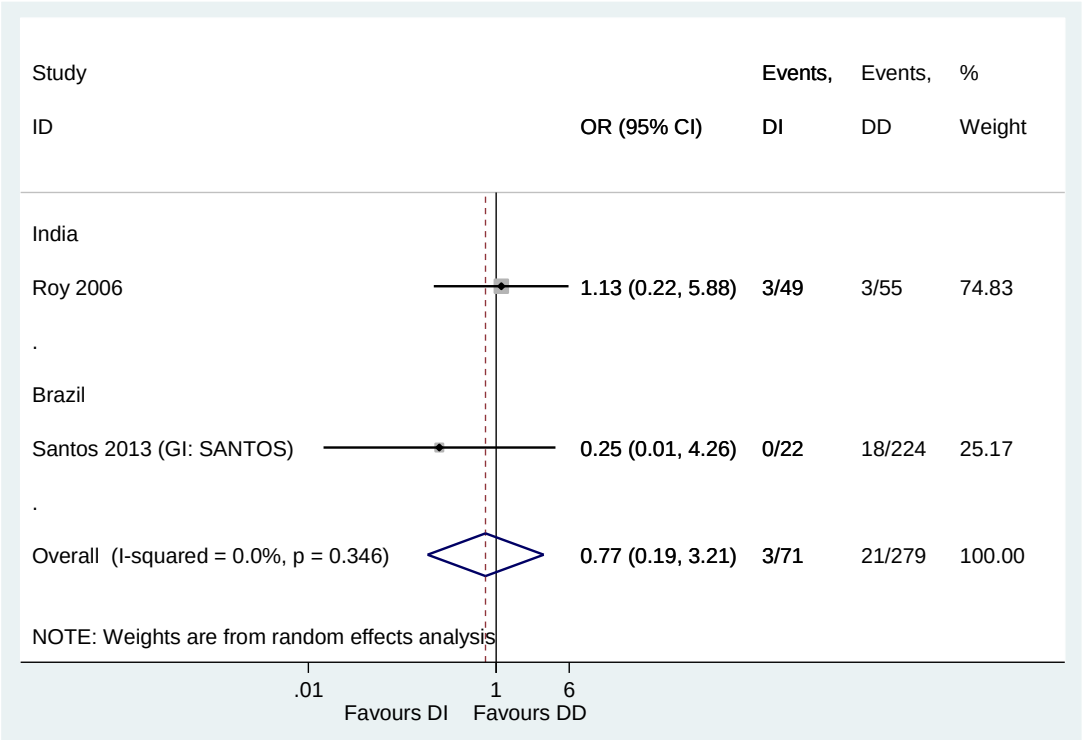

Figure 52 CYP2E1 *96-bp SNP and hepatotoxicity: heterozygous (DI)* versus *homozygous wild-type (DD)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (II) versus *homozygous wild-type (DD)*

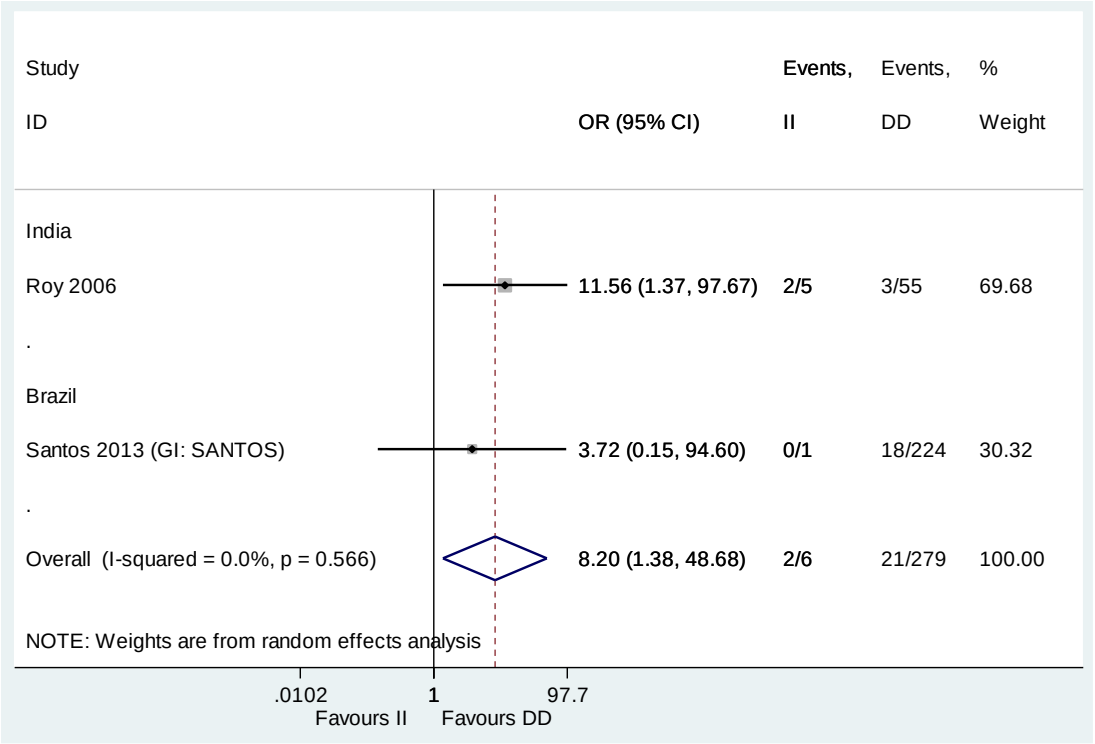

Figure 53 CYP2E1 *96-bp SNP and hepatotoxicity: homozygous mutant-type (II)* versus *homozygous wild-type (DD)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *CYP2C9* rs4918758

Heterozygous (TC) versus *homozygous wild-type (TT)*

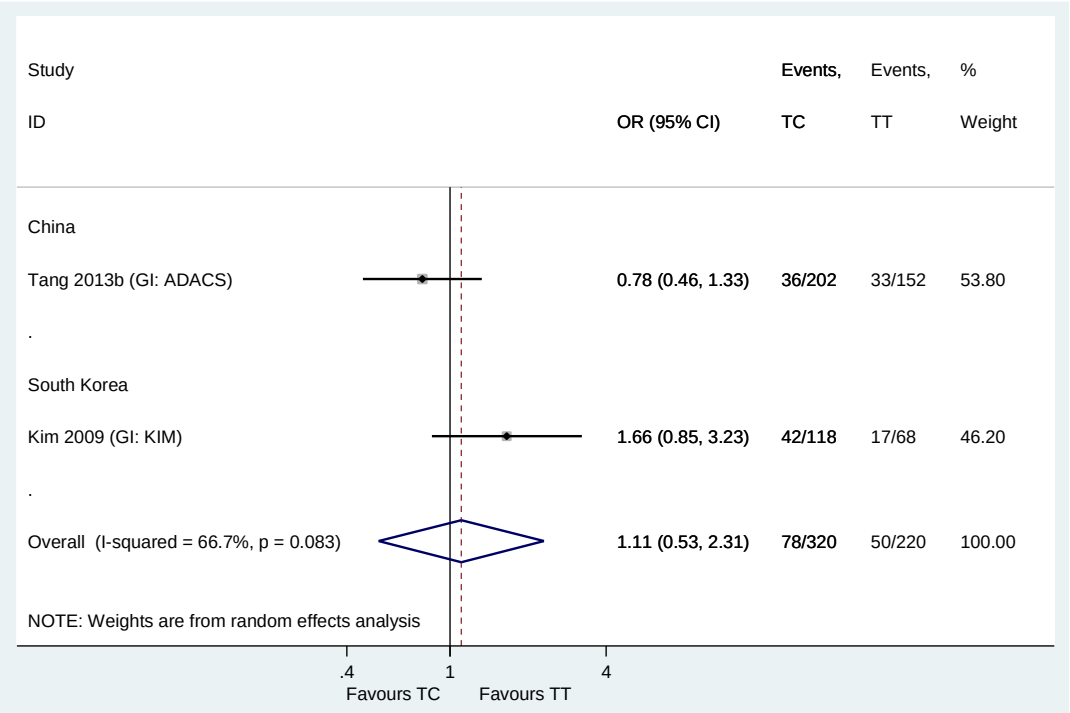

Figure 54 CYP2C9 *rs4918758 and hepatotoxicity: heterozygous (TC)* versus *homozygous wild-type (TT)*

One of the studies (Kim 2009) reports WT to be C and MT to be T, but the other study (Tang 2013b), and the data, suggest that WT is T and MT is C.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Homozygous mutant-type (CC) versus *homozygous wild-type (TT)*

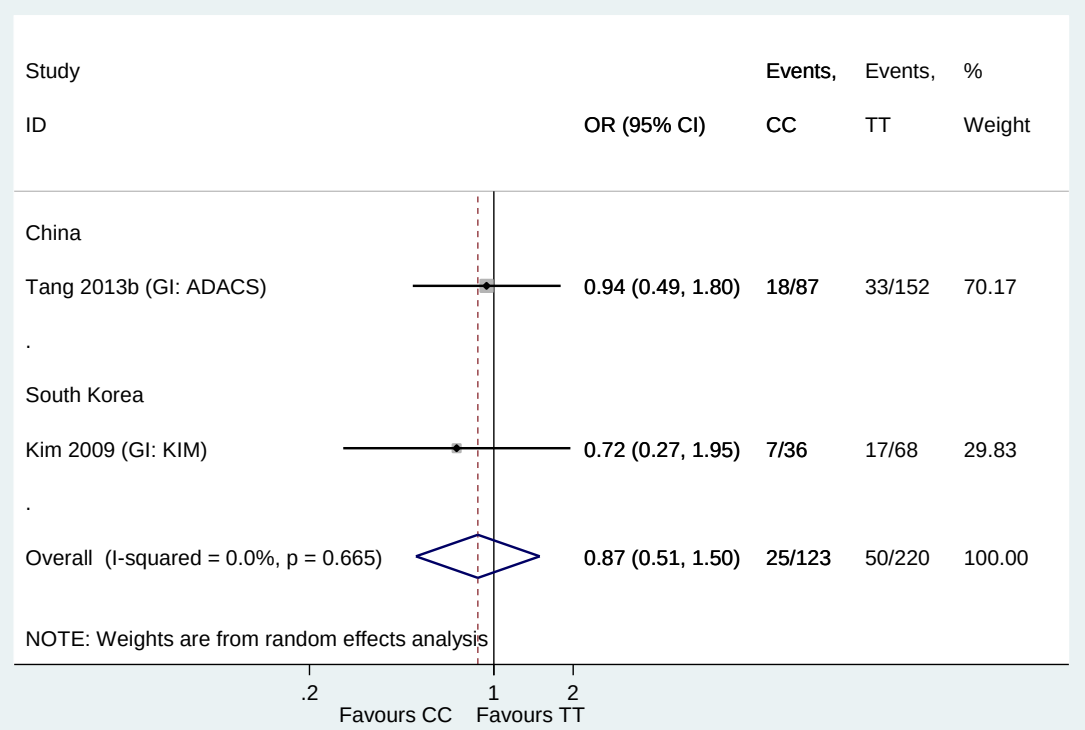

Figure 55 CYP2C9 *rs4918758 and hepatotoxicity: homozygous mutant-type (CC)* versus *homozygous wild-type (TT)*

One of the studies (Kim 2009) reports WT to be C and MT to be T, but the other study (Tang 2013b), and the data, suggest that WT is T and MT is C.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Pairwise comparisons for *CYP2B6* rs3745274

Heterozygous (GT) versus *homozygous wild-type (GG)*

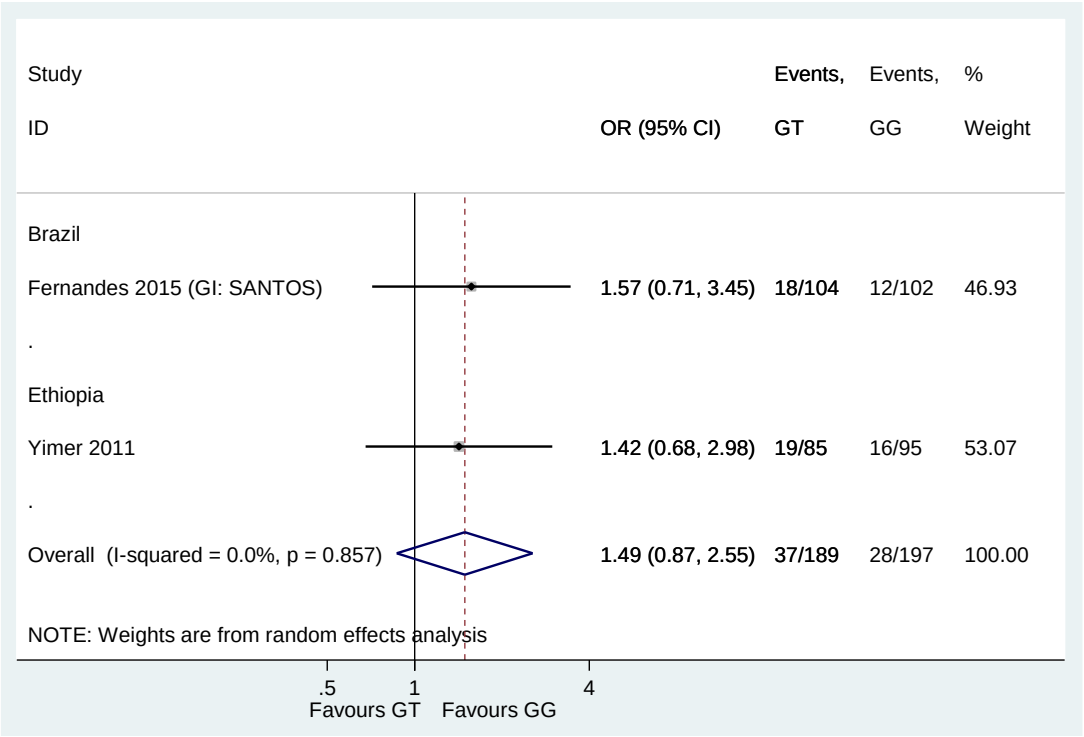

Figure 56 CYP2B6 *rs3745274 and hepatotoxicity: heterozygous (GT)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) versus *homozygous wild-type (GG).*

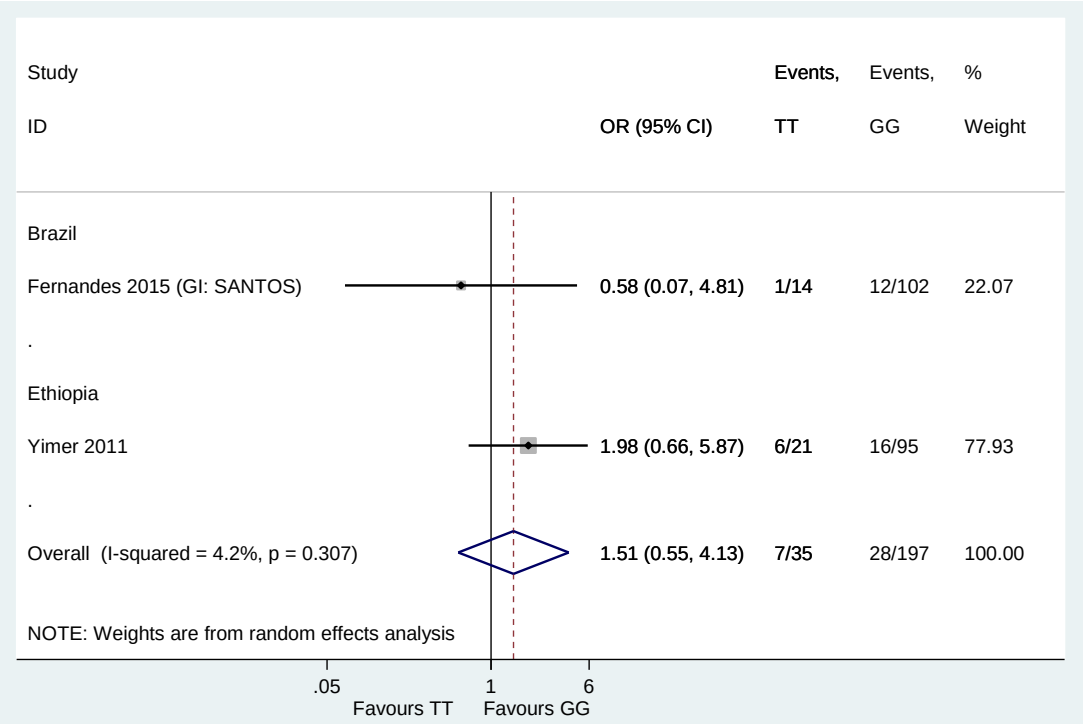

Figure 57 CYP2B6 *rs3745274 and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *ABCB1* rs1045642

Heterozygous (CT) versus *homozygous wild-type (CC)*

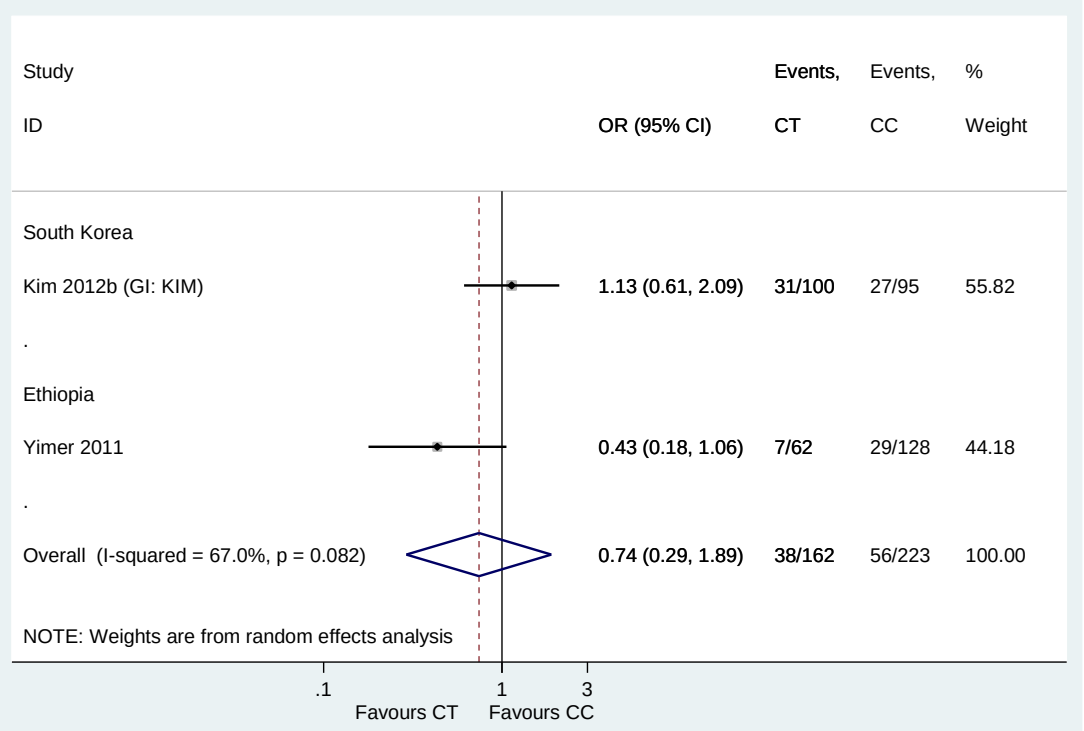

Figure 58 ABCB1 *rs1045642 and hepatotoxicity: heterozygous (CT)* versus *homozygous wild-type (CC)*

One of the studies (Kim 2012b) reports WT to be T and MT to be C, but the other study (Yimer 2011), and the data, suggest that WT is C and MT is T.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

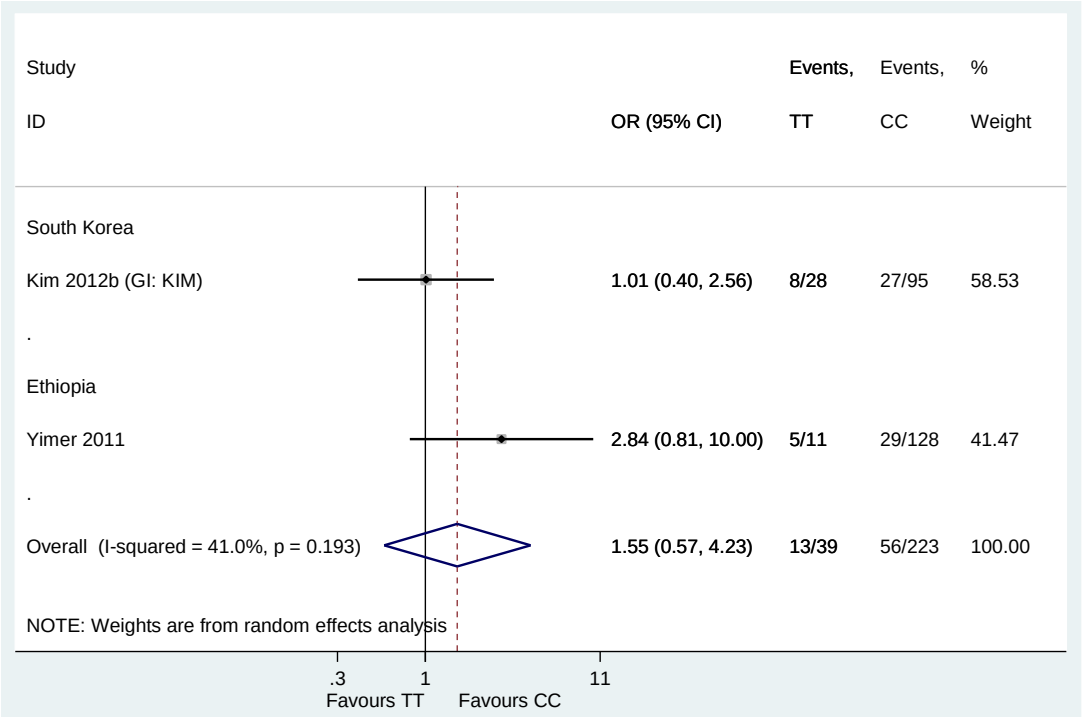

Figure 59 ABCB1 *rs1045642 and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

One of the studies (Kim 2012b) reports WT to be T and MT to be C, but the other study (Yimer 2011), and the data, suggest that WT is C and MT is T.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Pairwise comparisons for *NQO1* 609C-T (rs1800566)

Heterozygous (CT) versus *homozygous wild-type (CC)*

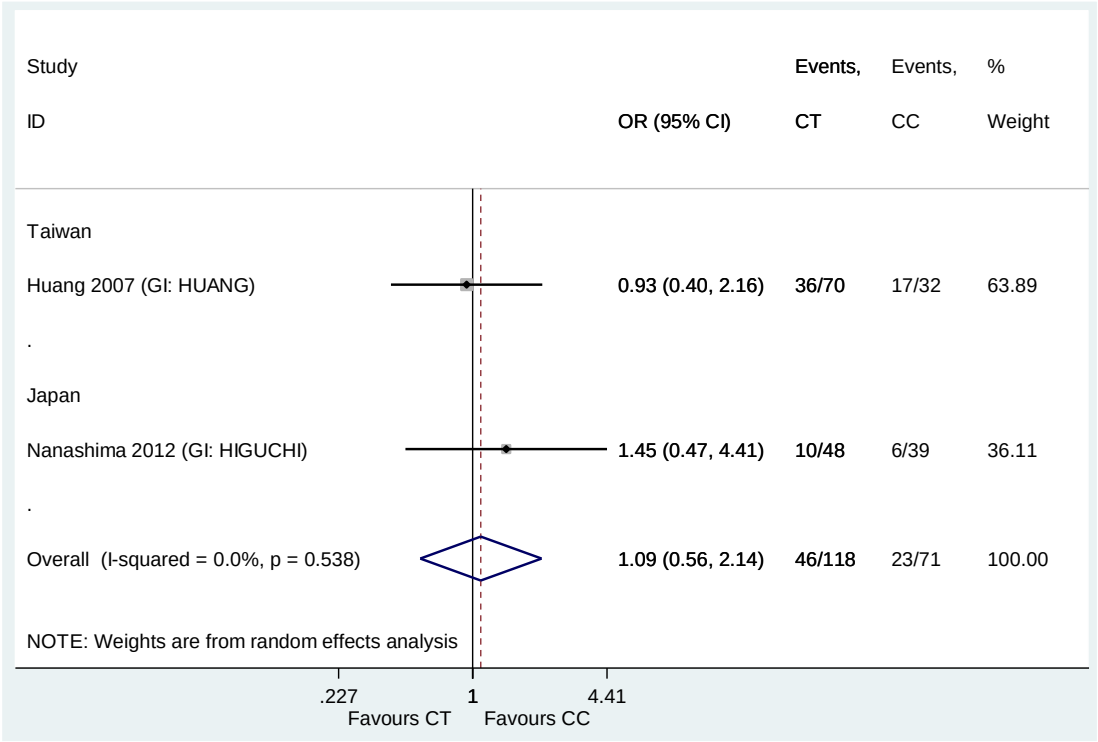

Figure 60 NQO1 *609C-T (rs1800566) and hepatotoxicity: heterozygous (CT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

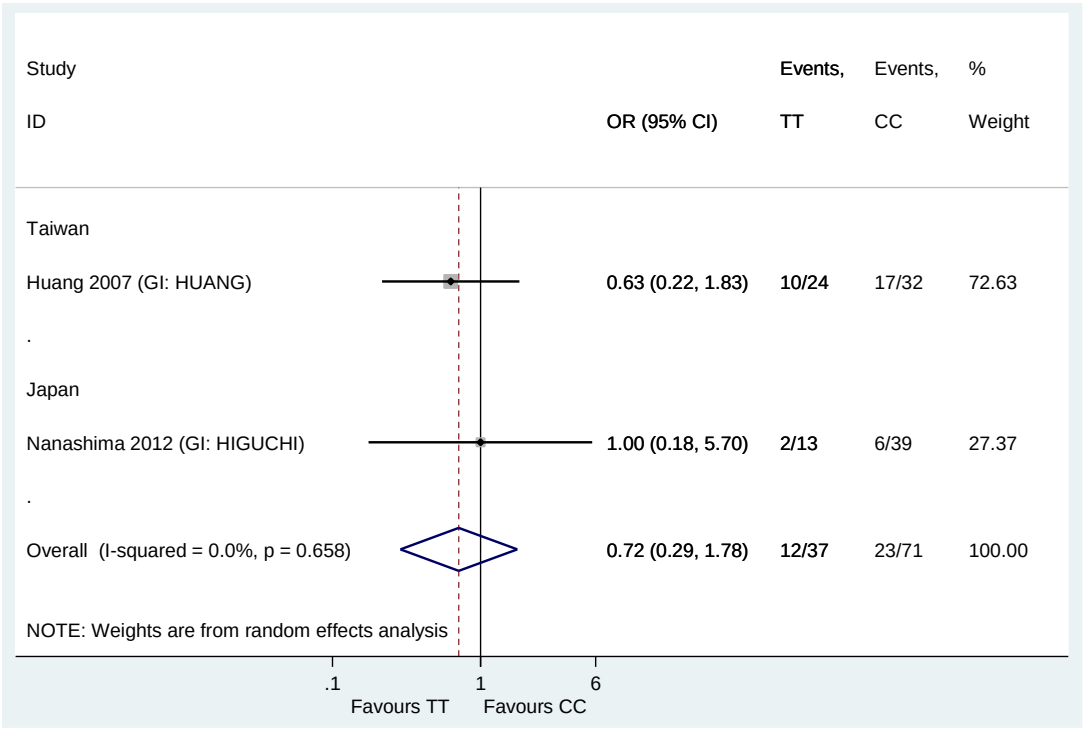

Figure 61 NQO1 *609C-T (rs1800566) and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *PXR* rs3814055

Heterozygous (CT) versus *homozygous wild-type (CC)*

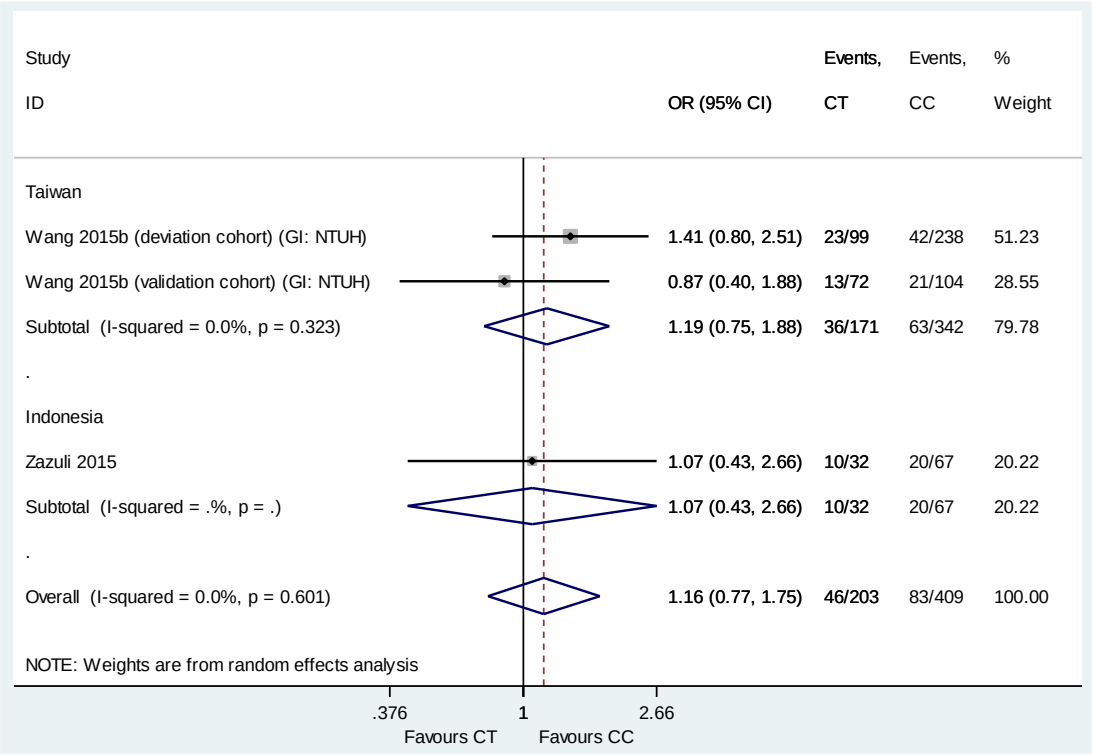

Figure 62 PXR *rs3814055 and hepatotoxicity: heterozygous (CT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

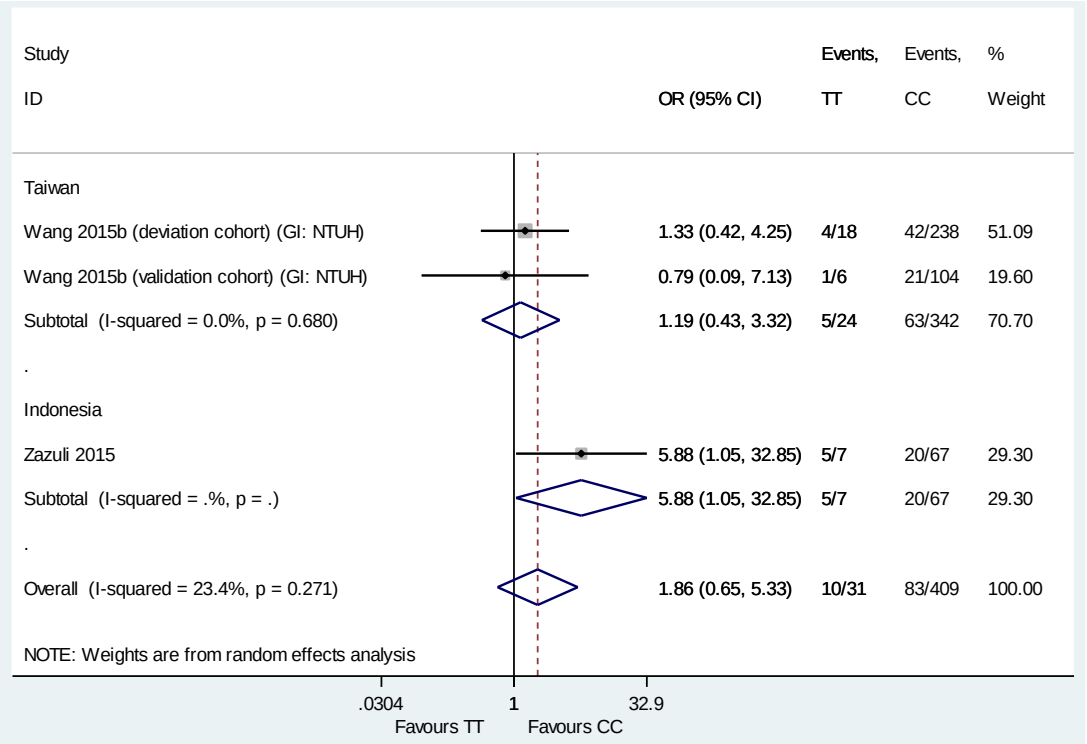

Figure 63 PXR *rs3814055 and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *PXR* rs12488820

Heterozygous (CT) versus *homozygous wild-type (CC)*

No meta-analysis was performed as no patients had heterozygous genotype in either patient cohort.

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

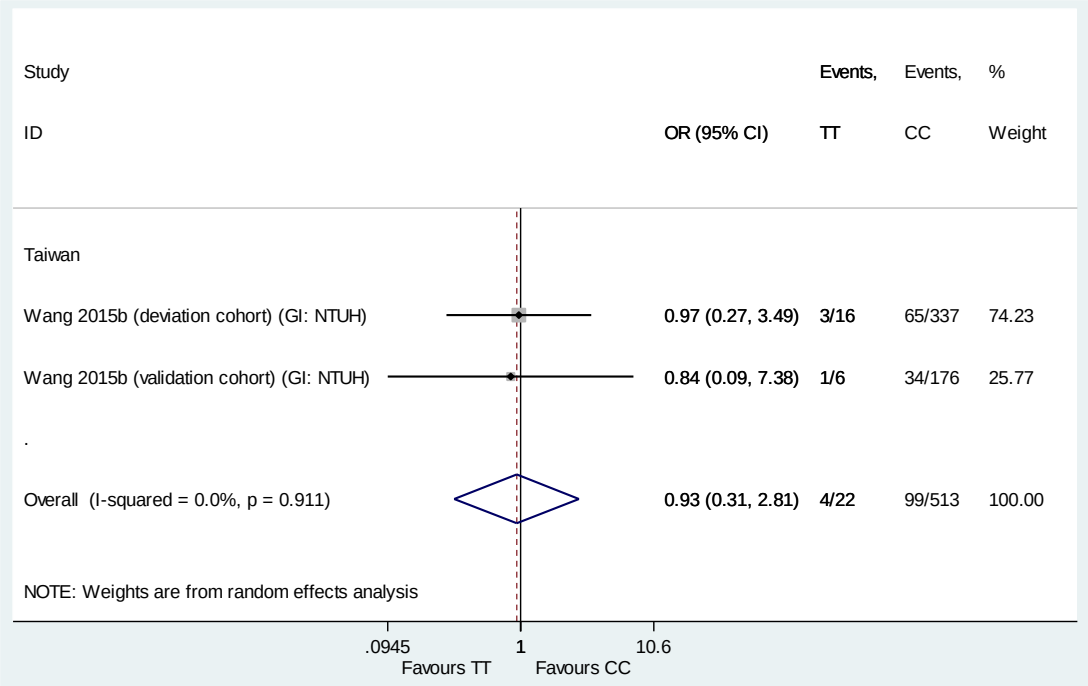

Figure 64 PXR *rs12488820 and hepatotoxicity: homozygous mutant-type (TT)* versus *homozygous wild-type (CC)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *PXR* rs2461823

Heterozygous (GA) versus *homozygous wild-type (GG)*

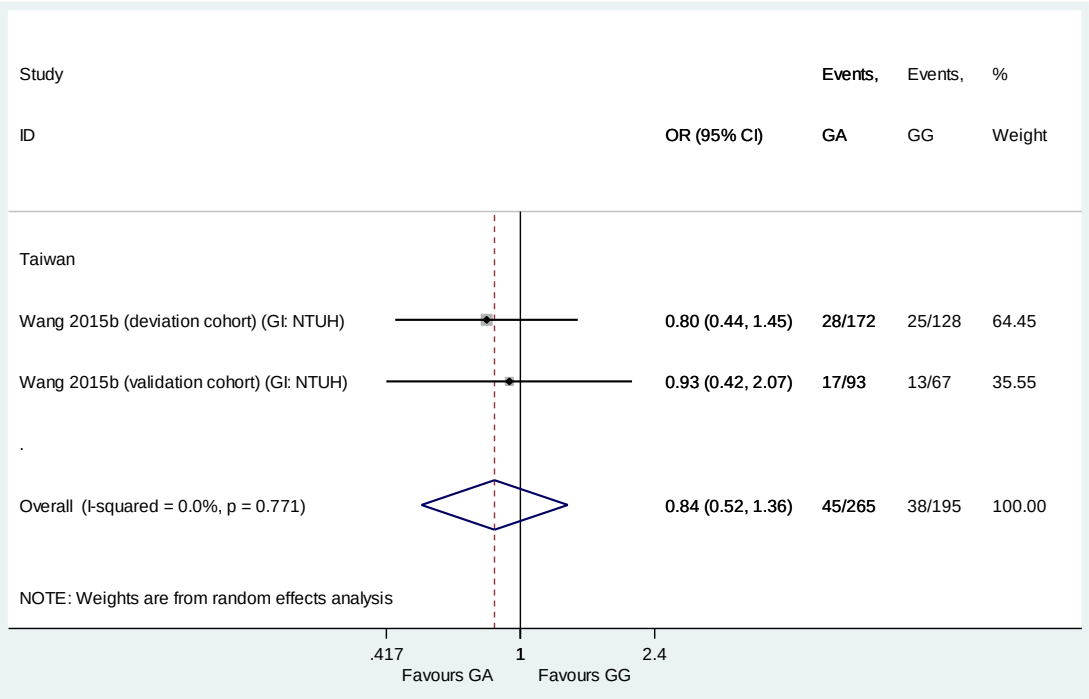

Figure 65 PXR *rs2461823 and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (AA) versus *homozygous wild-type (GG)*

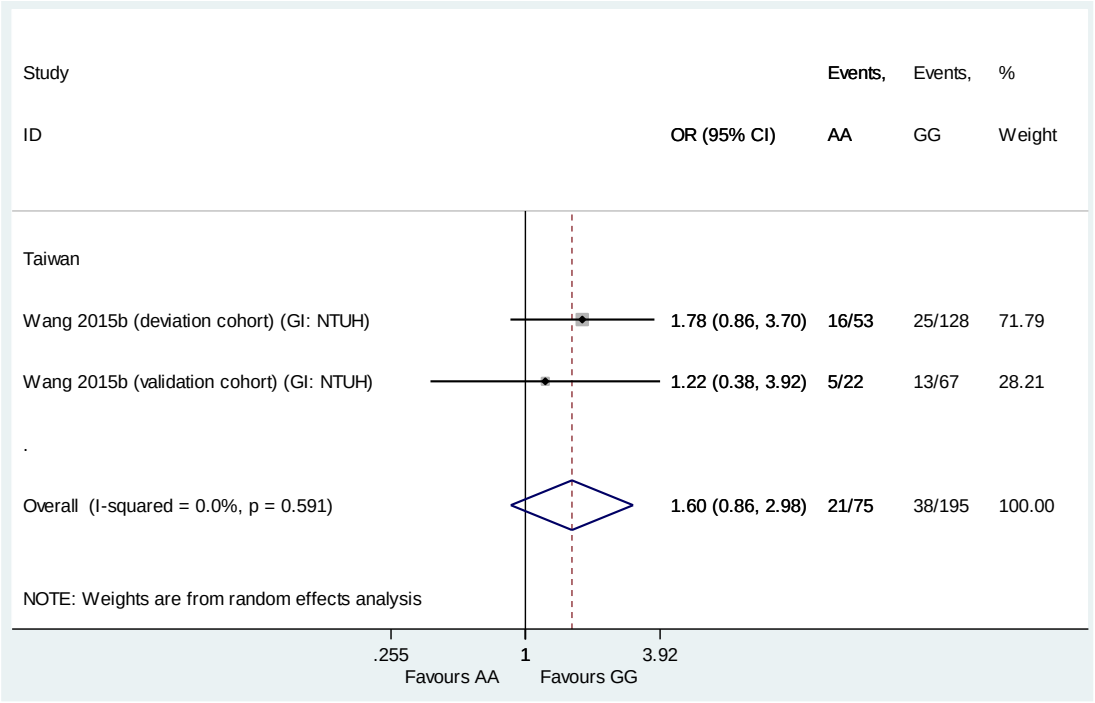

Figure 66 PXR *rs2461823 and hepatotoxicity: homozygous mutant-type (AA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *PXR* rs7643645

Heterozygous (AG) versus *homozygous wild-type (AA)*

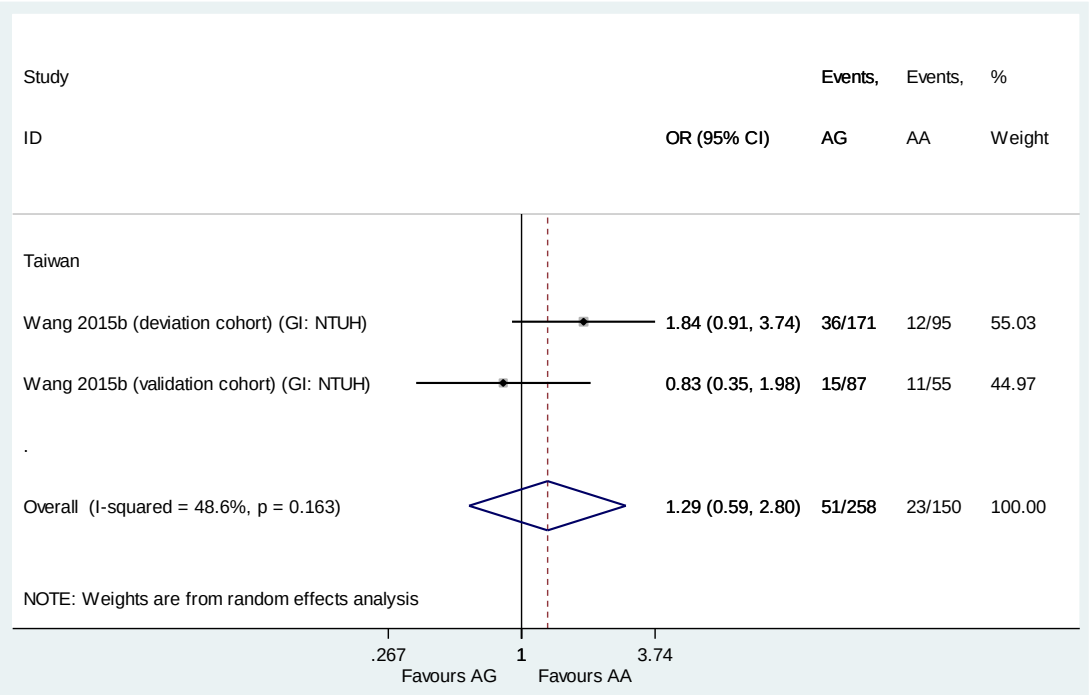

Figure 67 PXR *rs7643645 and hepatotoxicity: heterozygous (AG)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (GG) versus *homozygous wild-type (AA)*

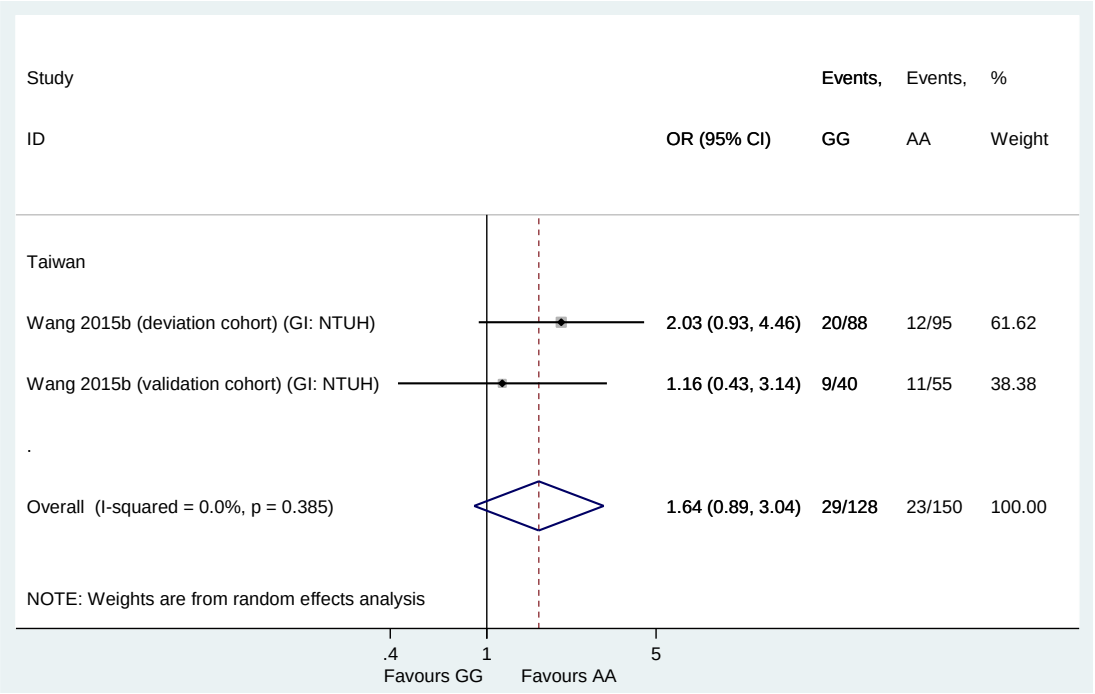

Figure 68 PXR *rs7643645 and hepatotoxicity: homozygous mutant-type (GG)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *PXR* rs6785049

Heterozygous (GA) versus *homozygous wild-type (GG)*

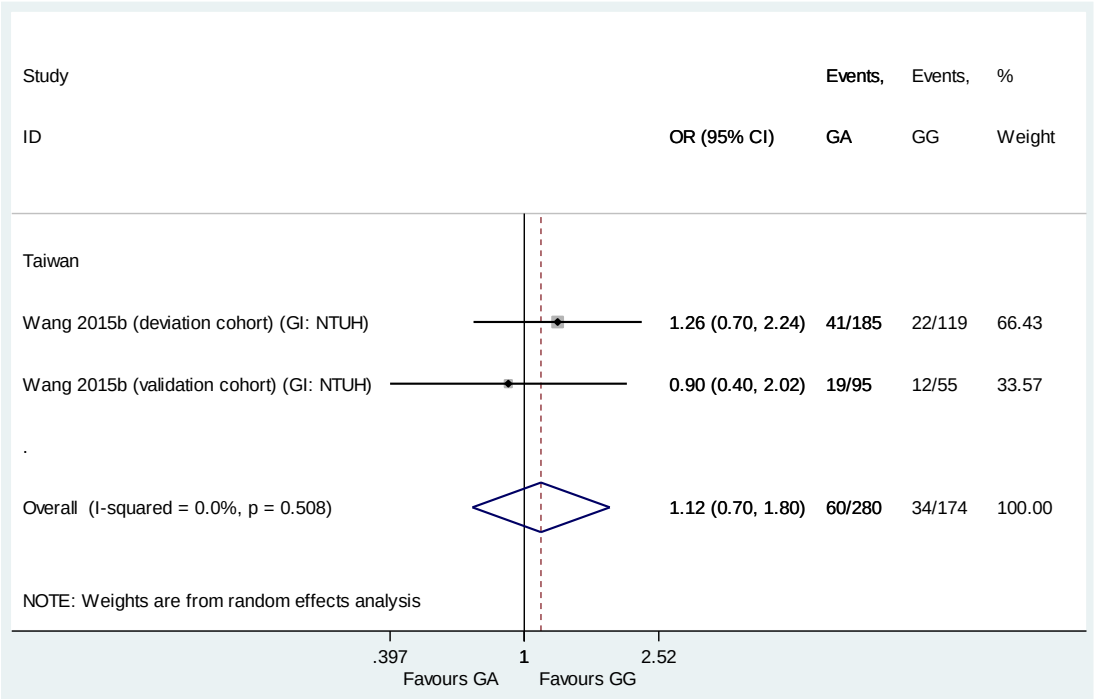

Figure 69 PXR *rs6785049 and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (AA) versus *homozygous wild-type (GG)*

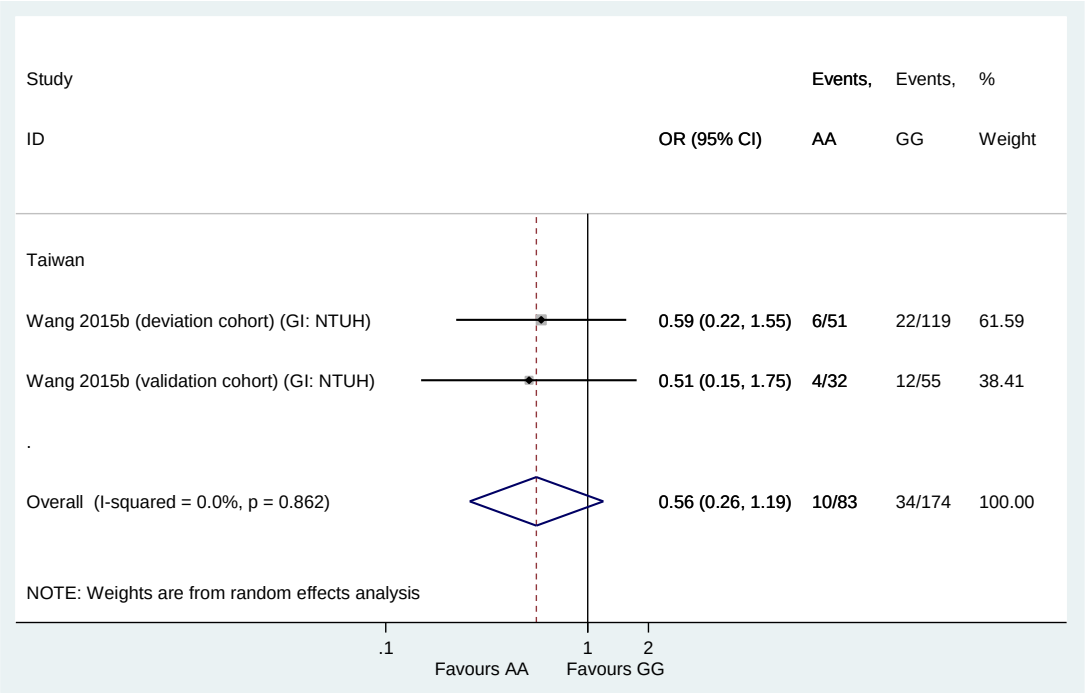

Figure 70 PXR *rs6785049 and hepatotoxicity: homozygous mutant-type (AA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *PXR* rs3814057

Heterozygous (AC) versus *homozygous wild-type (AA)*

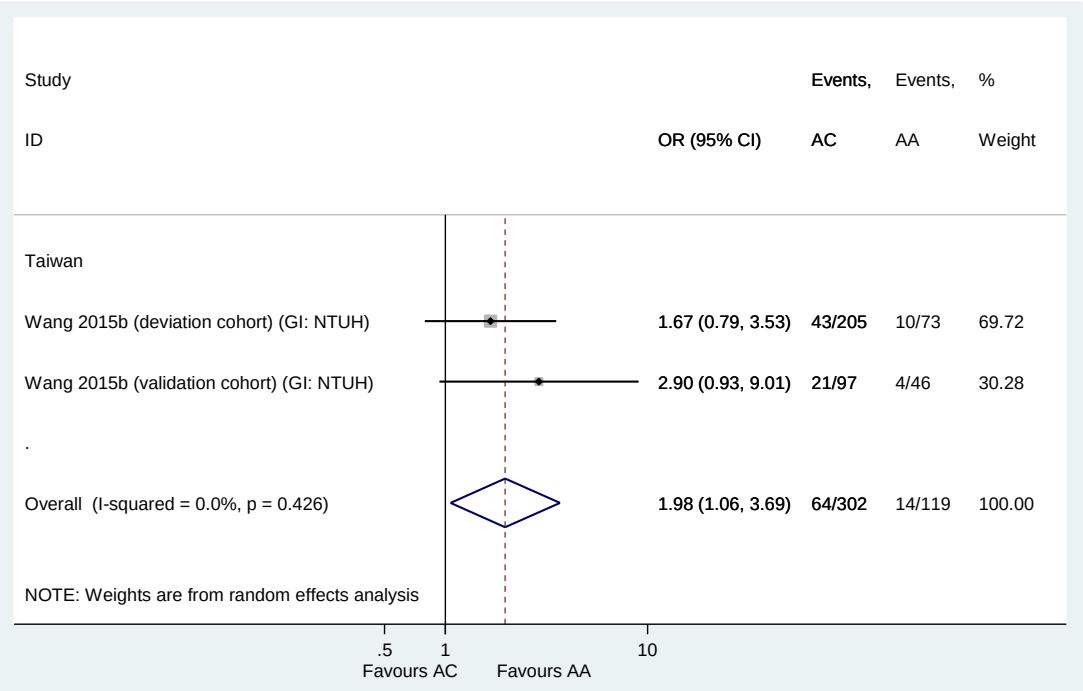

Figure 71 PXR *rs3814057 and hepatotoxicity: heterozygous (AC)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (CC) versus *homozygous wild-type (AA)*

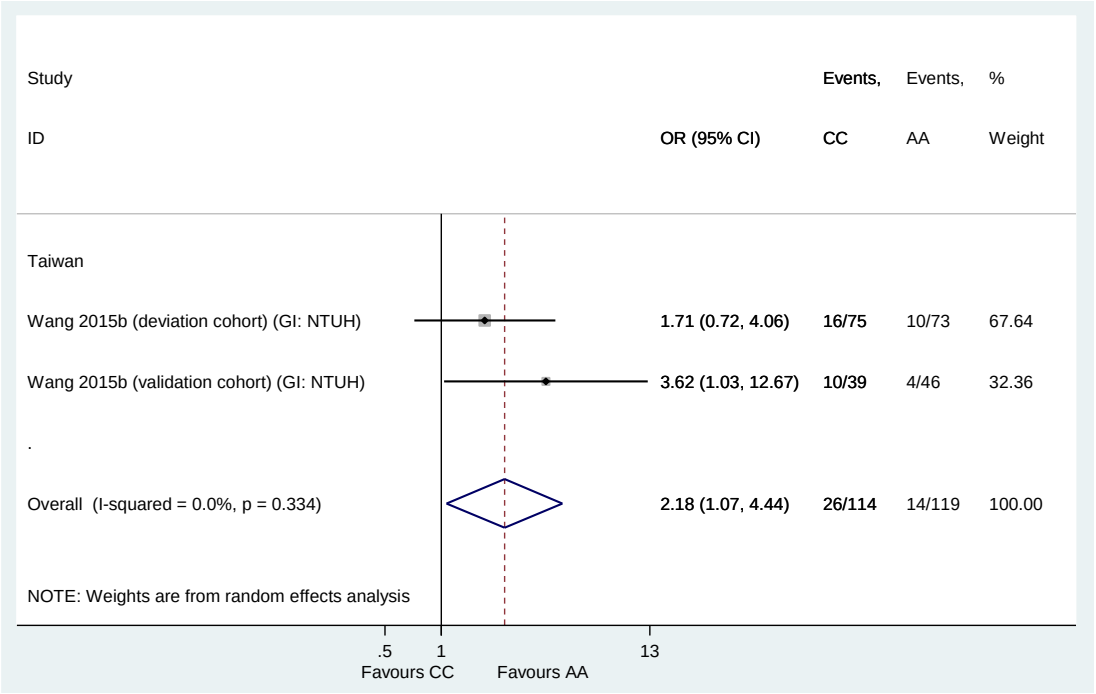

Figure 72 PXR *rs3814057 and hepatotoxicity: homozygous mutant-type (CC)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *SLCO1B1* rs4149013

Heterozygous (AG) versus *homozygous wild-type (AA)*

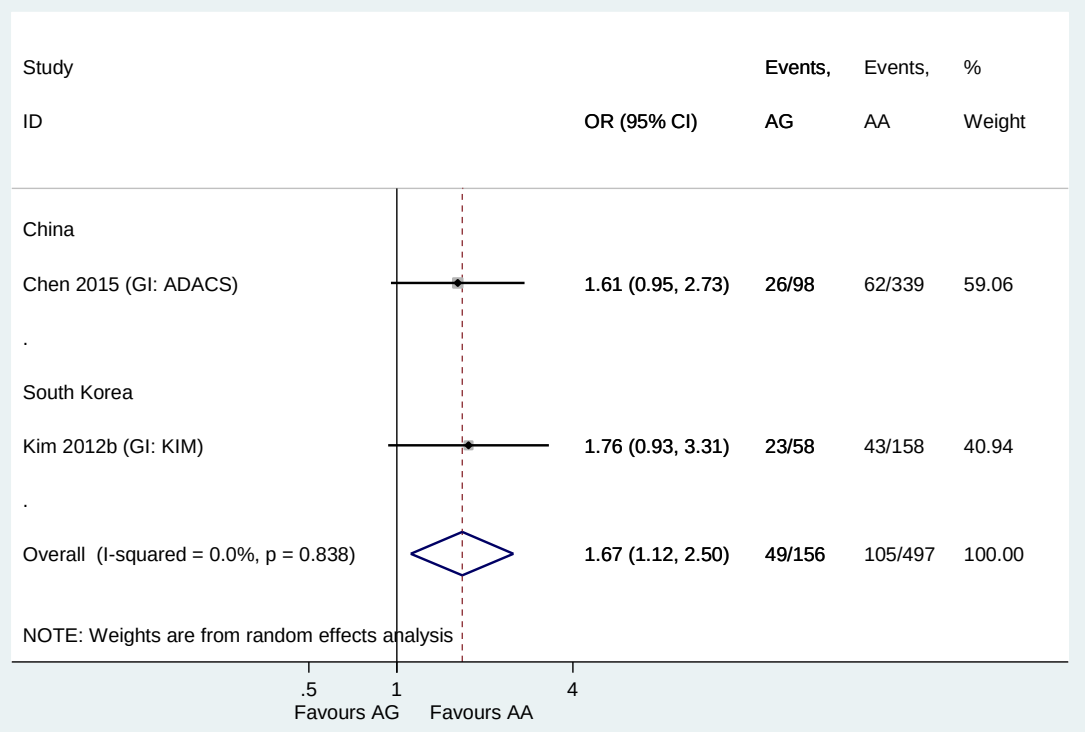

Figure 73 SLCO1B1 *rs4149013 and hepatotoxicity: heterozygous (AG)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (GG) versus *homozygous wild-type (AA)*

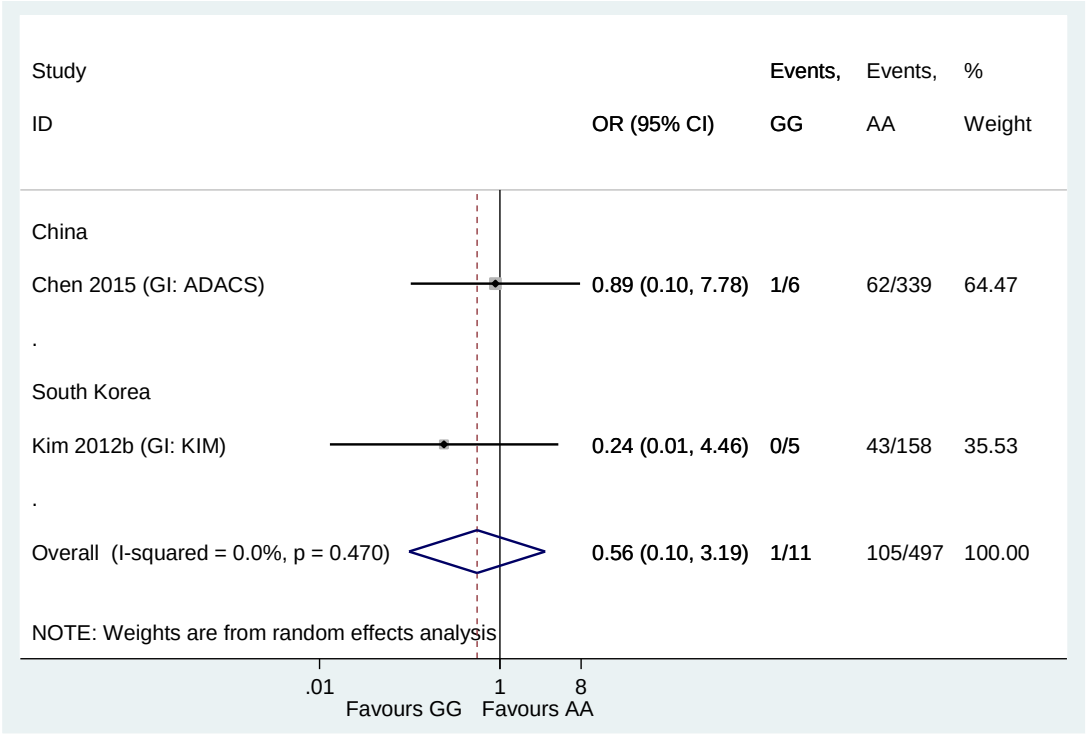

Figure 74 SLCO1B1 *rs4149013 and hepatotoxicity: homozygous mutant-type (GG)* versus *homozygous wild-type (AA)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *SLCO1B1* rs4149014

Heterozygous (GT) versus *homozygous wild-type (TT)*

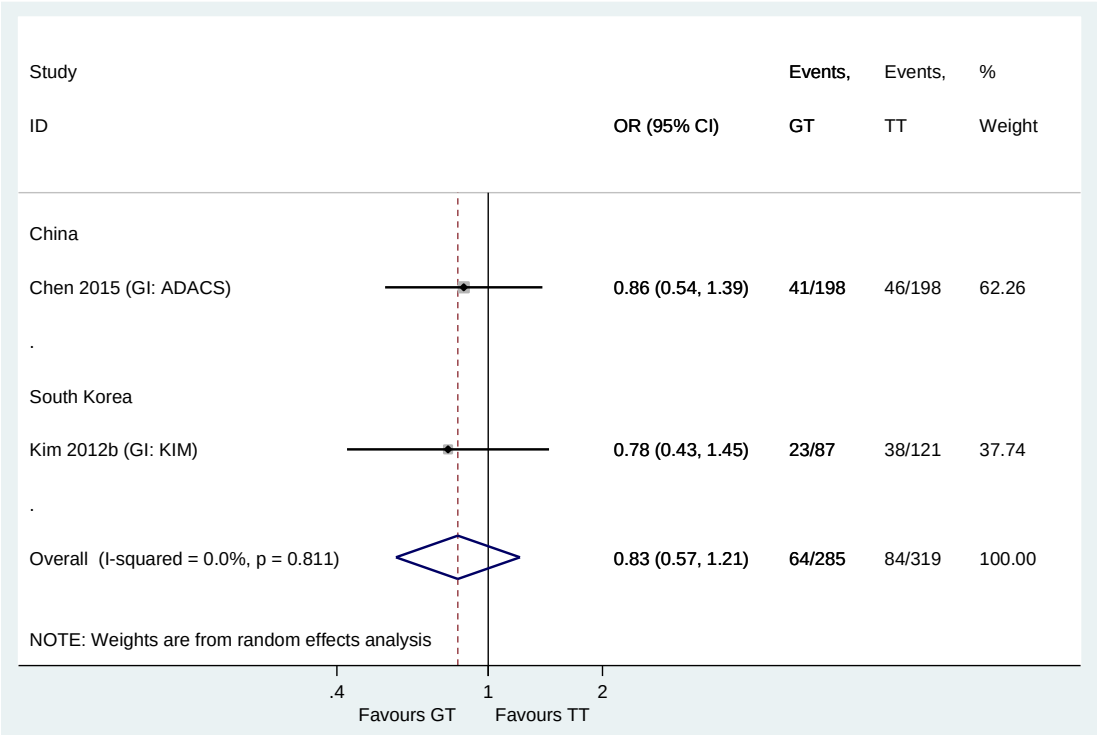

Figure 75 SLCO1B1 *rs4149014 and hepatotoxicity: heterozygous (GT)* versus *homozygous wild-type (TT)*

One of the studies (Kim 2012b) reports WT to be G and MT to be T, but the other study (Chen 2015), and the data, suggest that WT is T and MT is G.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Homozygous mutant-type (GG) versus *homozygous wild-type (TT)*

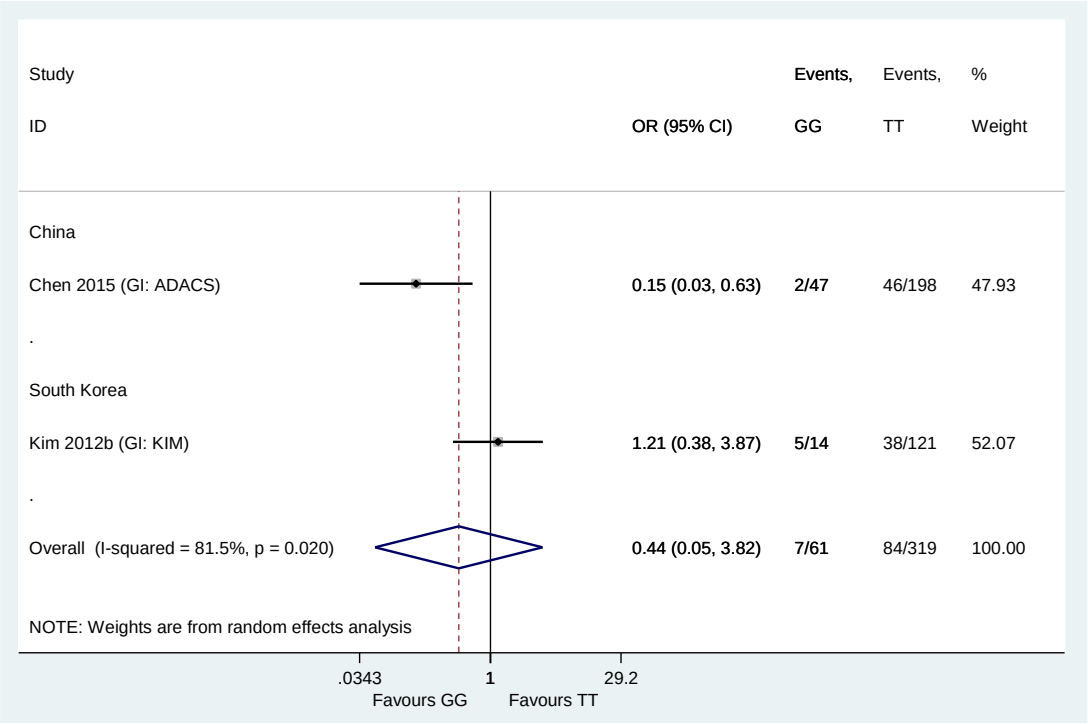

Figure 76 SLCO1B1 *rs4149014 and hepatotoxicity: homozygous mutant-type (GG)* versus *homozygous wild-type (TT)*

One of the studies (Kim 2012b) reports WT to be G and MT to be T, but the other study (Chen 2015), and the data, suggest that WT is T and MT is G.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Pairwise comparisons for *SLCO1B1* rs2306283

Heterozygous (GA) versus *homozygous wild-type (GG)*

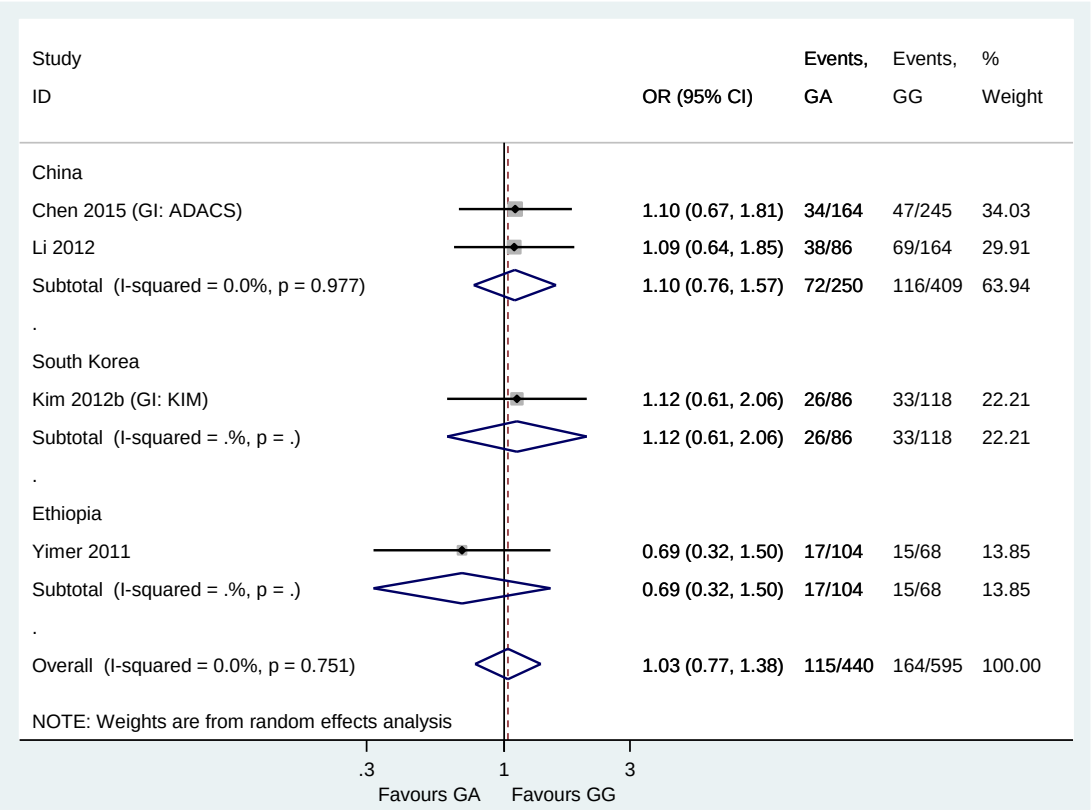

Figure 77 SLCO1B1 *rs2306283 and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (GG)*

Three of the studies (Chen 2015, Li 2012 and Yimer 2011) report WT to be A and MT to be G, but the other study (Kim 2012b), and the data, suggest that WT is G and MT is A.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Homozygous mutant-type (AA) versus *homozygous wild-type (GG)*

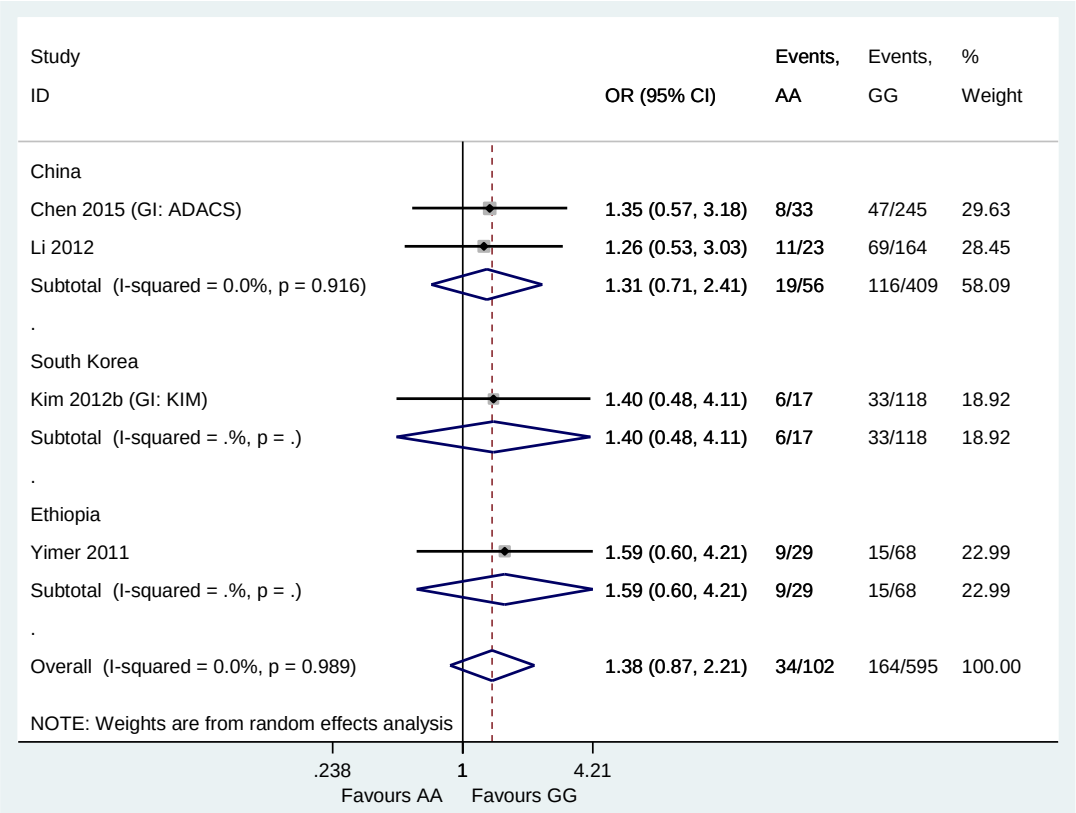

Figure 78 SLCO1B1 *rs2306283 and hepatotoxicity: homozygous mutant-type (AA)* versus *homozygous wild-type (GG)*

3 of the studies (Chen 2015, Li 2012 and Yimer 2011) report WT to be A and MT to be G, but the other study (Kim 2012b), and the data, suggest that WT is G and MT is A.
Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; MT: mutant-type; OR: odds ratio; WT: wild-type

Pairwise comparisons for *SLCO1B1* rs4149056

Heterozygous (TC) versus *homozygous wild-type (TT)*

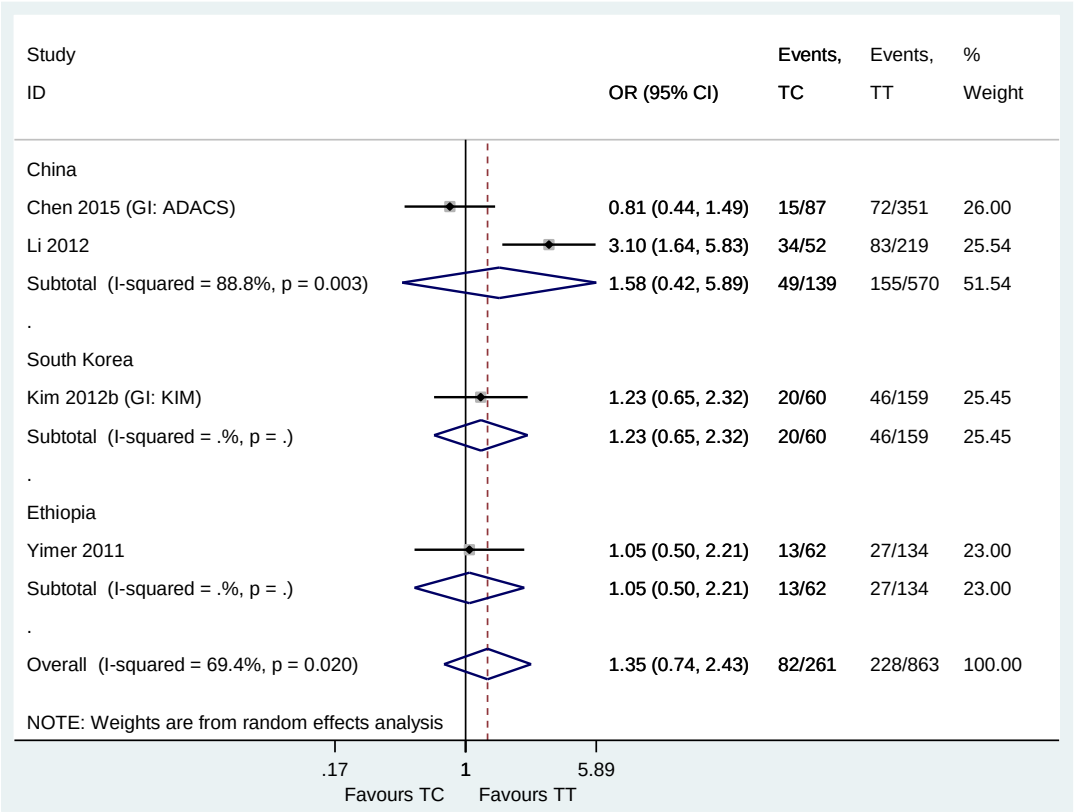

Figure 79 SLCO1B1 *rs4149056 and hepatotoxicity: heterozygous (TC)* versus *homozygous wild-type (TT)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (CC) versus *homozygous wild-type (TT)*

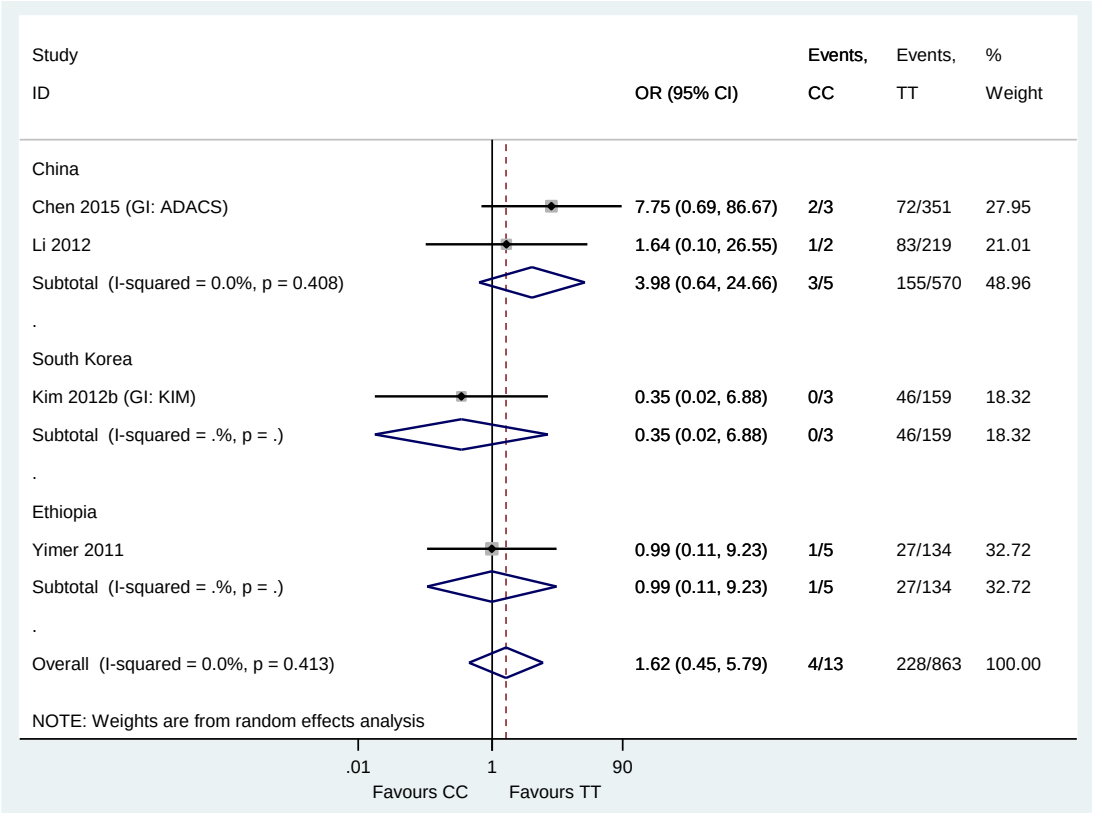

Figure 80 SLCO1B1 *rs4149056 and hepatotoxicity: homozygous mutant-type (CC)* versus *homozygous wild-type (TT)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

SOD2 rs4880: Homozygous mutant-type (CC) or heterozygous (CT) *versus* homozygous wild-type (TT)

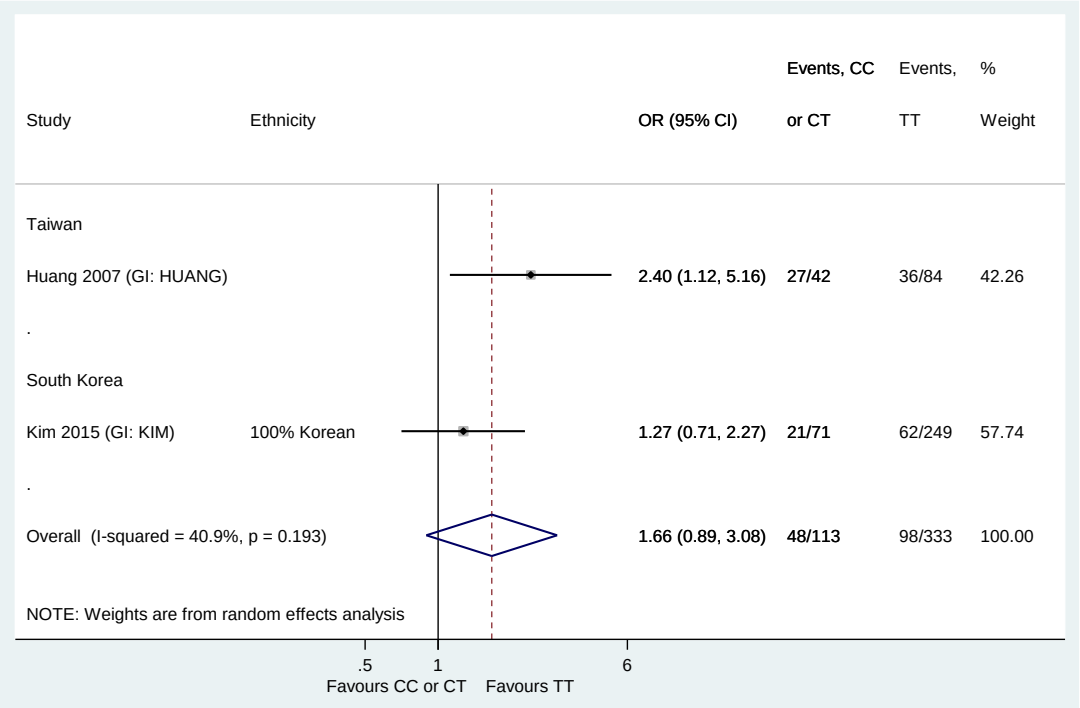

Figure 81 SOD2 *rs4880 and hepatotoxicity: homozygous mutant-type (CC) or heterozygous (CT)* versus *homozygous wild-type (TT)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Pairwise comparisons for *UGT1A1* rs4148323

Heterozygous (GA) versus *homozygous wild-type (GG)*

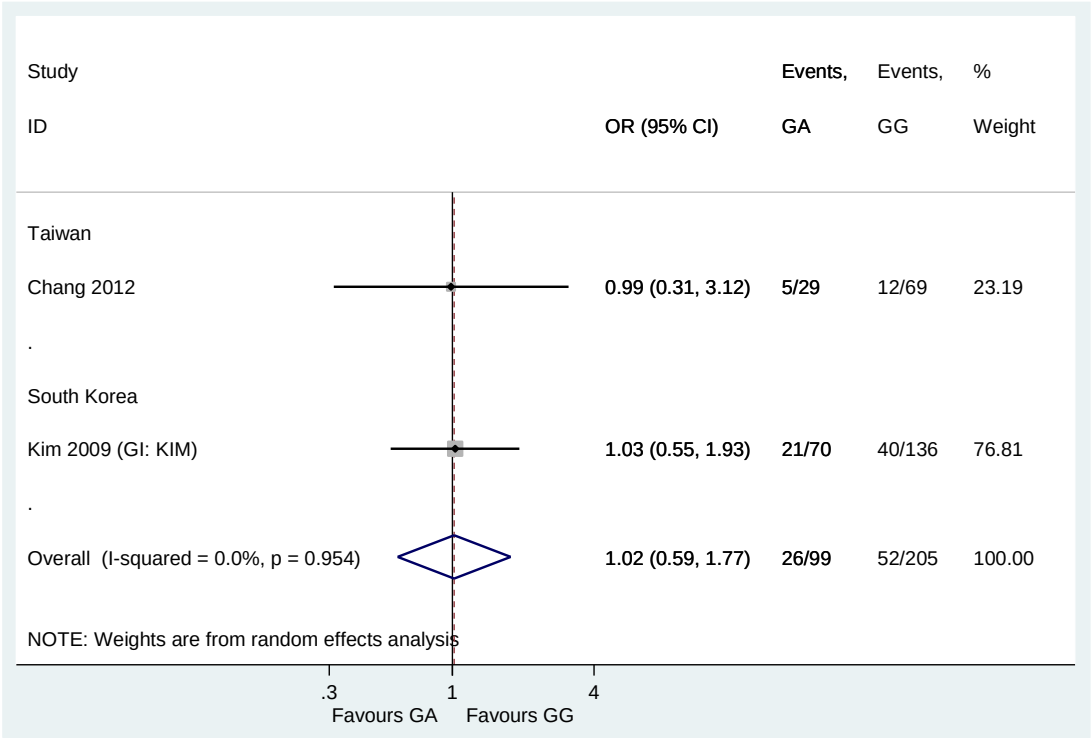

Figure 82 UGT1A1 *rs4148323 and hepatotoxicity: heterozygous (GA)* versus *homozygous wild-type (GG)*

Red dashed line indicates the value of the pooled effect estimate from the meta-analysis.
CI: confidence interval; GI: group identifier; OR: odds ratio

Homozygous mutant-type (TT) versus *homozygous wild-type (CC)*

No meta-analysis was performed as only one study identified patients with homozygous mutant-type genotype.

Appendix 5. Further analyses of the association between *NAT2* genetic variants and hepatotoxicity: Investigating the assumption of constant λ

NAT2 acetylator status

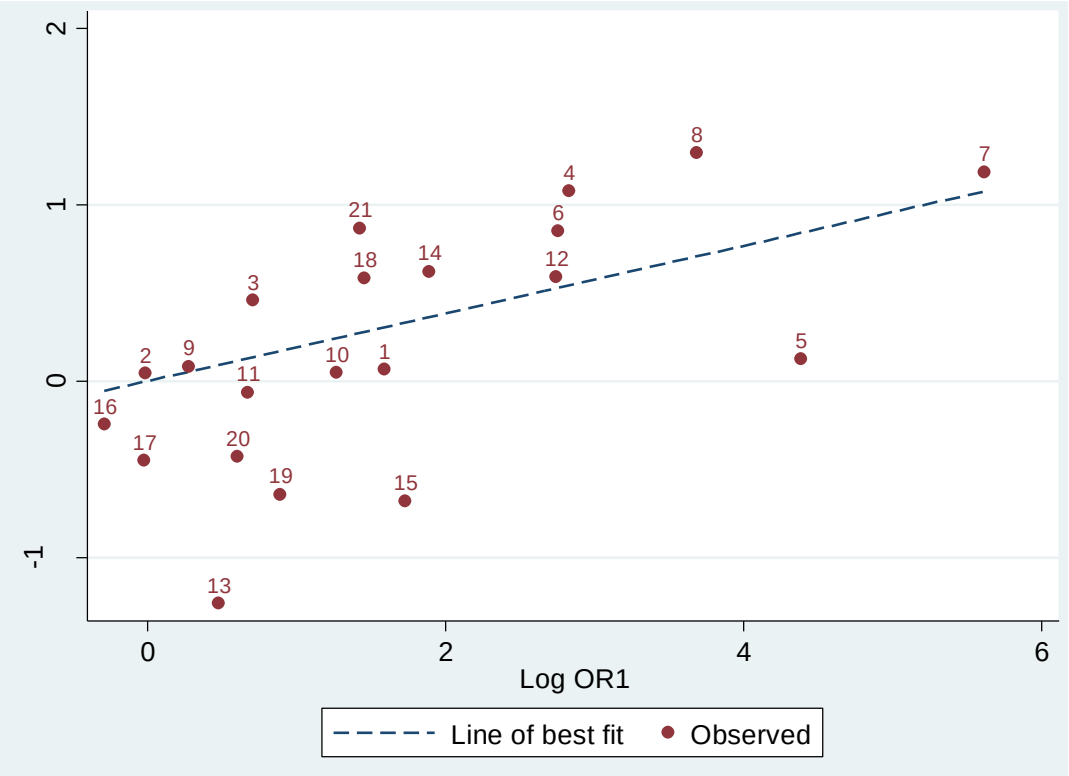

Figure 83 NAT2 *acetylator status: Graph showing the log OR for intermediate* versus *rapid acetylators against the log OR for slow* versus *rapid acetylators for each study*

Equation for line of best fit: y=0.19x
Labels for each study are provided in Table 43.
Log OR1: log odds ratio for intermediate *versus* rapid acetylators; log OR2: log odds ratio for slow *versus* rapid acetylators; OR: odds ratio

Table 43 Labels for each study included in Figure 83 and corresponding λ

Graph label	Study	λ
1	An 2012	0.04
2	Lv 2012 (GI: ADACS)	-3.79
3	Xiang 2014	0.64
4	Ho 2013	0.38
5	Azuma 2013	0.03
6	Higuchi 2007 (GI: HIGUCHI)	0.31
7	Ohno 2000	0.21
8	Shimizu 2006	0.35
9	Sotsuka 2011	0.28
10	Yuliwulandari 2016	0.04
11	Gupta 2013b (GI: GUPTA)	-0.10
12	Khalili 2011	0.21
13	Fredj 2016	-2.63
14	Mahmoud 2012	0.33
15	Çetintaş 2008	-0.40
16	Vuilleumier 2006	0.86
17	Leiro-Fernandez 2011 (GI: LEIRO)	24.72
18	Teixeira 2011	0.40
19	Santos 2013 (GI: SANTOS)	-0.72
20	Chamorro 2013	-0.71
21	Yamada 2009 (GI: YAMADA)	0.61

GI: group identifier

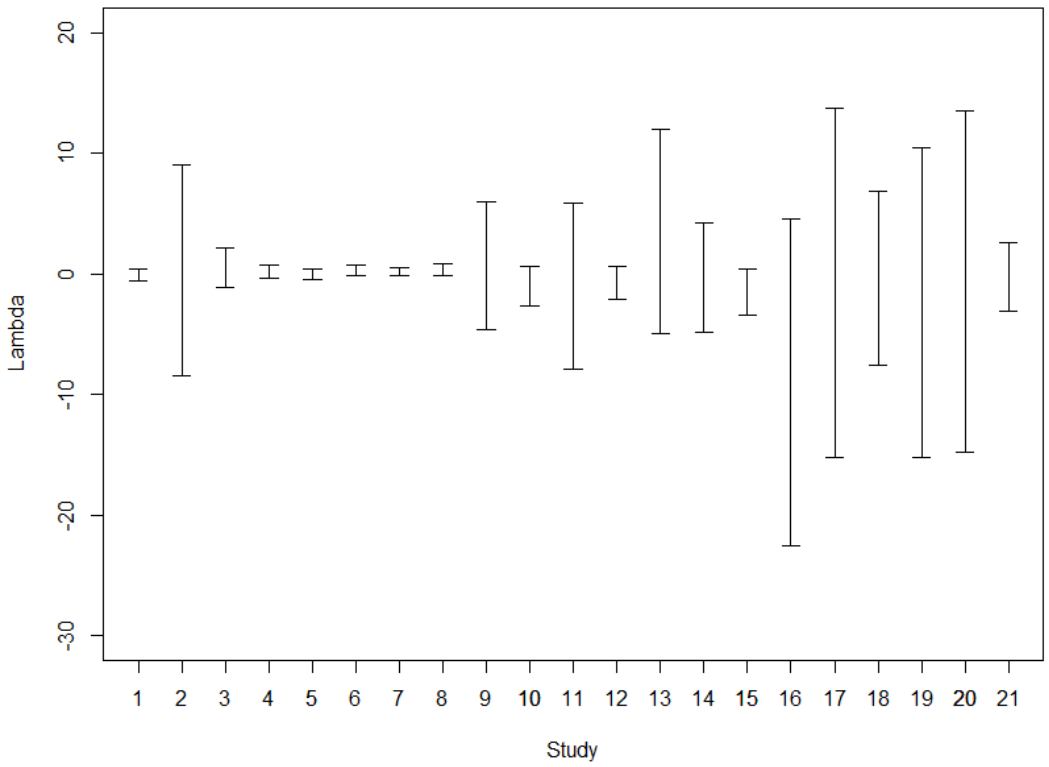

Figure 84 NAT2 *acetylator status: Graph showing the bootstrapped 95% CIs for λ for each study*

Labels for each study are provided in Table 44.
CI: confidence interval

Table 44 Labels for each study included in Figure 84 and corresponding limits of the bootstrapped 95% CI for λ

Graph label	Study	Lower limit for λ	Upper limit for λ
1	An 2012	-0.54	0.40
2	Lv 2012 (GI: ADACS)	-8.48	9.07
3	Xiang 2014	-1.13	2.10
4	Ho 2013	-0.40	0.70
5	Azuma 2013	-0.53	0.38
6	Higuchi 2007 (GI: HIGUCHI)	-0.20	0.74
7	Ohno 2000	-0.10	0.49
8	Shimizu 2006	-0.20	0.85
9	Sotsuka 2011	-4.65	6.00
10	Yuliwulandari 2016	-2.62	0.64
11	Gupta 2013b (GI: GUPTA)	-7.92	5.89
12	Khalili 2011	-2.16	0.57
13	Fredj 2016	-4.98	12.02
14	Mahmoud 2012	-4.87	4.19
15	Çetintaş 2008	-3.44	0.38
16	Vuilleumier 2006	-22.52	4.58
17	Leiro-Fernandez 2011 (GI: LEIRO)	-15.21	13.76
18	Teixeira 2011	-7.60	6.79
19	Santos 2013 (GI: SANTOS)	-15.19	10.42
20	Chamorro 2013	-14.83	13.50
21	Yamada 2009 (GI: YAMADA)	-3.08	2.60

CI: confidence interval; GI: group identifier

Interpretation

A reasonable amount of studies report data for this SNP, and the bootstrapped CIs overlap, so the assumption of constant λ seems reasonable here.

Conclusion

Perform genetic model-free analysis, and bivariate analysis to investigate the robustness of results.

NAT2 282C-T

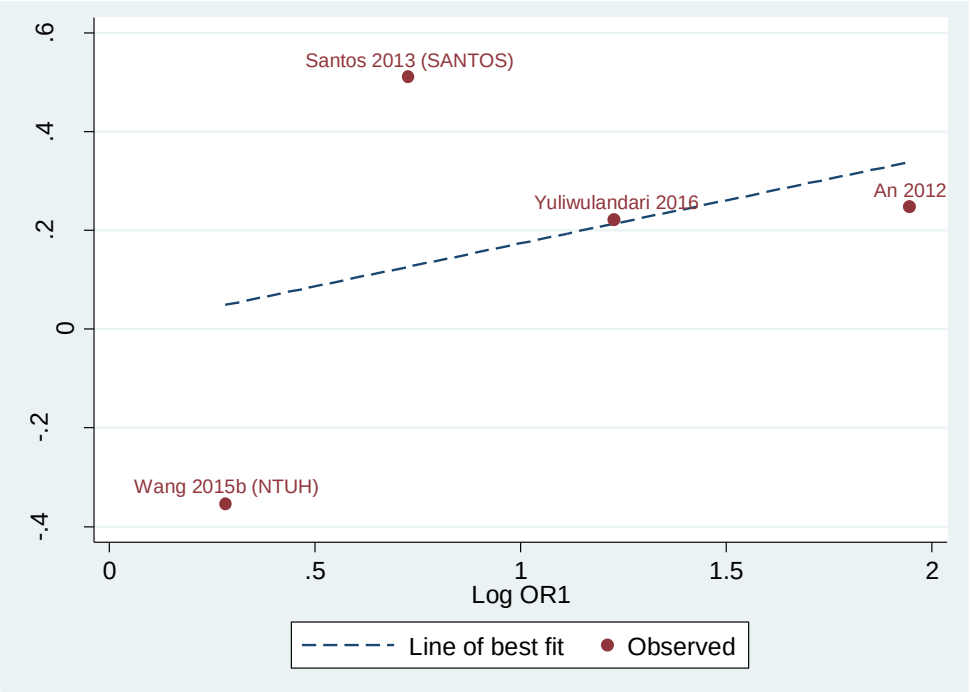

Figure 85 NAT2 *282C-T: Graph showing the log OR for CT* versus *CC against the log OR for TT* versus *CC for each study (all studies)*

Equation for line of best fit: y=0.17x
Log OR1: log odds ratio for CT *versus* CC; log OR2: log odds ratio for TT *versus* CC; OR: odds ratio

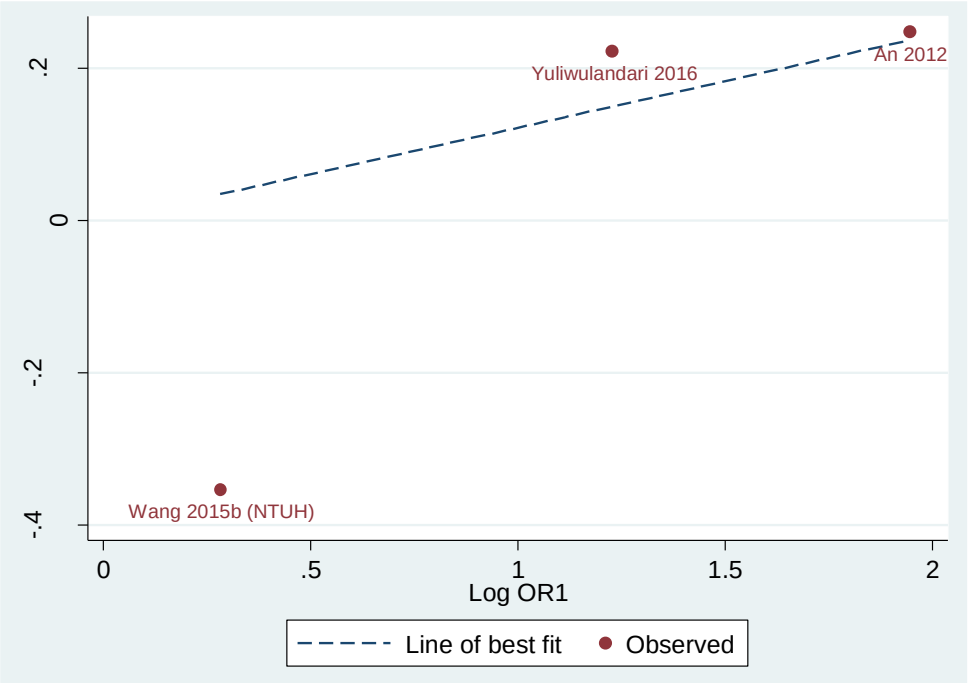

Figure 86 NAT2 *282C-T: Graph showing the log OR for CT* versus *CC against the log OR for TT* versus *CC for each study (sensitivity analysis excluding Santos 2013 due to deviation from HWE)*

Equation for line of best fit: y=0.12x
HWE: Hardy-Weinberg equilibrium; log OR1: log odds ratio for CT *versus* CC; log OR2: log odds ratio for TT *versus* CC; OR: odds ratio

Table 45 NAT2 *282C-T: λ for each study*

Study	λ
An 2012	0.13
Wang 2015b (GI: NTUH)	-1.25
Yuliwulandari 2016	0.18
Santos 2013 (GI: SANTOS)	0.70

GI: group identifier

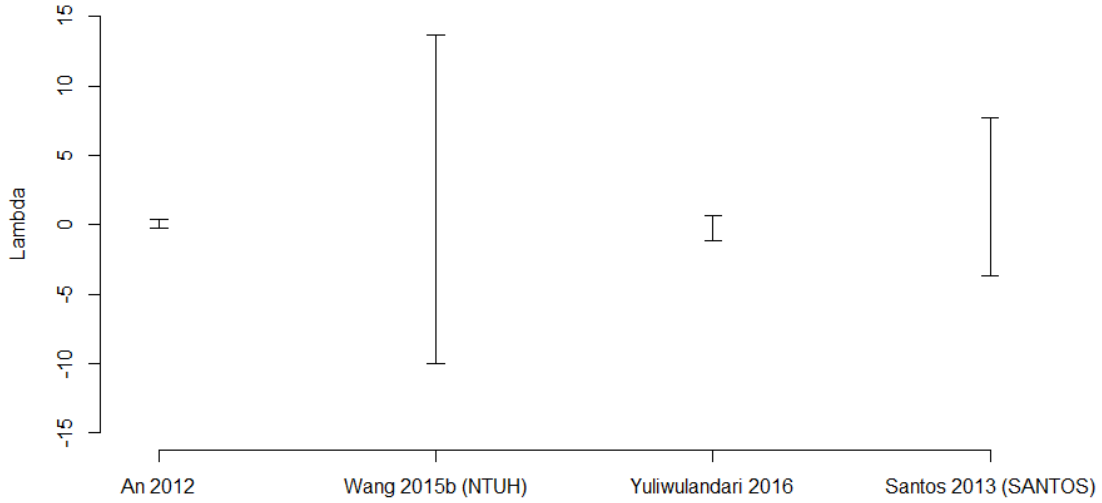

Figure 87 NAT2 *282C-T: Graph showing the bootstrapped 95% CIs for λ for each study*

CI: confidence interval

Table 46 NAT2 *282C-T: Limits of the bootstrapped 95% CI for λ for each study*

Study	Lower limit for λ	Upper limit for λ
An 2012	-0.26	0.42
Wang 2015b (GI: NTUH)	-9.96	13.67
Yuliwulandari 2016	-1.14	0.71
Santos 2013 (GI: SANTOS)	-3.66	7.73

CI: confidence interval; GI: group identifier

Interpretation

Although there is overlap in the bootstrapped CIs for λ, there are few studies and it is therefore difficult to assess the validity of the assumption of constant λ.

Conclusion

Perform both bivariate and genetic model-free analyses.

NAT2 341T-C

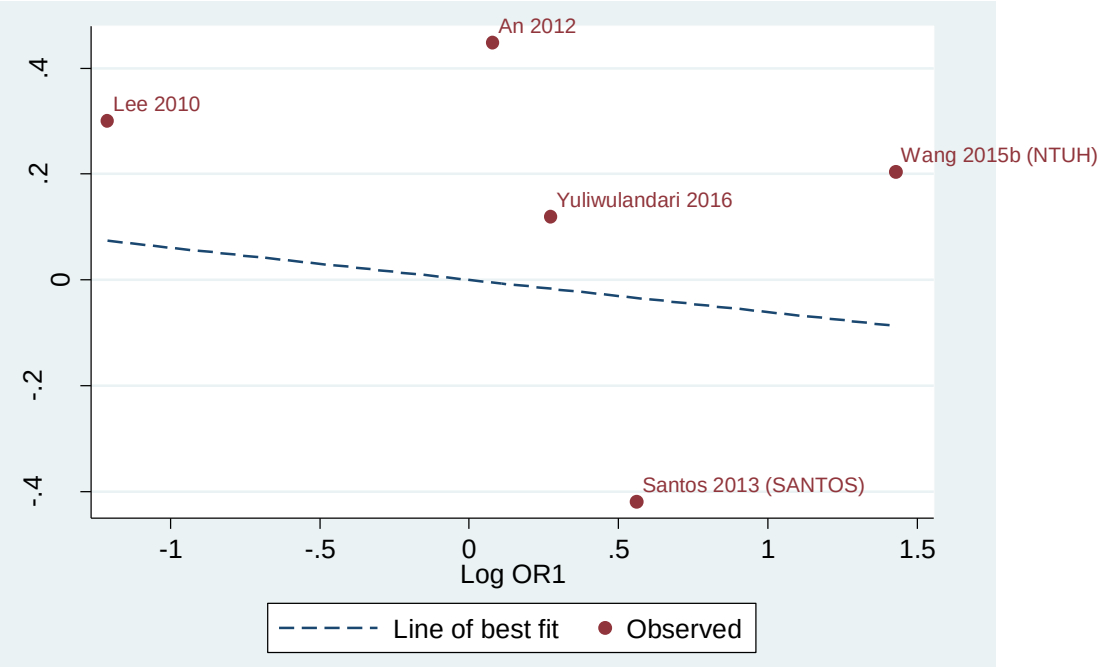

Figure 88 NAT2 *341T-C: Graph showing the log OR for TC* versus *TT against the log OR for CC* versus *TT for each study (all studies)*

Equation for line of best fit: y=-0.06x
Log OR1: log odds ratio for TC *versus* TT; log OR2: log odds ratio for CC *versus* TT; OR: odds ratio

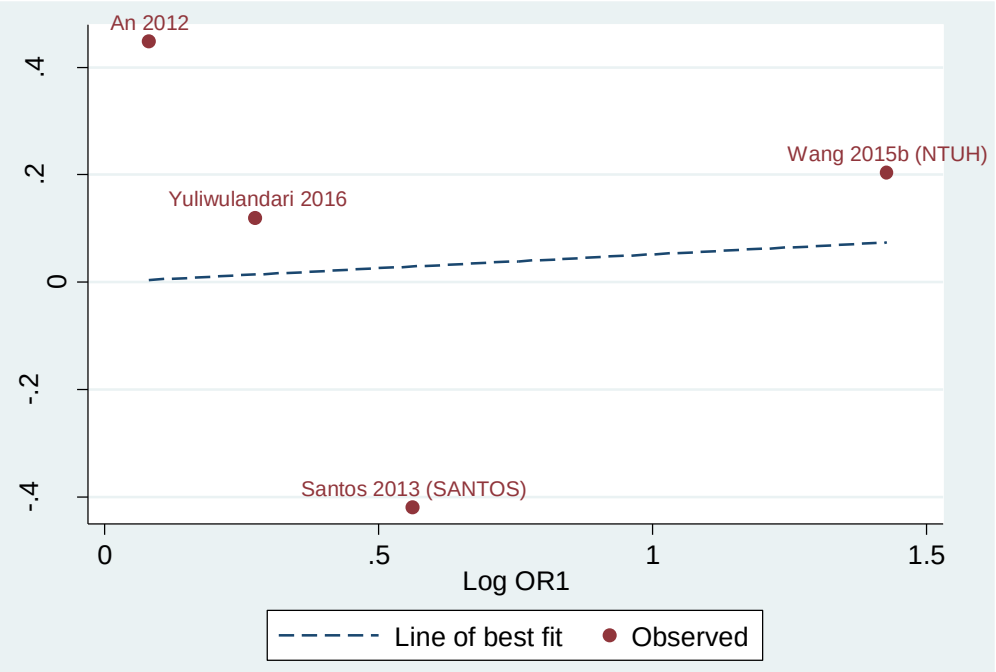

Figure 89 NAT2 *341T-C: Graph showing the log OR for TC* versus *TT against the log OR for CC* versus *TT for each study (sensitivity analysis excluding Lee 2010 due to deviation from HWE)*

Equation for line of best fit: y=0.05x
HWE: Hardy-Weinberg equilibrium; log OR1: log odds ratio for TC *versus* TT; log OR2: log odds ratio for CC *versus* TT; OR: odds ratio

Table 47 NAT2 *341T-C: λ for each study*

Study	λ
An 2012	5.57
Lee 2010	-0.25
Wang 2015b (GI: NTUH)	0.14
Yuliwulandari 2016	0.44
Santos 2013 (GI: SANTOS)	-0.75

GI: group identifier

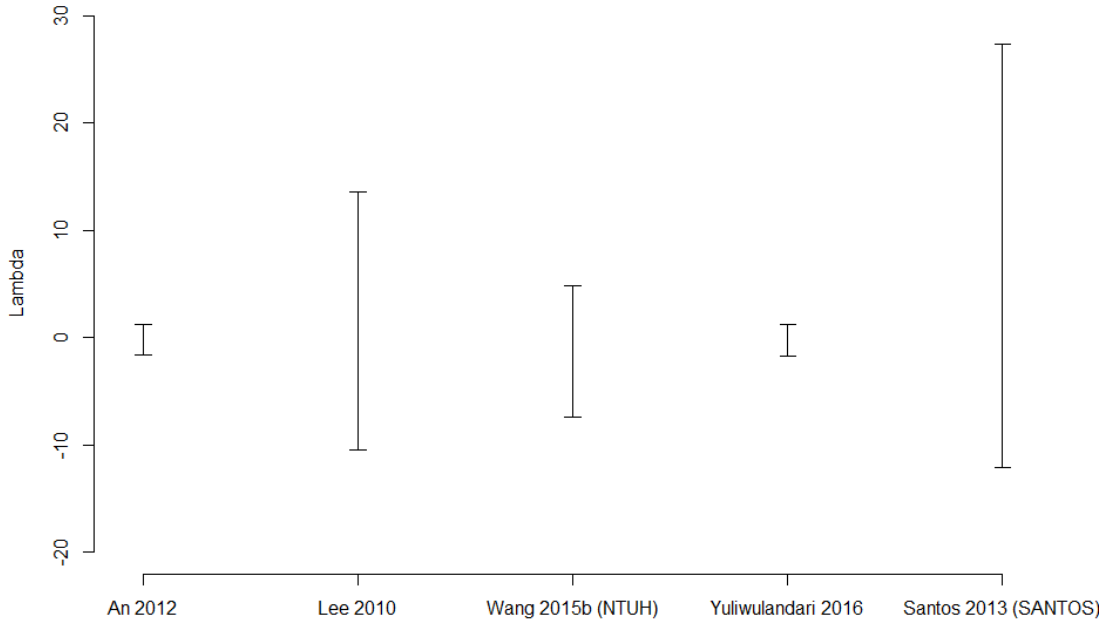

Figure 90 NAT2 *341T-C: Graph showing the bootstrapped 95% CIs for λ for each study*

CI: confidence interval

Table 48 NAT2 *341T-C: Limits of the bootstrapped 95% CI for λ for each study*

Study	Lower limit for λ	Upper limit for λ
An 2012	-1.62	1.22
Lee 2010	-10.43	13.60
Wang 2015b (GI: NTUH)	-7.38	4.86
Yuliwulandari 2016	-1.68	1.18
Santos 2013 (GI: SANTOS)	-12.07	27.30

CI: confidence interval; GI: group identifier

Interpretation

Although there is overlap in the bootstrapped CIs for λ, there are few studies and considerable variability in the estimates of λ from each of these studies, so the validity of the assumption of constant λ is unclear.

Conclusion

Perform both bivariate and genetic model-free analyses.

NAT2 481C-T

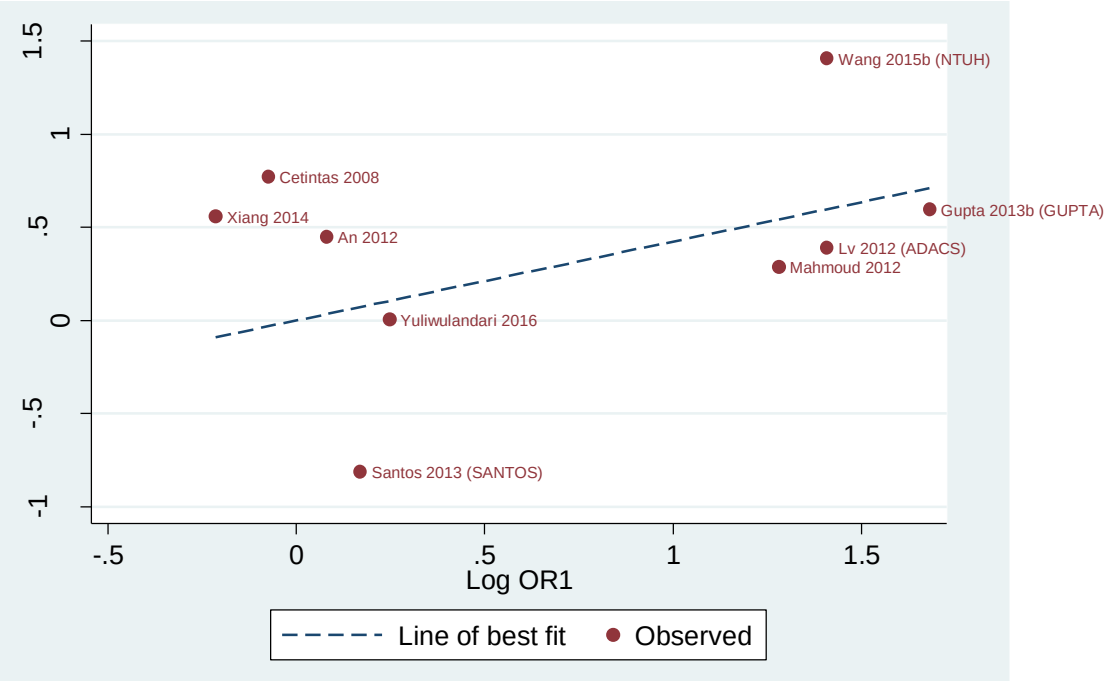

Figure 91 NAT2 *481C-T: Graph showing the log OR for CT* versus *CC against the log OR for TT* versus *CC for each study*

Equation for line of best fit: y=0.42x
Log OR1: log odds ratio for CT *versus* CC; log OR2: log odds ratio for TT *versus* CC; OR: odds ratio

Table 49 NAT2 *481C-T: λ for each study*

Study	λ
An 2012	5.57
Lv 2012 (GI: ADACS)	0.28
Xiang 2014	-2.61
Wang 2015b (GI: NTUH)	1.00
Yuliwulandari 2016	0.03
Gupta 2013b (GI: GUPTA)	0.36
Mahmoud 2012	0.22
Çetintaş 2008	-10.43
Santos 2013 (GI: SANTOS)	-4.77

GI: group identifier

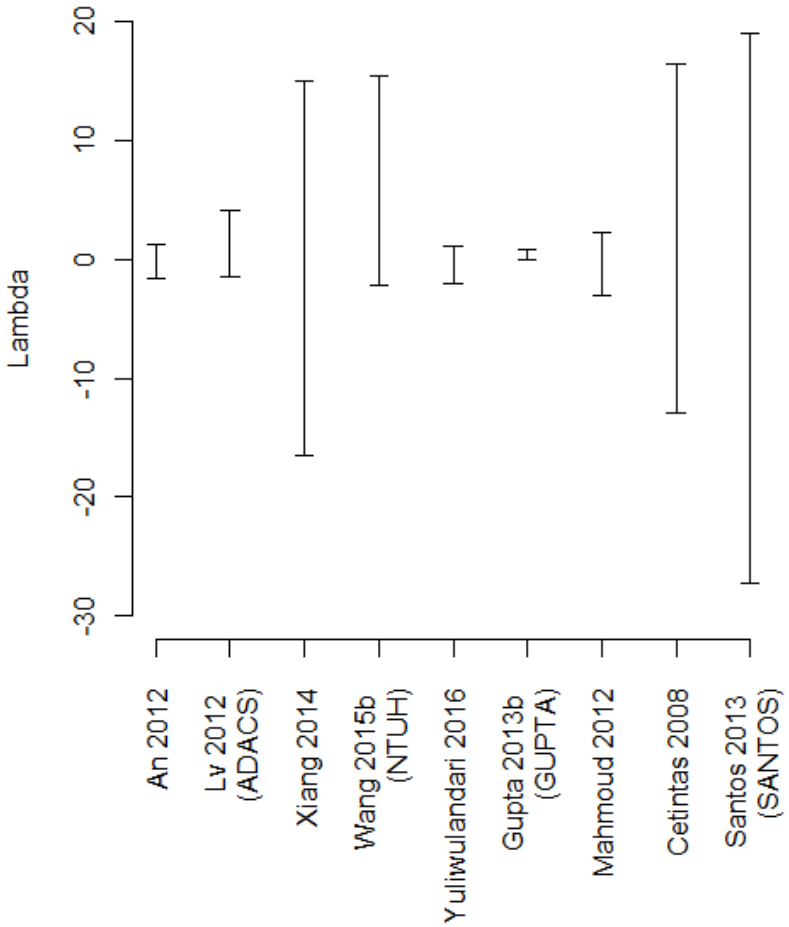

Figure 92 NAT2 *481C-T: Graph showing the bootstrapped 95% CIs for λ for each study*

CI: confidence interval

Table 50 NAT2 *481C-T: Limits of the bootstrapped 95% CI for λ for each study*

Study	Lower limit for λ	Upper limit for λ
An 2012	-1.62	1.22
Lv 2012 (GI: ADACS)	-1.44	4.08
Xiang 2014	-16.51	15.04
Wang 2015b (GI: NTUH)	-2.24	15.48
Yuliwulandari 2016	-2.04	1.05
Gupta 2013b (GI: GUPTA)	-0.08	0.86
Mahmoud 2012	-3.02	2.21
Çetintaş 2008	-12.89	16.47
Santos 2013 (GI: SANTOS)	-27.30	19.05

CI: confidence interval; GI: group identifier

Interpretation

A reasonable amount of studies report data for this SNP, and the bootstrapped CIs overlap, so the assumption of constant λ seems reasonable here.

Conclusion

Perform genetic model-free analysis, and bivariate analysis to investigate the robustness of results.

NAT2 590G-A

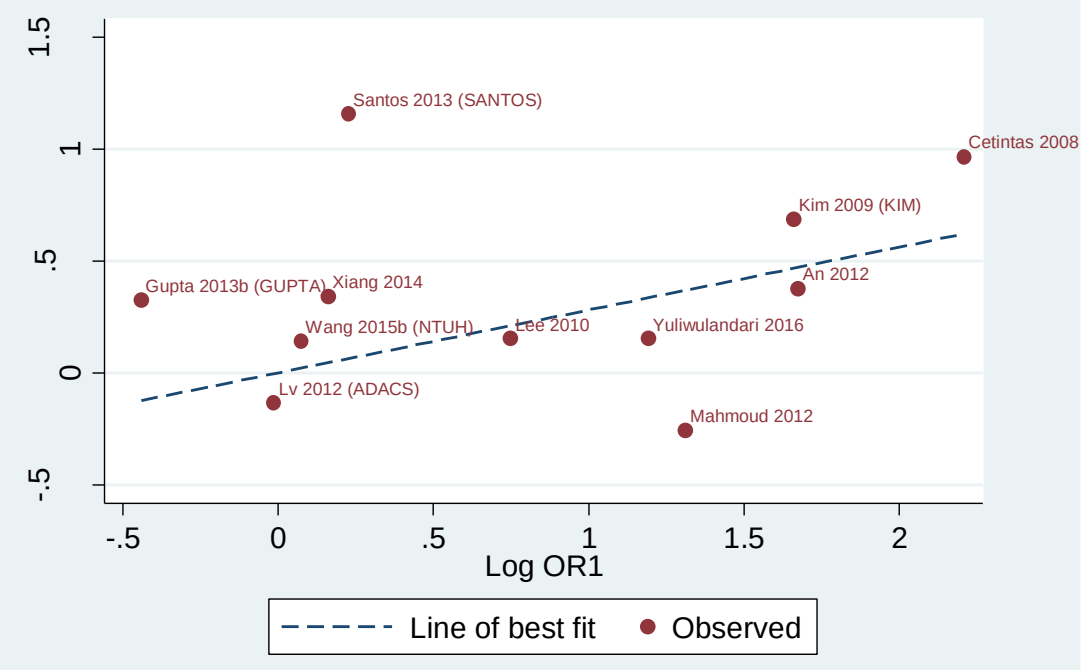

Figure 93 NAT2 *590G-A: Graph showing the log OR for AG* versus *GG against the log OR for AA* versus *GG for each study (all studies)*

Equation for line of best fit: y=0.28x
Log OR1: log odds ratio for AG *versus* GG; log OR2: log odds ratio for AA *versus* GG; OR: odds ratio

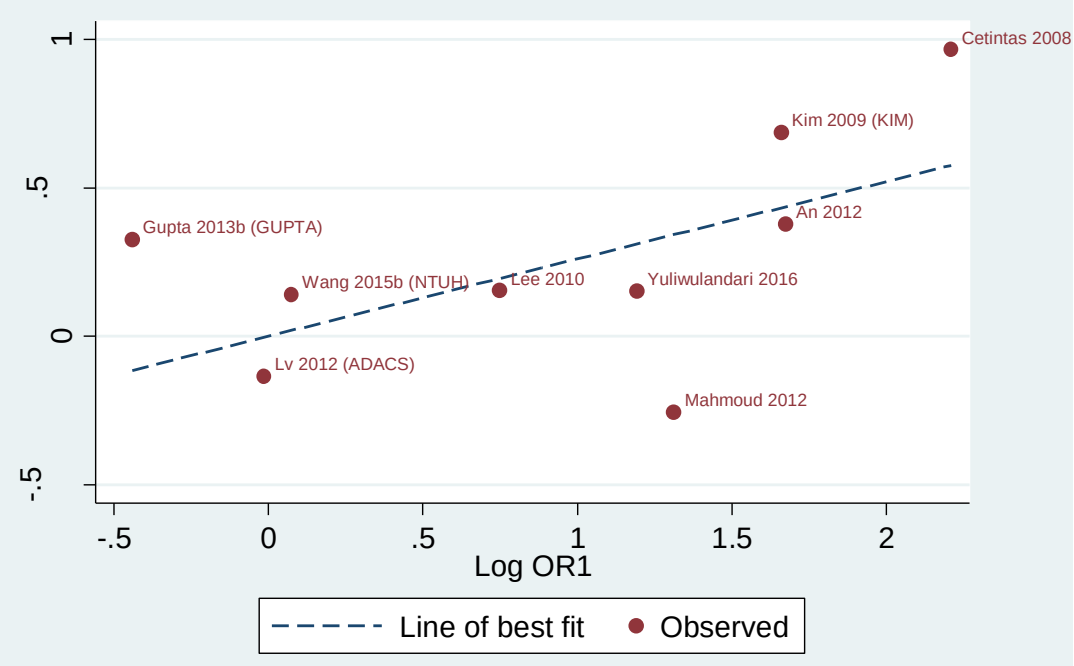

Figure 94 NAT2 *590G-A: Graph showing the log OR for AG* versus *GG against the log OR for AA* versus *GG for each study (sensitivity analysis excluding Xiang 2014 and Santos 2013 (due to deviation from HWE)*

Equation for line of best fit: y=0.26x
HWE: Hardy-Weinberg equilibrium; log OR1: log odds ratio for AG *versus* GG; log OR2: log odds ratio for AA *versus* GG; OR: odds ratio

Table 51 NAT2 *590G-A: λ for each study*

Study	λ
An 2012	0.23
Lv 2012 (GI: ADACS)	9.73
Xiang 2014	2.13
Lee 2010	0.21
Wang 2015b (GI: NTUH)	1.90
Kim 2009 (GI: KIM)	0.41
Yuliwandari 2016	0.13
Gupta 2013b (GI: GUPTA)	-0.74
Mahmoud 2012	-0.20
Çetintaş 2008	0.44
Santos 2013 (GI: SANTOS)	5.11

GI: group identifier

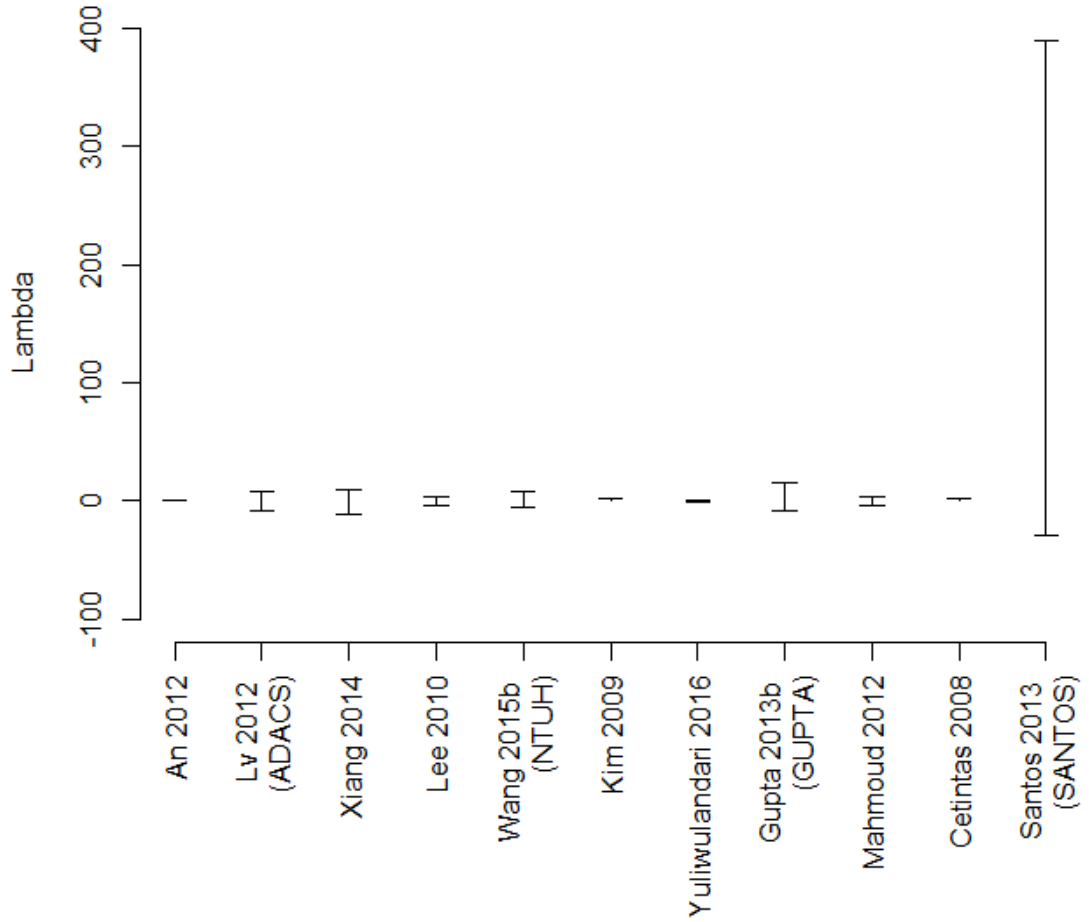

Figure 95 NAT2 *590G-A: Graph showing the bootstrapped 95% CIs for λ for each study*

CI: confidence interval

Table 52 NAT2 *590G-A: Limits of the bootstrapped 95% CI for λ for each study*

Study	Lower limit for λ	Upper limit for λ
An 2012	-0.17	0.85
Lv 2012 (GI: ADACS)	-9.24	7.61
Xiang 2014	-11.97	8.80
Lee 2010	-3.69	3.42
Wang 2015b (GI: NTUH)	-5.74	7.84
Kim 2009 (GI: KIM)	0.03	1.25
Yuliwandari 2016	-0.86	0.76
Gupta 2013b (GI: GUPTA)	-8.68	14.92
Mahmoud 2012	-4.36	3.91
Çetintaş 2008	0.01	1.16
Santos 2013 (GI: SANTOS)	-29.57	388.87

CI: confidence interval; GI: group identifier

Interpretation

A reasonable amount of studies report data for this SNP, and the bootstrapped CIs overlap, so the assumption of constant λ seems reasonable here.

Conclusion

Perform genetic model-free analysis, and bivariate analysis to investigate the robustness of results.

NAT2 803A-G

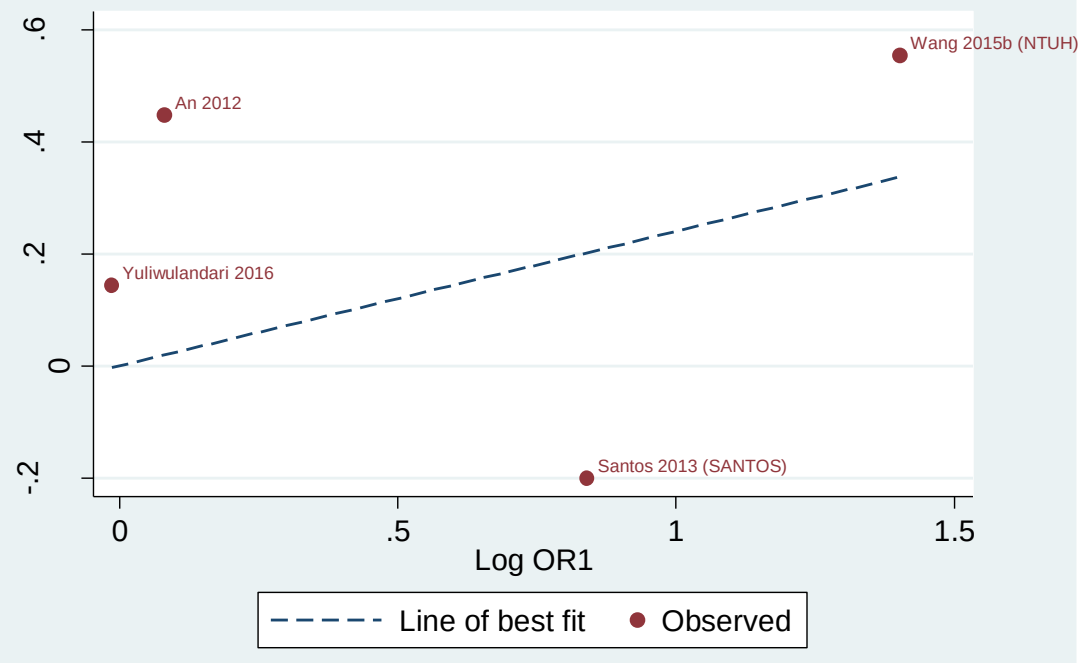

Figure 96 NAT2 *803A-G: Graph showing the log OR for GA* versus *AA against the log OR for GG* versus *AA for each study*

Equation for line of best fit: y=0.24x
Log OR1: log odds ratio for GA *versus* AA; log OR2: log odds ratio for GG *versus* AA; OR: odds ratio

Table 53 NAT2 *803A-G:* λ *for each study*

Study	λ
An 2012	5.57
Wang 2015b (GI: NTUH)	0.40
Yuliwulandari 2016	-10.59
Santos 2013 (GI: SANTOS)	-0.24

GI: group identifier

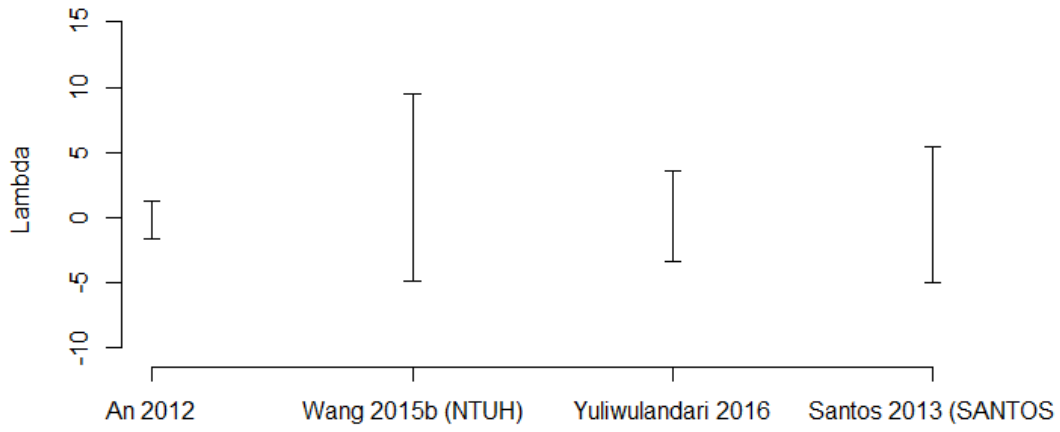

Figure 97 NAT2 *803A-G: Graph showing the bootstrapped 95% CIs for λ for each study*

CI: confidence interval

Table 54 NAT2 *803A-G: Limits of the bootstrapped 95% CI for λ for each study*

Study	Lower limit for λ	Upper limit for λ
An 2012	-1.62	1.22
Wang 2015b (GI: NTUH)	-4.91	9.46
Yuliwulandari 2016	-3.37	3.56
Santos 2013 (GI: SANTOS)	-5.09	5.47

CI: confidence interval; GI: group identifier

Interpretation

Although there is overlap in the bootstrapped CIs for λ, there are few studies and considerable variability in the estimates of λ from each of these studies, so the validity of the assumption of constant λ is unclear.

Conclusion

Perform both bivariate and genetic model-free analyses.

NAT2 857G-A

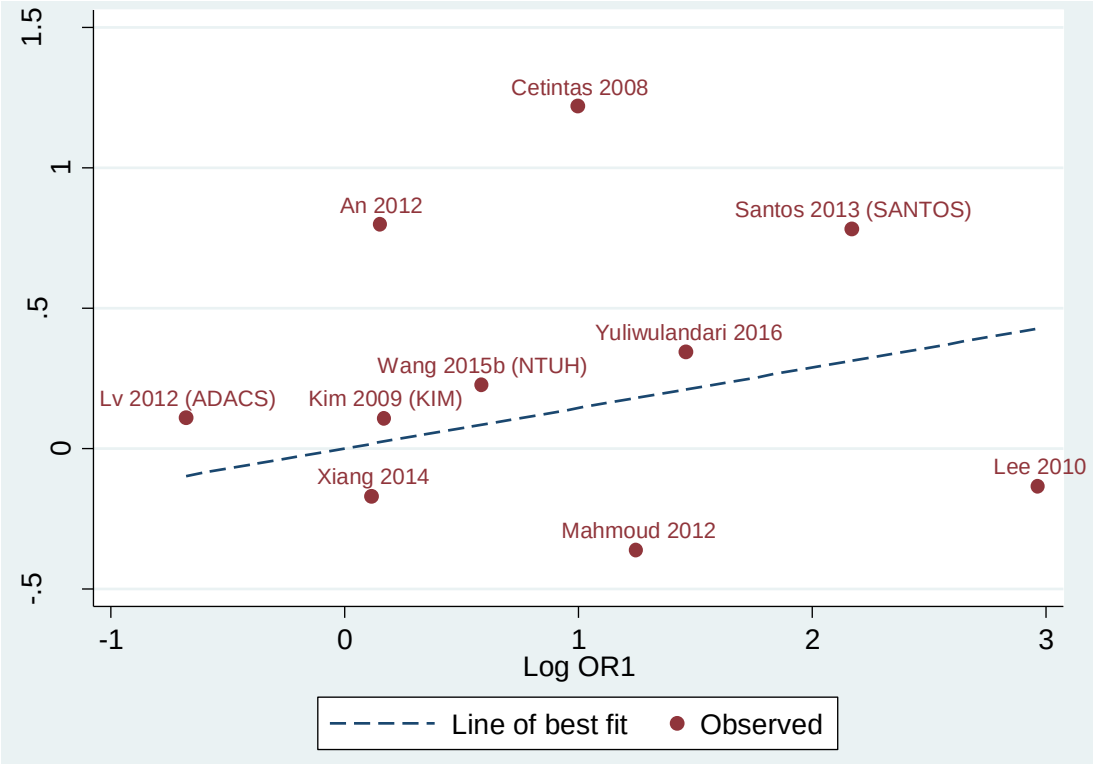

Figure 98 NAT2 *857G-A: Graph showing the log OR for AG* versus *GG against the log OR for AA* versus *GG acetylators for each study*

Equation for line of best fit: y=0.14x
Log OR1: log odds ratio for AG *versus* GG; log OR2: log odds ratio for AA *versus* GG; OR: odds ratio

Table 55 NAT2 *857G-A: λ for each study*

Study	*λ*
An 2012	5.28
Lv 2012 (GI: ADACS)	-0.16
Xiang 2014	-1.48
Lee 2010	-0.05
Wang 2015b (GI: NTUH)	0.39
Kim 2009 (GI: KIM)	0.64
Yuliwulandari 2016	0.24
Mahmoud 2012	-0.29
Çetintaş 2008	1.22
Santos 2013 (GI: SANTOS)	0.36

GI: group identifier

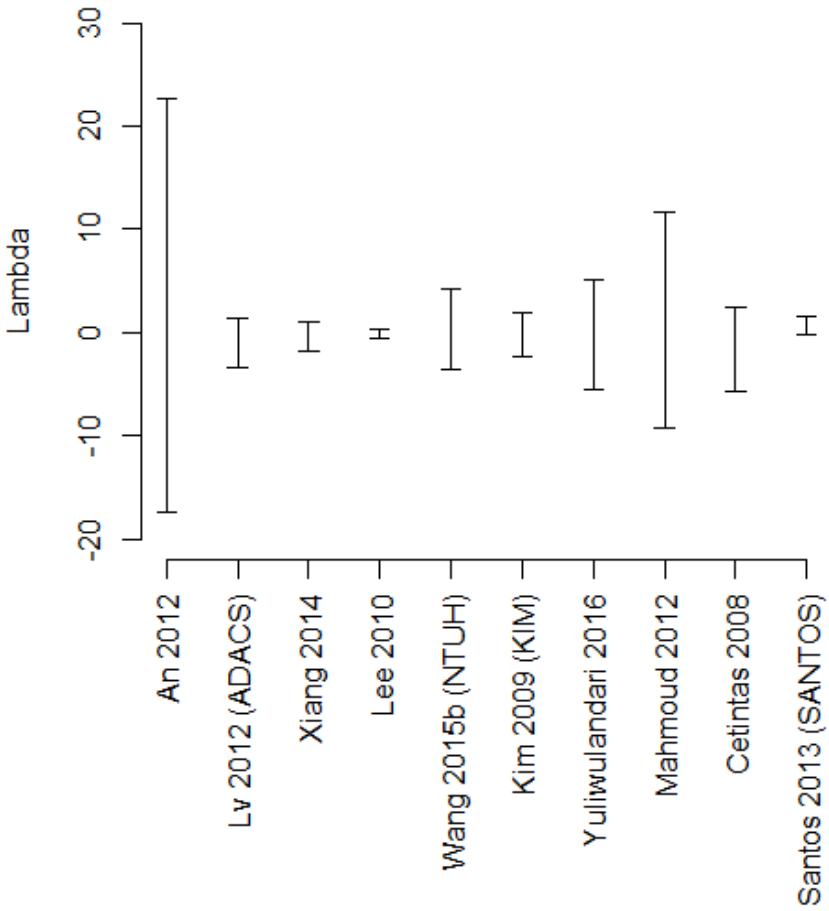

Figure 99 NAT2 *857G-A: Graph showing the bootstrapped 95% CIs for λ for each study*

CI: confidence interval

Table 56 NAT2 *857G-A: Limits of the bootstrapped 95% CI for λ for each study*

Study	Lower limit for λ	Upper limit for λ
An 2012	-17.32	22.66
Lv 2012 (GI: ADACS)	-3.47	1.46
Xiang 2014	-1.83	1.01
Lee 2010	-0.50	0.27
Wang 2015b (GI: NTUH)	-3.50	4.32
Kim 2009 (GI: KIM)	-2.39	1.87
Yuliwulandari 2016	-5.43	5.06
Mahmoud 2012	-9.27	11.69
Çetintaş 2008	-5.68	2.53
Santos 2013 (GI: SANTOS)	-0.28	1.57

CI: confidence interval; GI: group identifier

Interpretation

A reasonable amount of studies report data for this SNP, and the bootstrapped CIs overlap, so the assumption of constant λ seems reasonable here.

Conclusion

Perform genetic model-free analysis, and bivariate analysis to investigate the robustness of results.

Appendix 6. Further analyses of the association between *NAT2* genetic variants and hepatotoxicity: Results of the sensitivity analyses

NAT2 282C-T

Table 57 Association between NAT2 *282C-T and hepatotoxicity: sensitivity analysis excluding Santos 2013*

Analysis approach	λ (95% CI)	OR for CT *vs* CC (95% CI)	I^2	OR for TT *vs* CC (95% CI)	I^2
Genetic model-free	0.13 (-0.19 to 0.44)	1.21 (0.73 to 1.98)	0.0%	4.30 (2.38 to 7.77)	0.0%
Bivariate	N/A	1.21 (0.73 to 1.98)	0.0%	4.30 (2.38 to 7.77)	0.0%
Pairwise comparisons	N/A	1.20 (0.72 to 2.00)	0.0%	4.68 (2.52 to 8.70)	0.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

NAT2 341T-C

Table 58 Association between NAT2 *341T-C and hepatotoxicity: sensitivity analysis excluding Lee 2010*

Analysis approach	λ (95% CI)	OR for TC *vs* TT (95% CI)	I^2	OR for CC *vs* TT (95% CI)	I^2
Genetic model-free	0.14 (-0.44 to 0.73)	1.12 (0.68 to 1.84)	0.0%	2.26 (0.95 to 5.36)	0.0%
Bivariate	N/A	1.11 (0.68 to 1.83)	0.0%	2.25 (0.95 to 5.36)	0.0%
Pairwise comparisons	N/A	1.12 (0.68 to 1.85)	0.0%	1.80 (0.67 to 4.84)	0.0%

CI: confidence interval; N/A: not applicable; OR: odds ratio

NAT2 590G-A

Table 59 Association between NAT2 *590G-A and hepatotoxicity: sensitivity analysis excluding Xiang 2014 and Santos 2013*

Analysis approach	λ (95% CI)	OR for AG *vs* GG (95% CI)	I^2	OR for AA *vs* GG (95% CI)	I^2
Genetic model-free	0.30 (0.06 to 0.55)	1.29 (1.01 to 1.64)	0.0%	2.30 (1.32 to 4.03)	54.7%
Bivariate	N/A	1.30 (1.00 to 1.69)	0.0%	2.29 (1.28 to 4.10)	54.7%
Pairwise comparisons	N/A	1.27 (1.02 to 1.59)	0.0%	2.29 (1.28 to 4.12)	54.7%

CI: confidence interval; N/A: not applicable; OR: odds ratio

Appendix 7. Development of the STROPS guideline: Reporting items scored in the Delphi survey

Table 60 Reporting items scored at Round 1 and Round 2 of the Delphi survey

Reporting item	Help text
1. Indicate the study's pharmacogenetic design in the title and the abstract.	N/A
2. Provide in the abstract an informative and balanced summary of what was done and what was found.	Provide the key information that enables readers to understand the research question, methods, results and conclusions of the study.
3. Explain the scientific background and rationale for the investigation being reported.	Provide the rationale for conducting the pharmacogenetic study in the context of existing research in this health area, i.e. what is known on a topic and what gaps in current knowledge are addressed by the study.
4. Provide reasons for choosing the genes and SNPs genotyped.	Explain how the investigated genes and SNPs were chosen, with reference to relevant functional/animal studies, previous association studies, and any procedures used such as the 'tagging SNP' approach, or by assessing the likelihood of each individual SNP affecting the gene function with priority given to those with the most likely functional effect.
5. If reasons for #4 include previous association studies, provide key details from these studies (effect size and standard error/confidence interval).	N/A
6. State specific objectives, including any pre-specified hypotheses.	Provide the objectives for the study, specifying the relevant population, genetic variants, drugs and outcomes.
7. State if the study is the first report of a pharmacogenetic association, a replication effort, or both.	N/A
8. Present key elements of study design early in the paper.	Present key elements of study design so that readers can understand the basics of the study, e.g. for a cohort study, state that the study used a cohort design, describe the group of people that comprised the cohort and the time period for which they were followed; for a case-control design, state that the study used a case-control design, describe the cases and controls and their source population, etc; for a post-hoc pharmacogenetic analysis of a RCT, state how the subjects included in the analysis were chosen, including which arm of the RCT they were from.
9. Describe the setting, locations and relevant dates, including periods of recruitment, follow-up, and data collection.	Provide sufficient information to enable readers to assess the context and generalisability of a study's results.
10. Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.	Provide sufficiently detailed descriptions of the study participants to help readers understand the applicability of the results. Include details of follow-up procedures, including any procedures to minimise non-response/loss to follow-up.

Reporting item	Help text
11. Case-control study – Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. State whether true controls or population controls were used. Give the rationale for the choice of cases and controls.	Provide sufficiently detailed descriptions of cases and controls to help readers understand the applicability of the results. True controls are controls who have been exposed to the relevant treatment but have not developed the outcome of interest. Population controls are sometimes used in genetic studies for convenience; they are individuals who have already been genotyped that can be assumed to be controls, although we cannot ascertain whether they would have developed the outcome of interest if they had been exposed to the relevant treatment.
12. Cross-sectional study – Give the eligibility criteria, and the sources and methods of selection of participants.	Provide sufficiently detailed descriptions of the study participants to help readers understand the applicability of the results.
13. Report the drug and regime participants were exposed to, and the length of exposure.	Provide details of drug and regime, and length of exposure. This could be fixed across all participants, case/control specific, or variable. If variable, summarise in patient characteristics table.
14. Cohort study – For matched studies, give matching criteria and number in each genotype group.	Provide details of variables that were used to match individuals from each genotype group at the start of follow-up to make groups more comparable, and the numbers in each group.
15. Case-control study – For matched studies, give matching criteria and the number of controls per case.	Provide details of variables that were used to match cases and controls to ensure similarity between these groups, and the number of controls recruited per case.
16. Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	If one or more sub-samples from a larger study are used for the investigation of a genetic association, provide details of: inclusion and exclusion criteria, sources and methods of selection for these sub-samples, and state whether these methods were pre-specified or post-hoc.
17. If other publications report results for the same patient cohort, or a subset of the patient cohort, provide information on this patient cohort overlap and references to the relevant publications.	If other publications report results for the same patient cohort, or a subset of the patient cohort, report the numbers of patients in the current study for whom other publications report data for, rationale for the multiple publications, and provide references to the other publications.
18. Provide justification for choice of outcomes.	Explain why the outcomes are important, e.g. clinical importance, importance to patients, occurrence in previously developed core outcome sets, identification of a significant association in previous studies, etc.
19. Clearly define all outcomes, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Clearly define all outcomes, and all variables considered for and included in the analysis. Provide details of the diagnostic criteria for disease outcomes if applicable.
20. Clearly define genetic exposures (genetic variants) using a widely used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin).	The Human Gene Nomenclature Committee have published guidelines for human gene nomenclature.[1,2] Standard reference numbers are provided in dbSNP,[3] the National Center for Biotechnology Information's database of genetic variation. Guidelines are available for variations not listed in dbSNP.[4,5] Principle components analysis can be undertaken to infer continuous axes of genetic variation that reduce the data to a

Reporting item	Help text
	small number of dimensions.[6] The resulting principle components can be included as variables in analyses of association as a means of adjusting for population structure. 1. Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S. Guidelines for human gene nomenclature. *Genomics* 2002;79(4):464-70. 2. Wain HM, Lush M, Ducluzeau F, Povey S. Genew: the human gene nomenclature database. *Nucleic Acids Res* 2002;30(1):169-71. 3. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308-11. 4. Antonarakis SE, Group NW. Recommendations for a nomenclature system for human gene mutations. *Hum Mutat* 1998;11(1):1-3. 5. Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000;15(1):7-12. 6. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38(8):904-9.
21. Report the rs number of each genotyped SNP.	An 'rs' number (reference SNP ID number) is an identification tag assigned by NCBI to a group (or cluster) of SNPs that map to an identical location. The rs ID number, or rs tag, is assigned after submission of a SNP to dbSNP.
22. Report whether the outcomes measured (including definitions) are in line with core/preferred outcome sets for the particular topic of interest.	A core/preferred outcome set is an agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical studies in specific areas of health or health care. The COMET database[1] lists references to planned, ongoing and completed core outcome set work for a wide range of health topics. 1. COMET Initiative. *Core Outcome Measures in Effectiveness Trials* [cited 2020 10 July]. Available from: http://www.comet-initiative.org/.
23. For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.	Provide information on how all genetic exposures, confounders, and outcomes were measured. Report whether there were any differences in how data were collected in different patient groups (e.g. cases and controls).
24. Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify	Provide sufficient details to enable the reader to assess the potential extent of genotyping errors (a source of information bias). Report whether there were any differences in laboratory methods in different patient groups (e.g. cases and controls).

Reporting item	Help text
whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.	
25. If study is case-control, confirm whether patients were genotyped in mixed batches.	Report whether cases and controls were put into combined batches for genotyping purposes (to ensure genotyping quality is comparable across groups), rather than analysed in separate batches.
26. Confirm whether genotyping personnel were blinded to outcome status.	N/A
27. Describe the primers used.	Report details of any primers that were used, or if primers were inherent to the assay used, state this and provide information on the assay.
28. Describe genotype quality control methods.	Provide details of genotype quality control methods, for example, using negative controls, or re-genotyping/re-sequencing in all or a random sample of patients.
29. Describe findings of genotype quality control methods	Provide sufficient information to enable readers to assess the risk of bias due to incorrect genotype allocation.
30. Describe any efforts to address potential sources of bias.	Bias is a systematic deviation of a study's result from a true value, due to flawed information or subject selection. Report all potential sources of bias, and any steps taken to reduce the likelihood of bias occurring.
31. For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.	Bias from pharmacotherapy may occur when quantitative outcome variables are affected by treatment with drugs other than the study drug (e.g. outcome variables include biochemical markers of hepatotoxicity, and several patients are taking concomitant hepatotoxic medications).
32. Report how adherence to treatment was assessed, and report the results of the assessment.	Provide details on assessments of patient adherence, including limitations of the chosen method.
33. Explain how the study size was arrived at, or provide details of the *a priori* power to detect effect sizes of varying degrees.	Report the calculation performed to obtain the study sample size, providing references to any specific methodology. Or, if sample size was predetermined (for example, secondary analyses of a published dataset), provide details of *a priori* power calculations for a range of plausible effect sizes.
34. Explain how quantitative variables (confounders and effect modifiers) were handled in the analyses. If applicable, describe which groupings were chosen, and why.	Explain how quantitative data (in relation to effect modifiers and confounders) were collected and analysed. Continuous variables may be grouped into categories to create a new categorical variable; explain why and how quantitative data were grouped.
35. If applicable, describe how effects of treatment on quantitative outcome variables were dealt with.	If any quantitative outcome variables may be affected by treatment with drugs other than the study drug, report whether any of the available methods for adjusting for treatment effects[1] were used to deal with this potential bias. 1. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits:

Reporting item	Help text
	antihypertensive therapy and systolic blood pressure. *Stat Med* 2005;24(19):2911-35.
36. Describe all statistical methods, including those used to control for confounding.	Report which analyses were pre-specified and which were exploratory based on data inspection. If groups being compared are not similar regarding some characteristics, adjustment should be made for possible confounding variables. Provide details of procedures of variable selection and model comparison.
37. State software version used and options (or settings) chosen.	Provide details of any specialized software/packages used to analyse the data.
38. Describe any methods used to examine subgroups and interactions.	Explain what methods were used to examine whether associations differed across subgroups, or to examine interactions. An 'interaction' occurs when one factor modifies the effect of another, and is also sometimes referred to as 'effect modification'. Report whether these analyses were pre-planned or not.
39. Explain how missing data were addressed.	Confirm whether analyses were restricted to individuals with complete data on the required variables, or whether any imputation of missing data was performed.
40. Report any methods used to assess the assumption of missingness at random and the finding of such assessments.	Data are said to be 'missing at random' if the fact that they are missing is unrelated to actual values of the missing data. Data that are missing at random may not be important. Analyses based on the available data will tend to be unbiased, although based on a smaller sample size than the original data set.
41. Cohort study – If applicable, explain how loss to follow-up was addressed.	Individuals who withdrew from the study before the end of their observation period are 'lost to follow-up'. Report how many individuals were lost to follow-up, and whether these individuals were excluded or whether censoring strategies were used. Describe any censoring strategies.
42. Case-control study – If applicable, explain how matching of cases and controls was addressed.	If a matched design was used, describe in detail what statistical methods were used to take into account the matching of cases and controls.
43. Cross-sectional study – If applicable, describe analytical methods taking account of sampling strategy.	Sampling may be more complex than taking a simple random sample from the source population. For example, it may include several stages and clustering of participants. If a complex sampling strategy is used, estimates of association may be more or less precise than those derived from a simple random sample. If a complex sampling strategy was used, clearly state the methods used to adjust for this, so that readers may understand how the chosen sampling method influenced the precision of the obtained estimates.
44. Describe any sensitivity analyses.	Provide details of any analyses that were performed to investigate whether the results of the main analysis are consistent with those obtained with alternative analysis strategies or assumptions. Report whether these analyses were pre-planned or not.
45. State whether HWE was considered and, if so, how.	Describe any statistical tests or measures of departure from HWE, and any methods used to allow for deviations from HWE.
46. Where HWE test is undertaken, quote the *p*-value	State how small a *p*-value from a HWE test had to be to indicate a statistically significant deviation from HWE.

Reporting item	Help text
threshold applied to determine deviation from HWE.	
47. Describe any methods used for inferring genotypes or haplotypes.	Provide statistical methods and software used to infer genotype phase and haplotypes. See the STREGA statement[1] for further information. 1. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, Von Elm E, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *Hum Genet* 2009;125(2):131-51.
48. Describe any methods used to assess or address population stratification.	Explicitly document any methods used to assess the presence of population stratification or adjust for population stratification in the analyses. If no methods were used, state this.
49. Describe any methods used to assess and correct for relatedness among subjects. Report results of assessments for relatedness.	For pharmacogenetic studies, it is not uncommon for some participants to be related. Authors should report any methods used to assess relatedness, results of these assessments, and any methods used to correct for relatedness.
50. Describe any assumptions made regarding mode of inheritance.	The mode of inheritance for a particular SNP determines the observed trait for each given combination of parental alleles - some different modes of inheritance are summarised by Jorgensen and Williamson.[1] If a particular mode of inheritance is assumed in the analyses, i.e. genotypes are grouped according to the resulting observed trait, state this. If more than one approach is used in the analysis, assuming different modes of inheritance, then report on all analyses performed. 1. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. *Stat Med* 2008;27(30):6547-69.
51. Provide justification for assumption of mode of inheritance or if no mode is assumed.	For each analysis where a particular mode of inheritance is assumed, or no mode of inheritance is assumed, justify this approach, for example by summarising previous research on the mode of inheritance for the relevant SNP. If multiple analyses are performed making different assumptions, justify this decision.
52a. Describe any methods used to address multiple comparisons or to control risk of false positive findings due to investigating multiple genetic variants.	Provide sufficient detail to enable the reader to assess the likelihood of false positive results (type 1 errors) being reported.
52b. Describe any methods used to address multiple comparisons or to control risk of false positive findings due to investigating multiple outcomes.	Provide sufficient detail to enable the reader to assess the likelihood of false positive results (type 1 errors) being reported.
52c. Describe any methods used to address multiple comparisons or to control risk of false positive findings due to investigating multiple assumptions regarding mode of inheritance.	Provide sufficient detail to enable the reader to assess the likelihood of false positive results (type 1 errors) being reported.
53. Describe any methods used to adjust for extent of adherence in the analyses.	N/A

Reporting item	Help text
54a. Report the numbers of individuals at each stage of the study – e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.	Give an account of the numbers of individuals considered at each stage of recruiting study participants, from the choice of a target population to the inclusion of participants' data in the analysis.
54b. Give reasons for non-participation at each stage.	Provide transparent information on reasons for non-participation/exclusion of participants at each stage, to allow the reader to judge whether the study population was representative of the target population, and whether bias was possibly introduced.
54c. Consider use of a flow diagram.	A flow diagram can be an efficient and transparent way to convey the information described in #54a and #54b, which may otherwise require a lengthy description in the text.
55. For each genetic variant, report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful.	N/A
56. Report any SNPs that were excluded from analysis, and provide reasons for these exclusions.	Examples of possible reasons for excluding SNPs are the failure of the HWE test, or excessive missing data, etc.
57. Give characteristics of study participants (e.g., demographic, clinical, social) and information on potential confounders.	Report participant characteristics with appropriate summary measures. For example, for continuous data, mean and standard deviation, or median and range; for dichotomous data, numbers and proportions.
58. Indicate the number of participants with missing data for each variable of interest.	Report the amount of missing genotype data, missing data for potential confounders, and missing data for other important patient characteristics and outcomes.
59. For a cohort study, consider giving information listed in #57 and #58 by genotype.	Where several genetic variants have been considered, this may not be practical.
60. For a case-control study, give the information listed in #57 and #58 for cases and controls separately.	N/A
61. Report reasons for missing genotype data.	Report any reasons that genotype data were missing and report how much of the missing data was attributed to each reason.
62. Cohort study – Summarize follow-up time, e.g. average and total amount.	Average follow-up can be summarised using the mean and/or median follow-up time. Total amount of follow-up may be reported using total person-years of follow-up, or some indication of the completeness of follow-up.[1] 1. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet* 2002;359(9314):1309-10.
63. Where HWE test undertaken, highlight SNPs that deviate from HWE.	Report any SNPs for which a HWE test indicated deviation for HWE.
64. Where population stratification is assessed, report the results.	Report the results of any tests performed to detect the presence of population stratification.

Reporting item	Help text
65a. For a cohort study, report all outcomes (phenotypes) investigated for each genotype category over time.	For outcomes that relate to the occurrence of some event, report the number of events that occurred. If the risk of an event occurring changes over follow-up time, present the numbers and rates of events in appropriate intervals of follow-up or as a Kaplan-Meier life table or plot. For other outcomes, present appropriate summary measures (e.g. means and standard deviations) over time.
65b. For a case-control study, report numbers in each genotype category for all outcomes investigated.	Report numbers of cases and controls in each genotype category.
65c. For a cross sectional study, report all outcomes (phenotypes) investigated for each genotype category.	For outcomes that relate to the occurrence of some event, report the number of events that occurred. For other outcomes, present appropriate summary measures (e.g. means and standard deviations).
66. If a study includes more than one ethnic group, provide the summary data specified in #65 per ethnic group.	If the study includes patients belonging to different ethnic groups, provide all outcome data stratified by ethnicity.
67. Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.	Provide both unadjusted measures of association and measures of association adjusted for potential confounders, to enable readers to compare both measures and assess how the measure of association is impacted by adjusting for confounders. List all potential confounder variables considered, and the criteria/rationale for excluding or including variables in statistical models.
68. Report category boundaries when continuous variables were categorized.	If continuous outcomes were categorised, report the range of values covered by each category.
69. If relevant, consider translating effect estimates to number needed to test to illustrate potential clinical utility of any significant findings.	In pharmacogenetic studies, a 'number needed to test' can be calculated to demonstrate the clinical relevance of study findings. For example, '12 tuberculosis patients (95% confidence interval: 7 to 23) would need to be tested for the genetic variant (and possibly put on alternative therapy dependent on the results of the test) to prevent one case of hepatotoxicity'.
70. Report results of any adjustments for multiple comparisons.	For example, report Bonferroni adjusted *p*-values, or false discovery rates.
71. Report precise *p*-values for all associations.	Report the precise *p*-values for all associations, as opposed to only indicating whether an association was found to be statistically significant or not. For example, stating $p<0.05$ or $p>0.05$, or indicating statistical significance (or a lack of) by using asterisks (*) is not sufficient to satisfy this criterion. Reporting of *p*-values may be subject to journal guidelines; *p*-values below a certain threshold e.g. $p<0.0001$ may need to be reported as such.
72. Report other analyses done – e.g., analyses of subgroups and interactions, and sensitivity analyses.	Report the results of any analyses performed in addition to the main analysis. It may be impractical to present detailed findings for all analyses performed; in this case, present detailed results for important results only. Less important results can be summarised briefly in the text i.e. 'results of the sensitivity analysis were consistent with the results of the main analysis' and detailed in full in supplementary materials.

Reporting item	Help text
73. If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken.	Report results for all genetic variants that were investigated in the study, rather than selectively reporting only 'interesting' or significant results. Full results can be provided in supplementary materials if necessary.
74. If detailed results are available elsewhere, state how they can be accessed.	Report what results are available, and where to find these results. Provide sufficient details that a reader would easily be able to locate these resources.
75. Summarize key results with reference to study objectives.	Remind the reader of the main findings of the study with a short summary. This helps the reader to assess whether the author's interpretation and suggested implications are supported by the findings.
76. Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	Help the reader to interpret the validity and health care relevance of the study findings. Limitations might relate to, for example, characteristics of included patients, methods of outcome measurement, multiplicity of analyses, missing data, etc.
77. Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	When interpreting results, authors should consider the position of the study on the discovery to verification continuum. Consider potential sources of bias, residual confounding (due to unmeasured variables or imprecise measurement of confounders), the results of relevant sensitivity analyses, the issue of multiplicity and subgroup analyses. Authors should address the real range of uncertainty in estimates, which is larger than the statistical uncertainty reflected in confidence intervals.
78. Report genotype frequencies from other studies.	If allelic frequencies have previously been reported for individuals from the same population, quote these for comparison purposes.
79. Discuss the generalisability (external validity) of the study results.	Consider the extent to which the results of the study can be applied to other circumstances, i.e. different populations/settings/countries.
80. Discuss, if pertinent, the health care relevance of the study results.	Consider 1) the efforts it takes to obtain the additional genotype information, 2) the impact the results might have on medical or public health decision making and on expected health benefits, and 3) the extent to which these benefits will outweigh the potential harms related to genetic testing and the cost of implementation. Describe what evidence is still needed before health care implementation can be considered.
81. State whether the protocol for the analysed data is publicly available and if so, how the protocol can be accessed.	If the protocol is available, provide sufficient details that a reader would easily be able to locate this resource.
82. State whether the study has been registered. If the study has been registered, provide details of the registry.	Studies can be registered on many different official platforms; the most widely used platform is ClinicalTrials.gov.
83a. Report whether ethical approval was obtained for the collection of genetic data.	N/A
83b. If ethical approval was obtained, report the committee that gave ethical approval and a reference ID.	N/A

Reporting item	Help text
84. Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	Role of funders: State which part of the study the funders took direct responsibility for, e.g., study design, data collection, analysis, drafting of manuscript, decision to publish.
85. State whether databases for the analysed data are or will become publicly available and if so, how they can be accessed.	If databases are available, provide sufficient details that a reader would easily be able to locate these resources.

COMET: Core Outcome Measures in Effectiveness Trials; HWE: Hardy-Weinberg equilibrium; N/A: not applicable; RCT: randomised controlled trial; SNP: single nucleotide polymorphism

Table 61 Additional reporting items suggested by Delphi participants, scored in Round 2 of the Delphi survey only

Reporting item	Help text
Clearly state how haplotypes or star alleles were defined.	Even when referring to well-defined star alleles or haplotypes, it is good practice to provide full details in the current publication using standard nomenclature. For example, 'NAT2*5A was defined as rs1801280 c.341 allele C and rs1799929 c.481 allele T'.
Clearly state on which chromosomal strand the alleles are reported.	If the chromosomal strand for which alleles are reported is not stated, it is difficult to know which allele is associated with the phenotype of interest for A/T or G/C SNPs. A statement such as 'the A allele (positive chromosomal strand) is associated with...' is clear.
If studying drug metabolites, provide references and links to structures and database identifiers.	e.g. PubChem Compound IDs (https://pubchem.ncbi.nlm.nih.gov/)
Report disease/clinical indication of patients using a standardised ontology.	e.g. SNOMED CT; Mesh
If referring to the minor, wild-type or mutant allele of a variant, state which allele this is and for which given population/cohort.	The minor (less frequent) allele in one population may be the major (more frequent) allele in a different population. The allele and population should be clearly stated if using the terms 'minor', 'wild-type' or 'mutant', e.g. 'the minor allele, T, in Gujarati Indians'.
Report on the risk of phenoconversion (genotype-phenotype mismatch) and its magnitude in the study population.	For more information, see: Shah RR, Smith RL. Addressing phenoconversion: the Achilles' heel of personalized medicine. *Br J Clin Pharmacol* 2015;79(2):222-40.
Confirm whether patients were blinded to their genotyping result.	N/A

N/A: not applicable; SNP: single nucleotide polymorphism

Appendix 8. Development of the STROPS guideline: Consensus matrix

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
Title and abstract								
Title and abstract	1	Indicate the study's pharmacogenetic design in the title and the abstract.	60%	60%	67%	50%	70%	55%
	2	Provide in the abstract an informative and balanced summary of what was done and what was found.	100%	95%	87%	100%	100%	91%
Introduction								
Background/ rationale	3	Explain the scientific background and rationale for the investigation being reported.	80%	88%	87%	90%	97%	91%
	4	Provide reasons for choosing the genes and SNPs genotyped.	87%	85%	87%	90%	97%	91%
	5	If reasons for (4) include previous association studies, provide key details from these studies (effect size and standard error/confidence interval).	80%	40%	43%	70%	45%	27%
Objectives	6	State specific objectives, including any pre-specified hypotheses.	73%	90%	87%	70%	94%	91%
	7	State if the study is the first report of a pharmacogenetic association, a replication effort, or both.	67%	66%	87%	60%	74%	91%
Methods								
Study design	8	Present key elements of study design early in the paper.	80%	76%	87%	70%	74%	91%
Setting	9	Describe the setting, locations and relevant dates, including periods of recruitment, follow-up, and data collection.	60%	73%	53%	40%	61%	36%
Participants	10	Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.	93%	90%	87%	90%	94%	91%
	11	Case-control study – Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. State whether true controls or population controls were used. Give the rationale for the choice of cases and controls.	87%	85%	93%	90%	94%	91%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
	12	Cross-sectional study – Give the eligibility criteria, and the sources and methods of selection of participants.	87%	88%	93%	90%	94%	91%
	13	Report the drug and regime participants were exposed to, and the length of exposure.	93%	83%	87%	90%	90%	100%
	14	Cohort study – For matched studies, give matching criteria and number in each genotype group.	87%	90%	86%	100%	94%	91%
	15	Case-control study – For matched studies, give matching criteria and the number of controls per case.	87%	88%	86%	100%	94%	91%
	16	Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	67%	83%	87%	70%	90%	91%
	17	If other publications report results for the same patient cohort, or a subset of the patient cohort, provide information on this patient cohort overlap and references to the relevant publications.	47%	66%	80%	50%	74%	91%
	18	Report disease/clinical indication of patients using a standardised ontology.	NS	NS	NS	44%	65%	73%
	19	Confirm whether patients were blinded to their genotyping result.	NS	NS	NS	80%	52%	64%
Variables	20	Provide justification for choice of outcomes.	80%	71%	80%	80%	84%	82%
	21	Clearly define all outcomes, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	87%	88%	93%	90%	87%	100%
	22	Clearly define genetic exposures (genetic variants) using a widely used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin).	87%	85%	93%	90%	97%	91%
	23	Report the rs number of each genotyped SNP.	87%	81%	87%	90%	94%	82%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
	24	Report whether the outcomes measured (including definitions) are in line with core/preferred outcome sets for the particular topic of interest.	73%	59%	67%	80%	66%	64%
	25	Clearly state how haplotypes or star alleles were defined.	NS	NS	NS	80%	73%	82%
	26	Clearly state on which chromosomal strand the alleles are reported.	NS	NS	NS	60%	59%	73%
	27	If referring to the minor, wild-type or mutant allele of a variant, state which allele this is and for which given population/cohort.	NS	NS	NS	80%	67%	82%
	28	If studying drug metabolites, provide references and links to structures and database identifiers.	NS	NS	NS	40%	48%	64%
Data sources/ measurement	29	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.	67%	49%	53%	70%	61%	55%
	30	Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.	73%	61%	53%	80%	81%	73%
	31	If study is case-control, confirm whether patients were genotyped in mixed batches.	40%	48%	50%	30%	58%	36%
	32	Confirm whether genotyping personnel were blinded to outcome status.	60%	49%	53%	40%	55%	46%
	33	Describe the primers used.	47%	33%	27%	50%	29%	9%
	34	Describe genotype quality control methods.	80%	43%	47%	90%	52%	46%
	35	Describe findings of genotype quality control methods.	60%	43%	47%	60%	42%	46%
Bias	36	Describe any efforts to address potential sources of bias.	60%	76%	53%	60%	84%	55%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
	37	For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.	80%	60%	54%	90%	73%	73%
	38	Report how adherence to treatment was assessed, and report the results of the assessment.	80%	59%	53%	90%	81%	73%
Study size	39	Explain how the study size was arrived at, or provide details of the *a priori* power to detect effect sizes of varying degrees.	87%	81%	64%	90%	84%	73%
Quantitative variables	40	Explain how quantitative variables (confounders and effect modifiers) were handled in the analyses. If applicable, describe which groupings were chosen, and why.	71%	78%	69%	80%	87%	73%
	41	If applicable, describe how effects of treatment on quantitative outcome variables were dealt with.	79%	71%	57%	90%	77%	46%
Statistical methods	42	Describe all statistical methods, including those used to control for confounding.	64%	90%	86%	80%	97%	82%
	43	State software version used and options (or settings) chosen.	57%	44%	33%	60%	42%	18%
	44	Describe any methods used to examine subgroups and interactions.	57%	73%	64%	60%	87%	64%
	45	Explain how missing data were addressed.	71%	68%	67%	80%	81%	73%
	46	Report any methods used to assess the assumption of missingness at random and the finding of such assessments.	69%	54%	43%	70%	66%	46%
	47	Cohort study – If applicable, explain how loss to follow-up was addressed.	64%	63%	43%	60%	77%	36%
	48	Case-control study – If applicable, explain how matching of cases and controls was addressed.	71%	68%	43%	70%	83%	27%
	49	Cross-sectional study – If applicable, describe analytical methods taking account of sampling strategy.	64%	63%	43%	80%	69%	27%
	50	Describe any sensitivity analyses.	31%	69%	53%	22%	67%	36%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
	51	State whether HWE was considered and, if so, how.	71%	71%	64%	90%	80%	64%
	52	Where HWE test is undertaken, quote the *p*-value threshold applied to determine deviation from HWE.	79%	64%	50%	80%	77%	36%
	53	Describe any methods used for inferring genotypes or haplotypes.	64%	78%	67%	70%	90%	55%
	54	Describe any methods used to assess or address population stratification.	57%	85%	67%	60%	97%	64%
	55	Describe any methods used to assess and correct for relatedness among subjects. Report results of assessments for relatedness.	43%	72%	40%	60%	81%	27%
	56	Describe any assumptions made regarding mode of inheritance.	50%	70%	47%	60%	73%	36%
	57	Provide justification for assumption of mode of inheritance or if no mode is assumed.	50%	65%	47%	60%	77%	55%
	58a	Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating multiple genetic variants.	86%	80%	80%	90%	97%	82%
	58b	Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating multiple outcomes.	86%	78%	80%	90%	93%	82%
	58c	Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating multiple assumptions regarding mode of inheritance.	57%	78%	64%	50%	86%	73%
	59	Describe any methods used to adjust for extent of adherence in the analyses.	67%	63%	50%	89%	74%	50%
Results								
Participants	60a	Report the numbers of individuals at each stage of the study – e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.	67%	85%	87%	90%	90%	91%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
	60b	Give reasons for non-participation at each stage.	40%	63%	53%	30%	58%	46%
	60c	Consider use of a flow diagram.	40%	51%	47%	30%	48%	36%
	61	For each genetic variant, report numbers of individuals in whom genotyping was attempted and numbers of individuals in whom genotyping was successful.	53%	60%	73%	50%	63%	73%
SNPs	62	Report any SNPs that were excluded from analysis, and provide reasons for these exclusions.	73%	68%	80%	50%	81%	82%
Descriptive data	63	Give characteristics of study participants (e.g., demographic, clinical, social) and information on potential confounders.	80%	88%	100%	100%	94%	100%
	64	Indicate the number of participants with missing data for each variable of interest.	73%	61%	73%	90%	71%	91%
	65	For a cohort study, consider giving information listed in (63) and (64) by genotype.	47%	50%	71%	30%	50%	73%
	66	For a case-control study, give the information listed in (63) and (64) for cases and controls separately.	67%	60%	71%	60%	67%	64%
	67	Report reasons for missing genotype data.	53%	46%	53%	40%	42%	36%
	68	Cohort study – Summarize follow-up time, e.g. average and total amount.	60%	78%	60%	70%	87%	64%
	69	Where HWE tests have been undertaken, highlight SNPs that deviate from HWE.	80%	66%	77%	90%	79%	73%
	70	Where population stratification is assessed, report the results.	93%	73%	80%	90%	87%	73%
Outcome data	71a	For a cohort study, report all outcomes (phenotypes) investigated for each genotype category over time.	57%	83%	100%	70%	80%	91%
	71b	For a case-control study, report numbers in each genotype category for all outcomes investigated.	79%	85%	100%	90%	87%	91%
	71c	For a cross sectional study, report all outcomes (phenotypes) investigated for each genotype category.	57%	85%	100%	70%	90%	91%
	72	If a study includes more than one ethnic group, provide the summary data specified in (71) per ethnic group.	86%	78%	73%	90%	84%	73%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
Main results	73	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g. 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.	86%	88%	92%	100%	97%	100%
	74	Report category boundaries when continuous variables were categorised.	69%	83%	100%	70%	97%	100%
	75	If relevant, consider translating effect estimates to number needed to test to illustrate potential clinical utility of any significant findings.	62%	57%	53%	60%	62%	46%
	76	Report results of any adjustments for multiple comparisons.	86%	81%	87%	80%	94%	100%
	77	Report precise *p*-values for all associations.	64%	73%	87%	70%	87%	91%
Other analyses	78	Report other analyses done – e.g., analyses of subgroups and interactions, and sensitivity analyses.	50%	75%	71%	50%	83%	91%
	79	If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken.	36%	76%	73%	20%	74%	73%
	80	If detailed results are available elsewhere, state how they can be accessed.	86%	73%	67%	80%	81%	82%
Discussion								
Key results	81	Summarize key results with reference to study objectives.	100%	85%	100%	100%	97%	100%
Limitations	82	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	93%	85%	100%	100%	97%	100%
	83	Report on the risk of phenoconversion (genotype-phenotype mismatch) and its magnitude in the study population.	NS	NS	NS	78%	50%	46%
Interpretation	84	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	86%	83%	93%	100%	90%	100%
	85	Report genotype frequencies from other studies.	57%	37%	47%	50%	23%	36%

Category	#	Criteria	Consensus Round 1			Consensus Round 2		
			JE (n=15)	PR (n=41)	SR (n=15)	JE (n=10)	PR (n=31)	SR (n=11)
Generalisability	86	Discuss the generalisability (external validity) of the study results.	71%	73%	73%	90%	81%	73%
	87	Discuss, if pertinent, the health care relevance of the study results.	71%	70%	60%	90%	67%	55%
Other information								
Study registration/protocol	88	State whether the protocol for the analysed data is publicly available and if so, how the protocol can be accessed.	64%	53%	53%	30%	58%	46%
	89	State whether the study has been registered. If the study has been registered, provide details of the registry.	79%	63%	53%	80%	71%	36%
Ethical approval	90a	Report whether ethical approval was obtained for the collection of genetic data.	86%	95%	80%	100%	97%	73%
	90b	If ethical approval was obtained, report the committee that gave ethical approval and a reference ID.	43%	54%	53%	50%	55%	27%
Funding	91	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	64%	63%	87%	80%	81%	82%
Databases	92	State whether databases for the analysed data are or will become publicly available and if so, how they can be accessed.	71%	54%	60%	70%	65%	46%

Consensus definition: Support from at least 70% of participants scoring 'Critical', i.e. score 7-9 (from a 1-9 scale). Green shading indicates the stakeholder group reached consensus for the item.

Participants were excluded from the calculations (denominators) if they did not score an item.

HWE: Hardy-Weinberg equilibrium; JE: journal editors; NS: not scored; PR: primary researchers; SNP: single nucleotide polymorphism; SR: systematic reviewers

Appendix 9: Development of the STROPS guideline: Explanation and elaboration document

Abstract

1. Provide in the abstract an informative and balanced summary of what was done and what was found.

Explanation

Study authors should provide the key information that enables readers to understand the research question, study design, methods, results and conclusions of the study. This item is from the STROBE statement[1] (Item 1b); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2] An example abstract is provided below; however it is important to note that journals may specify their own set of guidelines, which authors ought to follow in terms of abstract structure and content.

Example

"INTRODUCTION: Approximately 30% of patients with epilepsy are resistant to treatment with anti-epileptic drugs (AEDs). The ABC drug transporter proteins are hypothesized to mediate drug resistance in epilepsy. More recently, a non-ABC putative transporter, RLIP76, has also been proposed to be involved in the mechanism of pharmacoresistance. One previous association study of six polymorphisms in *RLIP76* failed to find any association with drug resistance in a retrospective cohort of epilepsy patients. We aimed to look for an association with outcomes reflecting drug response in a larger prospective cohort, with gene-wide coverage.

PATIENTS AND METHODS: We investigated the role of common polymorphisms in *RLIP76* in epilepsy pharmacoresistance by genotyping 23 common *RLIP76* polymorphisms in a prospective cohort of 503 epilepsy patients, from the standard and new anti-epileptic drugs (SANAD) prospective study of new and old AEDs. A total of 13 of these were tested for association with four outcomes reflecting response to drugs: time to first seizure, time to 12-month remission, time to withdrawal due to inadequate seizure control, and time to withdrawal due to unacceptable adverse drug events.

RESULTS: No significant associations, allowing for multiple testing, were found in the whole cohort. There was also no effect in a subgroup of patients on carbamazepine, which is thought to be a RLIP76 substrate, although two polymorphisms were associated with time to first seizure (p=0.007).

DISCUSSION: We failed to demonstrate any association between *RLIP76* polymorphisms and four different measures of drug response in the larger cohort, but a subgroup analysis of patients receiving carbamazepine suggested an association that should be investigated further.

CONCLUSIONS: Our data suggest that common variants in *RLIP76* are unlikely to contribute to epilepsy drug response".[3]

Introduction: Background/rationale

2. Explain the scientific background and rationale for the investigation being reported.

Explanation

Study authors should provide the rationale for conducting the pharmacogenetic study in the context of existing research in this health area, i.e. what is known on a topic and what

gaps in current knowledge are addressed by the study. This item is from the STROBE statement[1] (Item 2); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"Inhaled corticosteroids (ICS) are recommended for adults and children with asthma and for chronic obstructive pulmonary disease (COPD). Although ICS are generally well tolerated and have fewer systemic adverse effects than do oral corticosteroids, some patients can still develop systemic adverse effects. Adrenal suppression is a clinically important adverse effect, particularly in children with asthma, in whom the diagnosis of adrenal suppression can be challenging because presentation can range from asymptomatic biochemical changes to nonspecific lethargy to florid adrenal crisis and death (…).

Interindividual variation in susceptibility to adrenal suppression is striking (…). The reasons for the interindividual variability in both adults and children remain unclear, because clinical factors only account for a small proportion of the variance.

Previous pharmacogenomic studies in patients with asthma using corticosteroids have focused on efficacy. As far as we know, no studies examining the pharmacogenomics of corticosteroid-induced adrenal suppression have been reported. The aim of the Pharmacogenetics of Adrenal Suppression with Inhaled Steroids (PASS) study was to undertake a pharmacogenomics assessment of factors predisposing to corticosteroid-induced adrenal suppression among children with asthma using ICS as part of their treatment".[4]

3. Provide reasons for choosing the genes and SNPs genotyped.

Explanation

It is important that researchers conducting a candidate gene study choose the genes and SNPs to be investigated in a systematic way, using prior knowledge to guide their decisions. Study authors should explain how the investigated genes and SNPs were chosen, with reference to relevant functional/animal studies, previous association studies, and any procedures used, such as examining linkage disequilibrium patterns (the "tagging SNP" approach), or assessing the likelihood of each individual SNP affecting the gene function with priority given to those with the most likely functional effect. Clear rationale instils confidence in the reader that all analyses performed have been reported, rather than only statistically significant or interesting results (i.e. selective reporting of results).

It is important to note that for next generation sequencing (NGS), criteria may be applied to prioritise genetic variants for association analyses. In this case, filtering options ought to be specified.

The origin of this item is Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[5]

Example

"More recently, a novel putative mechanism of epilepsy multidrug resistance has been described. RLIP76, also known as RALBP1, is alleged to be a non-ABC multispecific transporter, which transports a variety of drugs, with a similar substrate specificity to PGP [P-glycoprotein]. A study by Awasthi et al. showed that RLIP76 was upregulated in brain tissue from drug-resistant epilepsy patients, and colocalized with PGP in endothelial cells. RLIP76 was shown to transport both phenytoin and carbamazepine in an isolated artificial liposome system and in crude membrane vesicles. Furthermore, RLIP76 blockade with anti-RLIP76 antibodies altered phenytoin and carbamazepine transport to a much greater extent

than did PGP blockade, indicating the potential importance of RLIP76 in AED [anti-epileptic drug] transport.

(...) All SNPs in the *RLIP76* gene in HapMap Phase 1, and exonic and untranslated region SNPs from dbSNP, were selected for genotyping. Primer design was successful for 23 polymorphisms (Figure 1), which were genotyped on the Sequenom MALDI-TOF mass extension platform at the Sanger Institute (...)".[3]

Introduction: Objectives

4. State specific objectives, including any pre-specified hypotheses.

Explanation

Study authors should provide the objectives for the study, specifying the relevant population, genetic variants, drugs and outcomes. This item is from the STROBE statement[1] (Item 3); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"The primary objective of this study was to determine, in a large patient cohort, whether c.516G>T and c.983T>C polymorphisms are predisposing factors for nevirapine hypersensitivity in a Malawian HIV-infected adult population. Secondly, we aimed to investigate whether carriage of *HLA-C*04:01* in combination with variants of *CYP2B6* increases the risk for nevirapine hypersensitivity".[6]

5. State if the study is the first report of a pharmacogenetic association, a replication effort, or both.

Explanation

This item is from the STREGA statement[7] (Item 3), although we modified the item to be more applicable to pharmacogenetic studies. In the STREGA statement, the item reads: "State if the study is the first report of a genetic association (...)". Further guidance and rationale for this item is detailed in the STREGA statement publication.

Example

"As far as we know, no studies examining the pharmacogenomics of corticosteroid-induced adrenal suppression have been reported. The aim of the Pharmacogenetics of Adrenal Suppression with Inhaled Steroids (PASS) study was to undertake a pharmacogenomic assessment of factors predisposing to corticosteroid-induced adrenal suppression among children with asthma using ICS [inhaled corticosteroids] as part of their treatment. Validation was undertaken in both a paediatric asthma cohort (enrolled to the PASS study) and an adult COPD [chronic obstructive pulmonary disease] cohort (enrolled to the Pharmacogenomics of Adrenal Suppression in COPD [PASIC] study)".[4]

Methods: Study design

6. Present key elements of study design early in the paper.

Explanation

Study authors should state the study design used, and present key features of the study design so that readers can understand the basics of the study, e.g. for a cohort study: describe the group of people that comprised the cohort and the time period for which they were followed; for a case-control design: describe the cases and controls and their source population; for a post-hoc pharmacogenetic analysis of a randomised controlled trial (RCT):

state how the subjects included in the analysis were chosen, including which arm of the RCT they were from. This item is from the STROBE statement[1] (Item 4); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"Blood samples, demographic and clinical data from patients initiating warfarin for venous thromboembolism or atrial fibrillation between November 2004 and March 2006 were collected, as described previously (…).

In this prospective cohort study (...), all patients received usual clinical care with doses being determined either by the anticoagulant clinic or attending physician. There were four fixed study visits for each patient, the first at the time of initiation of warfarin (index visit), then at 1 week, 8 weeks and 26 weeks of warfarin therapy".[8]

Methods: Setting

7. Describe the setting, locations and relevant dates, including periods of recruitment, follow-up, and data collection.

Explanation

Study authors should provide sufficient information to enable readers to assess the context and generalisability of a study's results. It is advisable to specify dates rather than length of time periods, i.e. the dates that recruitment began and ended, the dates that follow-up began and ended, and the date of data collection. This item is from the STROBE statement[1] (Item 5), although we modified the item to be applicable to pharmacogenetic studies. We removed the reference to periods of "exposure", as pharmacogenetic studies do not investigate the effects of "exposures", (other than genotype, which is fixed over time). Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"This prospective study was conducted at National Taiwan University Hospital, a tertiary-care center in Taiwan (...).

From March 2007 to February 2010, adult patients (>16 years) with culture-confirmed pulmonary TB [tuberculosis] were enrolled as the derivation cohort. Mycobacterial culture and drug susceptibility testing were performed as previously described. Subjects were excluded if they were pregnant, had a life expectancy 6 months, had abnormal baseline liver function test (LFT), or had Mycobacterium tuberculosis (MTB) isolates resistant to INH [isoniazid], RMP [rifampin], or both. From March 2010 to February 2013, TB patients fulfilling these criteria were enrolled as the validation cohort (...).

The LFT was checked at 2, 4, 6, 8, 12, and 16 weeks after the start of anti-TB treatment or whenever symptoms of hepatitis developed during the initial 6 months of anti-TB treatment".[9]

Methods: Participants

8. Give the eligibility criteria, and the sources and methods of selection of participants. For a cohort study, describe methods of follow-up. For a case-control study, state whether true controls or population controls were used. Give the rationale for the choice of cases and controls.

Explanation

Study authors should provide sufficiently detailed descriptions of the study participants to help readers understand the applicability of the results. In a case-control study, true controls are controls who have been exposed to the relevant treatment but have not developed the outcome of interest. If historical controls have been used, specify the setting in which this data was collected. Population controls are individuals who have already been genotyped that can be assumed to be controls, although we cannot ascertain whether they would have developed the outcome of interest if they had been exposed to the relevant treatment.

This item is derived from the STROBE statement[1] (item 6a); we added the specification that study authors should state whether true controls or population controls were used, as the use of population controls is a common feature in pharmacogenetic studies. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example: Cohort study

"Patients initiated onto warfarin irrespective of indication were recruited from the Royal Liverpool and Broadgreen University Hospitals Trust and University Hospital Aintree between November 2004 and March 2006. The only exclusion criterion was inability or refusal to give informed consent (…).

There were four fixed study visits for each patient, the first at the time of initiation of warfarin (index visit), then at 1 week, 8 weeks and 26 weeks of warfarin therapy (…). At the index visit, patient demographics were recorded and baseline INR [international normalised ratio], clotting factor activity and protein levels were measured (Table 1). At the remaining follow-up visits INR was again measured, and dose changes since the previous visit were recorded. In addition to the four fixed study visits, patients also attended anticoagulant clinic according to their clinical needs. This meant that, at the end of follow-up, data on warfarin dose changes and INR levels were available longitudinally for each patient, which together provided a complete picture of treatment progress from warfarin initiation onwards. For patients who missed one or more fixed follow-up visits, INR measurements and dose changes missing as a consequence were obtained from clinical records".[10]

Example: Case-control study with population controls

"From a cohort of ~600,000 patients receiving statins identified in the CPRD [the UK Clinical Practice Research Datalink] (www.cprd.com), a case-control design was used to identify suitable patients for the study, as previously described. Participation was restricted to white people ≥18 years of age and with the first ever statin prescription at least 1 year after the start of CPRD data collection.

All cases conformed to internationally agreed standards for statin-induced myopathy and rhabdomyolysis. Cases were categorized into two groups: (i) myopathy: patients who discontinued their implicated statin with a rise in CK [creatine kinase] > 4 × ULN [upper limit of normal]; and (ii) severe myopathy: individuals with a history of rhabdomyolysis or CK > 10 × ULN after statin exposure (…).

Population control genotype data for the initial discovery case-control GWAS [genome-wide association study] was obtained from the Wellcome Trust Case-Control Consortium 2 (WTCCC2) cohort of 2,501 individuals from the UK Blood Service".[11]

Example: Case-control study with true controls

"The study cohort consisted of older adults with polypharmacy and history of cardiovascular disease to ensure homogeneity of the study sample. The cases were presented by eligible individuals with the history of FH [frequent hospitalisations]. The controls included eligible patients with infrequent hospitalizations (IHs) randomly drawn from the study cohort, based on case–control matching criteria. On the basis of a previous work, cases (FH) were defined as individuals who were hospitalized at least 3 times during the past 2 years".[12]

This article provides a reference to previous work conducted to justify their choice of cases and controls.

9. Report the drug and regime participants were exposed to, and the length of exposure.

Explanation

The purpose of pharmacogenetic studies is to explore how genetic variants influence individuals' responses to drugs. Study authors should provide details of drug and regime, length of exposure, and route of administration. This could be fixed across all participants or variable. If variable, authors should provide this information in the text or in a table of patient characteristics. We recommend that authors use generic drug terms from a standardised database e.g. DrugBank where possible. This item was conceived by members of the Steering Committee.

Example

"All patients received oral INH (300 mg), rifampicin (600 mg), pyrazinamide (20 mg/kg body weight), and ethambutol (800 mg) daily for the first 2 months. Pyrazinamide was then discontinued, while INH, rifampicin and ethambutol were continued for another 4 months".[13]

10. For a matched case-control study, give matching criteria and the number of controls per case.

Explanation

Study authors should provide details of variables that were used to match individuals to make case and control groups more comparable. This item is from the STROBE statement[1] (item 6b), although we modified the item to be more applicable to pharmacogenetic studies. We removed the reference to matched cohort studies; following searches of the literature we found very few pharmacogenetic studies that used a matched cohort design, and this item would therefore be irrelevant to the vast majority of guideline users. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"For each case, two controls were recruited, within the same cohort and on the same ATD [anti-tuberculosis drug] (INH, RIF, PZA) [isoniazid, rifampicin, pyrazinamide] but with serum ALT [alanine aminotransferase] levels <3 times ULN [upper limit of normal], serum bilirubin <1 mg/dL and no history of severe nausea, vomiting within the first 3 months of initiation of therapy. Controls were matched with cases on the basis of age, sex, disease severity and drug dosage".[14]

11. Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.

Explanation

If one or more sub-samples from a larger study are used for the investigation of a pharmacogenetic association, authors should provide details of: inclusion and exclusion criteria, sources and methods of selection for these sub-samples, and state whether these methods were pre-specified or post-hoc. This item is from the STREGA statement[7] (Item 6a); rationale for this item is detailed in the STREGA statement publication.

Example

"In a post hoc analysis of a 4.3 year placebo-controlled randomized trial with 390 patients with Type 2 diabetes [T2D] already on insulin, we analyzed the influence of polymorphisms in genes coding for ATM and the transporters OCT1 and MATE1 (…). The HOME trial is a 4.3 year randomized placebo controlled trial that included 390 Caucasian patients aged 30–80 years with T2D treated with insulin. Patient selection, study design, data collection and power analysis have been described previously. Figure 1 shows the trial design and the recruitment plus retention of patients for the current study".[15]

Figure 1 of this publication provides information on the criteria for including patients in the post-hoc pharmacogenetic analysis of the RCT.

12. If other publications report results for the same patient cohort, or a subset of the patient cohort, provide information on this patient cohort overlap and references to the relevant publications.

Explanation

In pharmacogenetic research, it is common for multiple articles to report data for the same patient cohort (or for overlapping patient cohorts); different articles may report on different outcomes and genetic variants. To aid interpretation, it is useful to highlight overlapping or identical cohorts across articles. This item was conceived by members of the Steering Committee.

Example

"Given the association between *CYP4F2* and warfarin dose requirements, and the emerging evidence that this P450 isoform is involved in the metabolism of vitamin K1, we have undertaken a comprehensive analysis of *CYP4F2* SNPs and haplotypes in a prospectively recruited cohort of patients from two UK clinics (...).

Patients (n = 311) were recruited prospectively as they were initiated onto warfarin at two hospitals in Liverpool, the Royal Liverpool and Broadgreen University Hospitals Trust and University Hospital Aintree (...). Analyses of association between the *CYP2C9* and *VKORC1*, and 27 other genes, and the response to warfarin are reported in the accompanying manuscript".[16]

The article provides a reference to the accompanying manuscript in the text.

13. Report disease/clinical indication of patients using a standardised ontology when possible.

Explanation

To aid interpretation, authors should use controlled vocabularies such as MeSH and SNOMED to describe the disease/clinical indication of patients. This item was suggested by Delphi participants at Round 1 of the survey.

Example

"We included children aged 5–18 years with asthma using ICS [inhaled corticosteroids] as part of their treatment".[4]

Methods: Variables

14. Clearly define all outcomes, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.

Explanation

Study authors should define all outcomes, and all variables considered for and included in the analysis. If outcomes are categorised into "primary" and "secondary" outcomes, this should be specified. Authors should use controlled vocabularies such as MeSH and SNOMED to describe phenotypes.

This item is from the STROBE statement[1] (item 7), although we modified the item to be applicable to pharmacogenetic studies. We removed reference to "predictors" from the item as this is not relevant to pharmacogenetic studies, and to "exposures", as the exposure in a pharmacogenetic study is the genetic variant; definition of genetic variants is covered in the STROPS guideline item 16. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"For this GWAS [genome-wide association study], we used the following outcome measures: (1) Mean weekly dose (MWD): mean dose received weekly during a minimum follow-up time of 14 days post-loading; the loading period, that is, the first 3 days of treatment, was not included in the calculations. (2) Stable mean weekly dose (SMWD): mean weekly dose for at least three consecutive visits where INRs [international normalised ratios] were within the targeted range, spanning a minimum of 14 days and with at least 7 days separating the first and middle INR measurements, and the middle and last one. (3) INR >4.0 in the first week on warfarin (...).

Non-genetic variables used for testing univariately for association with each outcome were age, height, weight, BMI [body mass index], gender, loading dose, total follow-up time, dosing method (manual or computerised), mean target INR, blood count (haemoglobin, platelets, white cells, neutrophils, basophils, lymphocytes, monocytes, eosinophils), potassium, bicarbonate, chloride, urea, creatinine, triglycerides, albumin, total protein, bilirubin, ALT [alanine transaminase], alkaline phosphate, gamma GT, fibrinogen, coagulation factors II, V, VII, IX and X, Proteins C and S, current smoking status, number of cigarette smoked per day, ex-smoker status, alcohol consumption, interacting co-medication (binary), non-interacting co-medication (binary), sum of effect of interacting co-medications. The coagulation factors were measured as described by Jorgensen et al. For each variable, either a linear (quantitative outcomes) or logistic (binary outcome) regression was used to test for association with outcome in R, and variables found to be significant univariately (P ≤0.05) were included as covariates in the linear or logistic regressions used to test for association between each SNP and outcome in turn".[8]

15. Provide justification for choice of outcomes.

Explanation

Study authors should explain why the outcomes are important, e.g. clinical importance, importance to patients, inclusion in previously developed core outcome sets, identification of a significant association in previous studies, etc. A core outcome set is an agreed

standardised set of outcomes that should be measured and reported, as a minimum, in all clinical studies in specific areas of health or health care. The COMET database[17] lists references to planned, ongoing and completed core outcome set work for a wide range of health topics. Providing a clear justification for the choice of outcomes provides reassurance to the reader that selective reporting of results has not occurred, i.e. results have not been omitted from the report due to the significance or perceived importance of the estimate of association. The origin of this item is Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[5]

Example

"Guidelines for TB management recommend a combination regimen including isoniazid (INH), rifampin, ethambutol, and pyrazinamide as the first-line treatment. This regimen often causes adverse drug reactions, such as hepatitis, cutaneous reactions, gastrointestinal upset, and drug fever. Although mild reactions can be tolerated or managed with symptomatic therapy, serious cases require discontinuation of medication and prolongation of the treatment period (…).

Previous studies in an Indian and a Taiwanese population reported that homozygous null mutations in *GSTM1* increased the risk of ATD-induced hepatitis. In contrast to these findings, a subsequent study in Spain failed to validate this association, reporting instead that null mutations in *GSTT1* were associated with ATD-induced hepatotoxicity. Thus, the association between null mutations of *GSTT1* or *GSTM1* and ATD-induced hepatitis remains unclear and needs to be replicated in other ethnic groups.

ATD-induced cutaneous reactions, such as rashes, can be serious adverse reactions and their incidence is higher than that of hepatitis and gastrointestinal reactions. Despite the clinical significance of ATD-induced cutaneous reactions, not much is known about genetic predisposition to these reactions. It is suggested that, like drug-induced liver injury, hypersensitivity reactions to reactive metabolites underlie the mechanisms of drug eruption. Langerhans cells and epidermal keratinocytes have been suggested to play pivotal roles in the development of drug-induced hypersensitivity reactions in the skin. Drug metabolites transferred into or bioactivated in the skin can induce an immune response after haptenization. Detoxification by GST enzymes may be involved in the development of cutaneous reactions to ATD. Therefore, we hypothesized that null mutations of *GSTT1* and *GSTM1* genes are associated with ATD-induced cutaneous reactions. To our knowledge, there is no published report on the association between genetic polymorphisms in GST enzymes and ATD-induced skin reactions. In this study, we examined whether null mutations in *GSTT1* and *GSTM1* were associated with the development of ATD-induced hepatitis and adverse cutaneous reactions in a Korean population".[18]

16. Clearly define genetic exposures (genetic variants) using a widely-used nomenclature system.

Explanation

Study authors should state all genetic variants that were screened (including SNPs, indels, copy number variations, and structural variations) and specify the tissue source of DNA, as recommended by McDonagh et al.[19] This information could be provided in supplementary materials if necessary. It is also useful for study authors to provide references to any databases and resources used for the selection of variants.

The Human Gene Nomenclature Committee have published guidelines for human gene nomenclature.[20,21] The Human Genome Variation Society has detailed information about how to describe variant locations (http://varnomen.hgvs.org/recommendations/general/), as recommended by McDonagh et al.[19] and Thorn et al.[22]

This item is from the STREGA statement[7] (Item 7b); further guidance is detailed in the STREGA statement publication.

Example

"Venous blood sample collection followed informed consent from the parents or guardians of the participants and DNA was extracted using a standard salting-out method (…).

SNPs were submitted to Illumina (CA, USA) Technical Support for evaluation using the Assay Design Tool. SNPs were scored (varying from 0–1) by the Assay Design Tool based on compatibility to successful GoldenGate genotyping. A total of 14 SNPs with a score above 0.6 were selected for genotyping (summarized in Table 1)".[23]

Table 1 of this article gives gene names, rs numbers, and HGVS names for all the genotyped SNPs.

17. Report the rs number of each genotyped SNP.

Explanation

An "rs" number (reference SNP ID number) is an identification tag assigned by NCBI (National Center for Biotechnology Information) to a group (or cluster) of SNPs that map to an identical location. The rs ID number, or rs tag, is assigned after submission of an SNP to dbSNP.[24] Reporting an rs number for each genotyped SNP allows the reader to identify the same SNP across multiple articles. Guidelines are available for variations not listed in dbSNP.[25,26] In particular, submission ID or the position of the SNP and reference sequence ID for the chromosome should be specified. This item was conceived by members of the Steering Committee.

Example

"A total of 448 individuals were genotyped for the rs4149056 SNP in *SLCO1B1* and rs4693075 in *COQ2*".[27]

18. Clearly state how haplotypes or star alleles were defined.

Explanation

Even when referring to well-defined star alleles or haplotypes, authors should provide full details in the current publication using standard nomenclature. This allows the reader to identify the same haplotypes/star alleles across multiple articles, and can provide reassurance that haplotypes/star alleles have been defined according to widely accepted nomenclature. Standard nomenclature for some haplotypes are available on the PharmVar database.[28,29] This item was suggested by Delphi participants at Round 1 of the survey.

Example

"*SULT4A1-1* status was assigned to all CATIE subjects (…) using rs2285162 (A) and rs2285167 (G) as the haplotype tagging SNPs".[30]

19. If referring to the minor, major, wild-type, mutant, reference, risk or effect allele of a variant, state which allele this is and for which given population/cohort.

Explanation

The allele and population should be clearly stated if using any of these terms. For many genetic variants, the minor/mutant/risk/effect (less frequent) allele in one population may be the major/wild-type/reference (more frequent) allele in a different population. This item was suggested by Delphi participants at Round 1 of the survey.

Example

"The risk allele (T) frequency was highest in East Asians…".[31]

Methods: Data sources/measurement

20. For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.

Explanation

Authors should provide information on how all confounders and outcomes were measured. It is also important to report any differences in how data were collected in different patient groups (e.g. cases and controls). This item is from the STROBE statement[1] (item 8); further guidance and rationale for this item is provided in the STROBE explanation and elaboration paper.[2]

Example

"Warfarin dose and INR [international normalised ratio] values for each POC [point of care] testing event is routinely collected for all patients using warfarin at our institution, and this data were extracted onto our study database. From this information, each patient's full dose and INR history could be determined.

Data were also collected on indication for treatment, target INR, age, gender, height, weight, BMI [body mass index], haemorrhagic complications, serial serum albumin concentration(s) (as many children were hypoalbuminaemic at onset of therapy), and height and weight measurements. Clinical data were collected from hospital notes and the cardiac liaison team database".[32]

21. Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory/centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.

Explanation

Authors should provide sufficient details to enable the reader to assess the potential extent of genotyping errors (a source of information bias). It is important to report any differences in laboratory methods in different patient groups (e.g. cases and controls). Furthermore, if the study uses a case-control design, authors should report whether cases and controls were put into mixed batches for genotyping purposes (to ensure genotyping quality is comparable across groups), rather than analysed in separate batches. It is important to note that for next generation sequencing (NGS), library preparation, instrument, coverage level, pipeline and tools for variant calling should also be specified. This item is from the STREGA statement[7] (Item 8b); further guidance and rationale for this item is detailed in the STREGA statement publication.

Examples

"Whole blood on FTA® cards was obtained from 37 deceased Finnish individuals (...). All subjects and toxicology data were collected according to the ethical handling of human subject practices of the University of Helsinki. Anonymized DNA samples were transferred to University of North Texas Health Science Center (UNTHSC) and handled according to the UNTHSC Institutional Review Board Protocol Number 2016-051 (...).

DNA was extracted from FTA cards using the QIAGEN® QIAamp® DNA Blood Mini Kit and total human DNA was quantitated using the ThermoFisher Scientific Quantifiler™ Trio DNA Quantification Kit according to the respective manufacturers' recommendations (...).

Genotyping was performed using the Illumina Infinium® LCG Assay and Infinium® Omni2.5Exome-8 v1.3 BeadChip according to the manufacturer's recommended protocol. Template DNA input ranged from 200 to 1 ng genomic DNA. Image acquisition was performed on the Illumina HiScan™ System using the iScan Control Software (...).

"BeadChip images were analyzed in GenomeStudio® Genotyping Module v2.0.2, following the manufacturer's recommended quality control procedures including a Genotype Call (GenCall) Score cutoff of 0.15."

This article[33] provides detailed information on call rates and evaluation of genotyping errors in supplementary materials to the publication.

"Samples pertaining to matched cases and controls were analyzed in the same batch, and laboratory personnel were unable to distinguish between cases and controls".[34]

22. Describe genotype quality control methods and findings.

Explanation

It is important to report any quality control methods and findings so that the reader is able to assess how reliable the genotyping results, and consequently, the findings of a study are. Genotype quality control methods include using negative controls, and re-genotyping all or a random sample of patients. The origin of this item is Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[5]

Example

"Samples were genotyped at Affymetrix's service laboratory on the Genome-Wide Human SNP Array 6.0. Genotype data quality control was via the standard protocol that was established for the WTCCC2 studies (supplementary methods). Specifically, concordance check was performed on 116 SNPs by 1779 individuals overlapped between this GWA [genome-wide association] data and the WTCCC1 T2D [type 2 diabetes] case control study. Based on the concordance rate of 99.73%, individuals with more than 10% discordance were removed from the current study. After such stringent QC [quality control], the clean data set included 705125 autosomal SNPs on 3736 samples, of whom 1024 have definable metformin response".[35]

23. For quantitative outcome variables, specify if any investigation of potential bias resulting from pharmacotherapy was undertaken. If relevant, describe the nature and magnitude of the potential bias, and explain what approach was used to deal with this.

Explanation

Bias from pharmacotherapy may occur when quantitative outcome variables are affected by treatment with drugs other than the study drug (e.g. outcome variables include biochemical markers of hepatotoxicity, and several patients are taking concomitant hepatotoxic medications). This item is from the STREGA statement[7] (Item 9b); further guidance and rationale for this item is detailed in the STREGA statement publication.

Example

"…use of drugs with anti- or pro-emetic effects may also influence the occurrence and intensity of nausea and vomiting in cancer patients, but have not been thoroughly investigated yet (…).

All regressions were stratified by country and use of antiemetics was included as a binary stratification variable (antiemetics used/not used) for each country".[36]

24. Report how adherence to treatment was assessed, and report the results of the assessment.

Explanation

In general, treatment adherence is not an issue of great concern for non-pharmacogenetic studies of drug efficacy. This is because the aim of these trials is to estimate how effective the drug will be when used in a real-world setting; in reality, patients are likely to occasionally be non-adherent with the prescribed regimen, so the trial results will be reflective of clinical practice. However, the aim of pharmacogenetic studies is to identify associations between genetic variants and drug response outcomes. Treatment adherence is therefore an important issue in pharmacogenetic studies; taking too much or too little of the prescribed drug undoubtedly may impact drug response outcomes. Providing information on treatment adherence assessment methods and results allows the reader to consider whether adherence may have had an impact on outcomes. The origin of this item is Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[5]

Example

"Treatment compliance was assessed by comparing the number of administered treatment doses to the number of treatment doses scheduled each month. There were no consistently missing doses corresponding to more than 10 days of monthly scheduled doses of medication for all the patients".[13]

Methods: Study size

25. Explain how the study size was arrived at, or provide details of the a priori power to detect effect sizes of varying degrees.

Explanation

Study authors ought to report the calculation performed to obtain the study sample size, providing references to any specific methodology. Or, if sample size was predetermined (for example, if the study reports secondary analyses of a published dataset), provide details of a priori power calculations for a range of plausible effect sizes. This item is derived from the STROBE statement[1] (item 10); we added the detail that providing power calculations for a range of plausible effect sizes is sufficient to address this item, as it is not uncommon for sample size to be predetermined in pharmacogenetic studies. For example, in a post-hoc pharmacogenetic analysis of a RCT, sample size would be limited by the number of participants included in the RCT who it would be possible to obtain genotype information for. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Examples

"As we are investigating a wide range of gene-outcome associations it is difficult to provide precise power calculations in advance. However, we can calculate power for some simplistic analyses. A key variable is the minor allele frequency (MAF) among controls. For rarer variants to be clinically important, their effect size (odds ratio (OR)) must be large. We therefore specify two benchmarks for the power analyses: we seek to have good power for (a) OR=3 and a rare variant (MAF=5%); (b) OR=2 and a common variant (MAF=20%). Assuming a type I error rate of 5% and 80% power, for scenario a) we would require 115 cases and 230 controls and for scenario b) we would require 123 cases and 246 controls. If we increased our sample size to 250 cases and 500 controls we would have 80% power to

observe and odds ratio of 2.2 for a MAF of 5%, and an odds ratio as small as 1.6 for a MAF of 20%. Please note that these effect sizes are for a single causal variant. We expect to realise much larger overall effect sizes via combinations of causal variants".[37]

"The target sample size for the primary analysis cohort was 500. To arrive at this estimate we considered two possible scenarios: (a) an odds ratio (OR) of 3 for association between a rare variant (minor allele frequency=5%) and the primary outcome; (b) an OR of 2 for association between a common variant (minor allele frequency=20%) and the primary outcome. A liberal type 1 error rate of 5% was assumed on the basis that validation cohorts would also be analysed to help eliminate false positives arising from the initial analyses. Assuming first of all prevalence of impaired adrenal response in children with asthma using inhaled steroid to be 17% (based on a the [*sic*] lower end of the range of previously published rates of adrenal suppression) the power for scenario a) was calculated as 77% and the power for scenario b) was calculated as 75%. If the prevalence was 40% (the upper end of the range of published estimates), the power increased to 91% for both scenarios. A prevalence of 20% would ensure power of at least 80% in both scenarios".[4] (Supplementary Appendix).

Methods: Quantitative variables

26. Explain how quantitative variables (confounders and effect modifiers) were handled in the analyses. If applicable, describe which groupings were chosen, and why.

Explanation

If continuous variables have been grouped into categories to create a new categorical variable, it is important to explain why and how quantitative data were grouped. This item is from STROBE[1] (item 11), although we added detail to make it clear that the item relates to confounders and effect modifiers, rather than quantitative outcome variables. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"The potential genetic risk scores ranged from 0 to 3 (…). The risk scores were dichotomized as low risk (0–1 points) and high risk (2–3 points) because only a small number of subjects had a score of 0 and 3, and they responded similarly to those with a score of 1 and 2, respectively".[38]

Methods: Statistical methods

27. Address the following:

a) Describe methods used to control for confounding.

b) Describe any methods used to examine subgroups and interactions.

c) Explain how missing data were addressed.

d) Cohort study – If applicable, explain how loss to follow-up was addressed.

e) Case-control study – If applicable, explain how matching of cases and controls was addressed.

f) Describe any sensitivity analyses.

Explanation

All statistical methods should be reported clearly, including details of which analyses were pre-specified and which were exploratory based on data inspection. Details of any software

used should also be provided. The STROBE explanation and elaboration document[2] advises that sufficient detail should be provided that a statistically competent reader with access to the data set would be able to verify reported results based on the reported methods. In particular:

a) If adjustments were made for confounding factors, authors ought to provide details of procedures of variable selection and model comparison.

b) Study authors should explain any methods used to examine whether associations differed across subgroups, or to examine interactions. An "interaction" occurs when one factor modifies the effect of another, and is also sometimes referred to as 'effect modification'.

c) Authors should confirm whether analyses were restricted to individuals with complete data on the required variables, or whether any imputation of missing data was performed. Data are said to be 'missing at random' if the fact that they are missing is unrelated to actual values of the missing data. Data that are missing at random may not be important. Analyses based on the available data will tend to be unbiased, but will also be based on a smaller sample size than the original data set. If authors assessed the assumption of missingness at random, the methods and findings of these assessments should be provided.

d) For a cohort study, authors ought to report how many individuals were lost to follow-up, and whether these individuals were excluded or whether censoring strategies were used.

e) For a matched case-control study, authors should describe in detail what statistical methods were used to account for the matching of cases and controls.

f) Provide details of any sensitivity analyses i.e. analyses performed to investigate whether the results of the main analysis are consistent with those obtained with alternative analysis approaches.

This item is from STROBE[1] (items 12a, 12b, 12c, 12d and 12e). Further guidance and rationale for this item is provided in the STROBE explanation and elaboration paper.[2]

Examples

a) "First, a univariate multinomial logistic regression model was fitted for each non-genetic factor in turn, to identify which non-genetic factors to adjust for in the SNP association analyses. Next, multivariable multinomial logistic regression models were fitted for each SNP in turn. For each SNP, two models were fitted. The first model included covariates to represent all non-genetic factors with $p<0.25$ univariately. Stepwise variable selection was applied to this baseline model to remove any covariates no longer significant in the multivariable model. The final model following variable selection was called the 'baseline model'. The second model ('the genetic model') was the same as the baseline model but also included a covariate to represent the SNP. The likelihood ratio test was applied to compare the two models and thus assess for statistical significance of the SNP".[39]

b) "In the multivariate logistic regression analysis, an interaction variable between sex and *PXR* genotypes and haplotypes was also included".[9]

c) "For some of the non-genetic variables there was a considerable amount of missing data (see Table 1). In order to minimise the impact of this, multiple imputation using chained equations, specifically the predictive mean matching method, was used to impute data for height, weight and albumin at the start of warfarin treatment (all variables had <30% missing observations). Multiple imputation was

not used for these variables at the time stable dose was achieved, as the amount of missingness was deemed too high (>40%). Instead, these variables were excluded from the list of potential covariates".[32]

d) "For analyzing EFS [event-free survival] and PFS [progression-free survival], patients who stopped imatinib or switched treatment during follow-up were censored at the time of stopping or switching. However, these patients could be informative, as their reason for stopping/switching is related to the events investigated. Therefore, 2 sensitivity analyses were undertaken to determine whether censoring was informative. The first assumed that these patients were at high risk of an event and that all censored observations were therefore EFS/PFS events occurring immediately after censoring. The second assumed that these patients were at low risk of an event and assumed that all EFS/PFS events happened after the latest follow-up; their censoring time was changed to the time of last follow-up".[40]

e) "The Wilcoxon signed-rank test was used while comparing the means of continuous variables from matched samples. The exact McNemar's test was used on paired nominal data".[12]

f) "Sensitivity analyses were also undertaken, in which each analysis was repeated, but cases without evidence of H. pylori infection were excluded".[41]

28. State whether Hardy-Weinberg equilibrium was considered and, if so, how.

Explanation

Authors should describe any statistical tests or measures of departure from Hardy-Weinberg equilibrium (HWE), and any methods used to allow for deviations from HWE. Where HWE tests have been undertaken, it is important to state the p-value threshold applied to determine deviation from HWE. This item is from the STREGA statement[7] (item 12f); further guidance and rationale for this item is provided in the STREGA statement publication.

Example

"Prior to analysis, each SNP was tested for Hardy–Weinberg Equilibrium (HWE) using Fisher's exact test. Those with a p-value of less than 0.001 were excluded from further analyses".[3]

29. Describe any methods used for inferring genotypes or haplotypes.

Explanation

Study authors ought to provide details of any statistical methods or software used to infer genotype phase and haplotypes. This item is from the STREGA statement[7] (item 12g); further guidance and rationale for this item is provided in the STREGA statement publication.

Example

"We used multiple imputation methods to infer remaining missing genotype values on the basis of the correlational structure of the observed genotypes (…).

2677G→T/A and 1236C→T were successfully genotyped in 96% and 94% of individuals, respectively (…). Based on these genotypes, haplotypes were inferred with PHASE".[42]

This article refers to the software program, PHASE,[43] which can be used to estimate haplotypes.

30. Describe any methods used to assess or address population stratification.

Explanation

Authors should explicitly document any methods used to assess the presence of population stratification or adjust for population stratification in the analyses. If no methods were used, this should be made clear in the study report. This item is from the STREGA statement[7] (item 12g); further guidance and rationale for this item is provided in the STREGA statement publication.

Example

"In the primary analysis cohort, to test for association, regression models assuming an additive genetic model were fitted in SNPtest using each SNP as a covariate in an independent model. To adjust for population substructure we included up to five principal components as covariates in this genome-wide analysis, subject to the principal components being significantly associated with the outcome univariately ($p<0.05$). If no principal components were significantly associated with the outcome, the first two principal components were included as covariates in genome-wide analysis".[4] (Supplementary Appendix).

This article refers to the software program, SNPtest,[44] which can be used to analyse single SNP associations in genome-wide studies.

31. Describe any methods used to assess and correct for relatedness among subjects. Report results of assessments for relatedness.

Explanation

It is not uncommon in pharmacogenetic studies for some participants to be related. Authors should report any methods used to assess relatedness, results of these assessments, and any methods used to correct for relatedness. This item is from the STREGA statement[7] (item 12j), although we added the specification that results of assessments for relatedness ought to be provided. Further guidance and rationale for this item is provided in the STREGA statement publication.

Example

"Patients within the primary analysis cohort were excluded from association analyses if any of the following criteria were met: (...) c) the pairwise identity by descent (IBD) statistic of relatedness was >0.1875 (patient with lowest call rate of the pair excluded)".[4] (Supplementary Appendix).

"92 (18%) children failed genotype quality control, of whom (...) nine did not meet identity-by descent criteria".[4]

32. Describe any methods used to address multiple comparisons or to control risk of false positive results due to investigating:

a) multiple genetic variants

b) multiple outcomes

c) multiple assumptions regarding mode of inheritance.

Explanation

Pharmacogenetic studies that perform a large number of statistical tests are at risk of type 1 errors. This item is derived from STREGA[7] (item 12i), but we modified the item to specify different possible sources of multiplicity in pharmacogenetic studies, i.e. multiple

outcomes, genetic variants, and assumptions regarding mode of inheritance. Mode of inheritance refers to the way in which a genetic trait is passed from one generation to the next, e.g. dominant, recessive, co-dominant inheritance. Authors may undertake multiple analyses, each making a different assumption about the underlying mode of inheritance, or making no assumption about the mode of inheritance. Study authors should provide sufficient detail to enable the reader to assess the likelihood of false positive results (type 1 errors) being reported. Rationale for this item is provided in the STREGA statement.

Example

"To account for multiple testing, the false discovery rate (FDR) was calculated in addition to the P value for each test of association. In calculating the FDR, all tests for association undertaken on the dataset, including those referred to in an accompanying manuscript, were taken into account".[16]

33. Describe any methods used to adjust for extent of adherence in the analyses.

Explanation

As discussed under item 24, treatment adherence is an important issue in pharmacogenetic studies. It is important to take treatment adherence into consideration in statistical analyses for pharmacogenetic studies, as recommended by Jorgensen and Williamson in their quality assessment checklist for pharmacogenetic studies[5] (the origin of this reporting item).

Generally, adjusting for a non-confounding covariate (such as treatment adherence) can explain variability in the outcome, consequently reducing noise and increasing power to detect pharmacogenetic associations.[45] However, it is important to note that when the drug response outcome is binary, and individuals are recruited according to case or control status, adjusting for the covariate can actually reduce power.[46] Nevertheless, methods have been developed to overcome this issue,[47-49] which account for non-confounding covariates while increasing power to detect genetic associations in case-control studies. It is therefore advisable to account for treatment adherence when investigating pharmacogenetic associations, even in the analyses of case-control studies, providing careful consideration is given to the choice of analysis method.

Example

"Repeating analyses of genetic association but adjusting for adherence: For any of the outcomes found significantly associated with nonadherence in the univariate analyses, the analyses of association with each of the 196 SNPs as previously reported in Jorgensen et al. were repeated, but this time after adjusting for adherence. To do this, for each SNP in turn, two proportional hazard regression models were compared using the likelihood ratio test. The first model included a covariate representing adherence status; the second was the same but also included a covariate to represent the SNP.

Finally, for the outcome of stable dose, an alternative approach was also adopted where the outcome itself was reduced by the estimated proportion of doses missed and the analyses for association with clinical and genetic factors as reported in Jorgensen et al. repeated with this revised outcome. For testing for association with genetic factors, two tests of association were undertaken for each SNP. The first made no assumptions regarding the underlying mode of inheritance and used ANOVA [analysis of variance] to test for association and the second assumed an additive mode of inheritance and used univariate linear regression. For testing for association with clinical factors, Student's t-test was used for binary variables, ANOVA for categorical variables and linear regression for continuous variables".[50]

Results: Participants

34. Report the numbers of individuals at each stage of the study – e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed.

Explanation

Authors ought to report the numbers of individuals considered at each stage of recruiting study participants, alongside reasons for non-participation at each stage. This allows the reader to judge whether the study population was representative of the target population, and whether bias was possibly introduced. A flow diagram can be an efficient and transparent way to convey this information. This item is from STROBE[1] (item 13a). Further guidance and rationale for this item is provided in the STROBE explanation and elaboration paper.[2]

Example

"From October 2007 to June 2008, a total of 4488 newly diagnosed patients with sputum smear positive pulmonary TB [tuberculosis] were recruited from four provinces (Zhejiang, Guangxi, Chongqing, Jilin) in China (…). A total of 4304 patients finished the follow-up (…).

Patients with any of the following were excluded from the present study: (i) abnormal serum ALT [alanine aminotransferase], AST [aspartate aminotransferase] or total bilirubin levels before anti-TB treatment; (ii) carriers of the hepatitis B or C virus; (iii) alcoholic liver disease or habitual alcohol drinking; (iv) the concomitant use of hepatotoxic drugs; and (v) a history of chronic liver disease or systemic diseases that may cause liver dysfunction. Among the remained patients, those fulfilled the criteria of ATDH [anti-TB drug-induced hepatotoxicity] were assigned into the case group. Incidence density sampling method was adopted to select controls from patients free of ATDH up to the date when the paired cases were diagnosed with ATDH. For each ATDH case, four controls were randomly selected and matched with age (within 5 years old), sex, treatment history, disease severity, drug dosage and place of sample collection. Finally, 89 patients with ATDH and 356 matched controls were included in the study".[51]

Results: SNPs

35. Report any SNPs that were excluded from analysis, and provide reasons for these exclusions.

Explanation

Authors should provide explicit statements of why variants considered important initially were excluded from analyses, for example, due to excessive missing data.[5] This provides assurance to the reader that no additional investigations have been undertaken, and that all analyses have been fully reported. The origin of this item is Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[5]

Example

"Of the 23 SNPs genotyped across *RLIP76* (Table 1), six SNPs failed successful genotyping, three SNPs were found to have a MAF of less than 1% (indeed one was monomorphic), and one SNP was found to deviate from HWE".[3]

Table 1 of this publication provides the rs number of each genotyped SNP, whether the SNP was included in analyses, and reasons for exclusion where applicable.

Results: Descriptive data

36. Give characteristics of study participants (e.g., demographic, clinical, social) and information on potential confounders.

Explanation

Study authors ought to report participant characteristics (e.g. age, sex, ethnicity, special characteristics such as pregnancy and concomitant medications) with appropriate summary measures. For example, for continuous data, mean and standard deviation, or median and range; for dichotomous data, numbers and proportions. For a case-control study, this information ought to be provided for cases and controls separately. It is also important to indicate the number of participants with missing data for each variable.

This item is from the STROBE statement[1] (Item 14a), although we removed the reference to "exposures", as the exposure in a pharmacogenetic study is the genetic variant. We would expect information on patient genotypes to be reported with the outcome data; it is not mandatory to provide this information as part of the baseline characteristics of the study population. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"The baseline demographic factors including underlying comorbidities are summarized in Table 1. The majority of patients (n=309), 60% of whom were inpatients, had a target INR [international normalised ratio] range of 2–3, with the remaining two patients having a target range of 3–4. The majority of the patients were White with atrial fibrillation being the most common indication for warfarin therapy. There was significant variation in the loading doses prescribed. Sixty-three percent were given 10 mg on the first 2 days, 17% were given 10 mg on day 1 followed by 5 mg on day 2, 7% were given 3 mg on both days whereas 5% were given 5 mg on both days".[10]

Table 3 of this publication summarises key baseline characteristics for patients included in this study; footnotes are used to indicate the amount of missing data for each variable.

37. Cohort study – Summarize follow-up time, e.g. average and/or total amount.

Explanation

Average follow-up can be summarised using the mean and/or median follow-up time. Total amount of follow-up may be reported using total person-years of follow-up, or some indication of the completeness of follow-up.[52] This item is derived from the STROBE statement[1] (Item 14c); we modified the item slightly to indicate that average and/or total follow-up time would satisfy this criteria. Further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"In our study, patients were entered into the study at the time they started antiepileptic drug treatment, and were followed-up prospectively to determine their response to medication, including both seizure control and adverse events (...).

Follow-up ranged from 84 days to 2296 days (median 934, mean 1041)".[42]

38. Where HWE tests have been undertaken, highlight SNPs that deviate from HWE.

Explanation

Although deviations from HWE may be due to disturbing factors which the researcher has no control over, it is also possible that deviation may be caused by genotyping errors,[53] by the existence of population stratification or by biased selection of controls.[54] In the quality assessment checklist for pharmacogenetic studies (the origin of this item),[5] Jorgensen and Williamson recommend that study authors highlight any SNPs that were found to deviate from HWE.

Example

"The rs3813867 SNP was not in Hardy-Weinberg equilibrium and was therefore excluded from (...) analysis".[55]

39. Where population stratification is assessed, report the results.

Explanation

Study authors ought to report the results of any tests performed to detect the presence of population stratification. If tests determine that population stratification is present, this indicates that the study is at risk of confounding. Therefore, any associations detected may be spurious findings. It is essential that readers are informed of the results of any tests for population stratification to enable appropriate interpretation of study findings. The origin of this item is Jorgensen and Williamson's quality assessment checklist for pharmacogenetic studies.[5]

Example

"The data collected showed that the African, European, and Amerindian ancestry mean ratios were not significantly different between the two groups examined ($p > 0.05$). Figure 1 shows the individual parental ethnic contribution of the case group (patients with hepatotoxicity) and control group (patients without hepatotoxicity) estimated through 48 AIMs [ancestry informative markers]".[56]

Results: Outcome data

40a) For a cohort study, report all outcomes (phenotypes) investigated for each genotype category over time.

40b) For a case-control study, report numbers in each genotype category for all outcomes investigated.

40c) For a cross sectional study, report all outcomes (phenotypes) investigated for each genotype category.

Explanation

All outcome data should be reported clearly, including the amount of missing genotype/outcome data. This item is derived from STREGA[7] (item 15); we modified the item to specify that all investigated outcomes ought to be reported, to emphasise the importance of not selectively reporting results. Specific guidance on how to report outcome data for each type of study design are as follows:

a) Cohort study: For outcomes that relate to the occurrence of some event, report the number of events that occurred. If the risk of an event occurring changes over follow-up time, present the numbers and rates of events in appropriate intervals of

follow-up or as a Kaplan-Meier life table or plot. For other outcomes, present appropriate summary measures (e.g. means and standard deviations) over time.

b) Case-control study: Report numbers of cases and controls in each genotype category.

c) Cross-sectional study: For outcomes that relate to the occurrence of some event, report the number of events that occurred. For other outcomes, present appropriate summary measures (e.g. means and standard deviations).

Example: Cohort study

In a cohort study conducted by Ramsey et al.,[30] the authors report change in PANSS-T (Positive and Negative Syndrome Scale total score) (summarised by means and standard deviations in Table 2), response rate and completer status (summarised by percentage of responders and completers in Table 3) and weight gain (summarised by mean and standard errors in Figure 1) for each haplotype group.

Example: Case-control study

"Associations between -308G/A and ATD [anti-tuberculosis drug]-induced hepatitis: The genotype frequencies in the case and control groups (...) are presented in Table 2".[57]

Table 2 of this publication reports the number and percentage of cases and controls in each genotype group (AA, AG, or GG).

Example: Cross-sectional

In a cross-sectional study conducted by Ebid et al.,[58] the numbers of responders and non-responders to metformin and glimepiride combination therapy in each genotype group is presented in Table 2 of the publication for the two investigated SNPs (*SLC22A1* rs622342 and *ABCC8* rs757110).

41. If a study includes more than one ethnic group, provide the summary data specified in (40) per ethnic group.

Explanation

Due to the possibility of confounding by population stratification, it is advisable to present results stratified by ethnicity. Furthermore, a significant association between a SNP and a treatment response outcome does not necessarily indicate a causal relationship; it is possible that the association only exists due to the SNP of interest being in strong linkage disequilibrium with the causal SNP. Patterns of linkage disequilibrium vary from one population to the next,[59] and therefore differences in estimates of association may be observed between different populations.. This provides further rationale for stratifying results by ethnicity. This item was conceived by members of the Steering Committee.

Example

In a case-control study conducted by Ng et al.,[60] the number of cases and controls in each *NAT2* acetylator group (assigned by genotype) is presented in Table 3 of the publication, for the overall study population, for patients from Europe, and for patients from the Indian subcontinent.

Results: Main results

42. Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.

Explanation

Providing both unadjusted measures of association and measures of association adjusted for potential confounders enables readers to compare both measures and assess how the measure of association is impacted by adjusting for confounders. Study authors ought to list all potential confounder variables considered, and the criteria/rationale for excluding or including variables in statistical models. It is also useful for authors to:

- Clearly state which allele/genotype of the variant is associated with the phenotype and the direction of the association, when associations are reported
- Report results of any adjustments for multiple comparisons, for example, Bonferroni adjusted p-values, or false discovery rates.
- Report precise p-values for all associations, as opposed to only indicating whether an association was found to be statistically significant or not. For example, stating $p<0.05$ or $p>0.05$, or indicating statistical significance (or a lack of) by using asterisks (*) is not appropriate.

This item is from the STROBE statement[1] (Item 16a), further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

In a retrospective cohort study conducted by Higashi et al.,[61] the authors present unadjusted and unadjusted estimates in Table 5, and specify that warfarin daily dose is the only covariate included in the adjusted model. The authors explain that covariates were included in the model if the hazard ratio changed by more than 5% upon inclusion of the covariate in the model.

43. Report category boundaries when continuous variables were categorised.

Explanation

If continuous outcomes were categorised, study authors ought to report the range of values covered by each category. This item is from the STROBE statement[1] (Item 16b), further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"(…) patients were divided according to the HbA1C level obtained at the patient clinic appointment and classified into responders or non-responders. Responders were defined as patients who received metformin and glimepiride combination therapy for at least 6 months, and their HbA1C was less than 7%. Non-responders were considered to be on combination therapy for at least 6 months, and their HbA1C was equal or higher than 7%".[58]

Results: Other analyses

44. Report other analyses done – e.g., analyses of subgroups and interactions, and sensitivity analyses.

Explanation

It may be impractical to present detailed findings for all analyses performed; in this case, authors should present detailed results for important results only. Less important results can be summarised briefly in the text and detailed in full in supplementary materials. This item is from the STROBE statement[1] (Item 17), further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"In a sensitivity analysis in which only those patients without any evidence of H. pylori infection (n=376) were analysed as cases, there was no significant interaction seen with *CYP2C19*17* (P=0.068) (see Supplementary Data online)".[41]

45. If numerous genetic exposures (genetic variants) were examined, summarize results from all analyses undertaken.

Explanation

This item encourages authors to report results for all genetic variants that were investigated in the study, rather than selectively reporting only "interesting" or statistically significant results. Full results can be provided in supplementary materials if necessary. This item is from the STREGA statement[7] (item 17b), further guidance and rationale for this item is detailed in the STREGA statement publication.

Example

"In the logistic regression analyses using an additive model, only one SNP, *CYP2C19*17*, was significantly associated with the presence of PUD [peptic ulcer disease] (odds ratio 1.47 (95% confidence interval (CI) 1.12 to 1.92); P=0.005, Table 3)".[41]

Table 3 of this publication provides the results of logistic regression analyses for all analysed SNPs.

46. If detailed results are available elsewhere, i.e. in supplementary materials, state how they can be accessed.

Explanation

Study authors ought to report what results are available, and provide sufficient details that a reader would easily be able to locate these resources. This item is derived from the STREGA statement[7] (item 17c), although we modified the item to indicate that we are referring to supplementary materials to the study publication. Further guidance and rationale for this item is detailed in the STREGA statement publication.

Example

"Supplementary material is linked to the online version of the paper at http://www.nature.com/cpt".[41]

Discussion: Key results

47. Summarize key results with reference to study objectives.

Explanation

A short summary of the main findings of the study helps the reader to assess whether the author's interpretation and suggested implications are supported by the findings. This item is from the STROBE statement[1] (item 18); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"In this large study in Malawian and Ugandan adults treated with nevirapine, we have been able to show an association between nevirapine-induced SJS/TEN [Stevens-Johnson syndrome/toxic epidermal necrolysis] and the c.983T>C polymorphism. The carriage frequency in nevirapine control individuals was 16% compared with 32% in those with nevirapine-induced SJS/TEN (P=0.0005, FDR [false discovery rate]=0.015) (Table 4). This association was not observed with any other nevirapine-induced hypersensitivity phenotype".[6]

Discussion: Limitations

48. Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.

Explanation

Discussing the limitations of the study helps the reader to interpret the validity and health care relevance of the study findings. Limitations might relate to, for example, characteristics of included patients, methods of outcome measurement, multiplicity of analyses, missing data, etc. This item is from the STROBE statement[1] (item 19); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"A limitation of our study is the size of the replication cohort, which was smaller than the discovery cohort and therefore lacked the statistical power to truly replicate the association observed between c.983T>C and nevirapine-induced SJS/TEN [Stevens-Johnson syndrome/toxic epidermal necrolysis]. However, nevirapine hypersensitivity is a rare phenotype and it was difficult to identify a larger number of patients. The replication cohort also consisted of both Malawian and Ugandan patients, which may introduce some population stratification. However, our data show that the frequency of c.983T>C polymorphism in the overall combined Malawian discovery and replication patients (0.18) was comparable to that observed in the Ugandan patients (0.15)".[6]

Discussion: Interpretation

49. Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.

Explanation

When interpreting results, authors should consider potential sources of bias, residual confounding (due to unmeasured variables or imprecise measurement of confounders), the results of relevant sensitivity analyses and subgroup analyses. Authors should discuss the real range of uncertainty with regards to reported results, which is greater than the statistical uncertainty demonstrated by confidence intervals. This item is from the STROBE

statement[1] (item 20); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"Utilizing a variety of outcomes reflecting drug response, we have not demonstrated a significant association with any of 13 common SNPs in the *RLIP76* gene in the whole cohort, either when investigated univariately or by way of multiple regression, taking the genetic region spanned by the SNPs as a whole. Three SNPs demonstrated low p-values when undertaking the univariate analyses, but these, with p-values of 0.02-0.04, would not survive correction for multiple testing to account for the number of analyses being undertaken. Furthermore, each of these SNPs was found to be ''nominally significant'' for only one of any of the four outcomes investigated. Following backward variable selection, one SNP was found to be nominally associated with time to first seizure (SNP rs167897) and another was found to be nominally associated with time to 12-month remission (SNP rs12457094), but once again, the resulting p-values would not survive correction for multiple testing.

In the subgroup analysis, including only those patients on carbamazepine, the only AED [anti-epileptic drug] used in the SANAD cohort that may be a substrate for RLIP76, univariate testing of the 13 SNPs demonstrated four SNPs with p-values less than 0.05, but once again, no SNP demonstrated a significant association with more than one of the four outcomes, and furthermore, once again these p-values would not survive correction for multiple testing. When investigating the genetic region as a whole, a nominally significant association was found with the outcome of time to first seizure (p=0.05), and the resulting models following backward variable selection gave p-values of less than 0.05 for all four outcomes (minimum: p=0.007). One SNP in particular, rs329017, was retained in the final model for three of the four outcomes. While once again these p-values would not survive correction for multiple testing, this is the strongest evidence for any influence of *RLIP76* genetic variation on drug response; the smaller size of the carbamazepine-treated subgroup may have limited the power to detect a stronger association.

The lack of definitively positive findings suggests that *RLIP76* genotypes probably have no influence on drug response in epilepsy patients as a whole, which is consistent with those reported recently. Subgroup analysis in patients on carbamazepine alone, however, does raise candidate polymorphisms for further analysis. Even under the assumption that none of these results are significant, this does not exclude a genetic influence on drug response mediated by RLIP76 since common polymorphisms analyzed may not represent all common variants throughout the gene. Alternatively, genetic contributions may arise from rare variants, in which case our study may have lacked power to detect an association. In addition, genetic factors influencing RLIP76 expression or function may be remote from the *RLIP76* gene itself".[3]

The article provides a reference to a previous study which also investigated the association between *RLIP76* genotypes and drug response in epilepsy patients, and makes comparisons between the findings of the two studies.

Discussion: Generalisability

50. Discuss the generalisability (external validity) of the study results.

Explanation

Study authors ought to consider the extent to which the results of the study can be applied to other circumstances, i.e. different populations/settings/countries. This item is from the

STROBE statement[1] (item 21); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"Our data show that the frequency of c.983T>C polymorphism in the overall combined Malawian discovery and replication patients (0.18) was comparable to that observed in the Ugandan patients (0.15), and to that reported in a Mozambican population (0.14). Although genetic differences do exist between these African cohorts, it would appear from our study that *CYP2B6* c.983T>C is likely to be generalizable across other sub-Saharan-African populations".[6]

Other information: Study registration

51. State whether the study has been registered. If the study has been registered, provide details of the registry.

Explanation

Studies can be registered on many different official platforms; the most widely used platform is ClinicalTrials.gov. This platform provides information on how registering studies fulfils a number of purposes and benefits many different groups of people. This item was conceived by members of the Steering Committee.

Example

"Clinical Trials.gov Identifier: NCT 00824772".[62]

Other information: Ethical approval

52. Report whether ethical approval was obtained for the collection of genetic data.

Explanation

If ethical approval was obtained, authors should also report the committee that gave ethical approval and a reference ID. This item was conceived by members of the Steering Committee.

Example

"Ethical approval for this study was granted by North West 3 Research Ethics Committee (10/H1002/57)".[39]

Other information: Funding

53. Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.

Explanation

The role of the funders relates to which part of the study the funders took direct responsibility for, e.g., study design, data collection, analysis, drafting of manuscript, decision to publish. This item is from the STROBE statement[1] (item 22); further guidance and rationale for this item is detailed in the STROBE explanation and elaboration paper.[2]

Example

"Role of the funding source: The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication (…).

Acknowledgments: We thank the NHS Research and Development Health Technology Assessment Programme as sponsors of the SANAD study, all clinicians involved in the collection of blood samples, all involved at the Sanger Institute in DNA preparation, genotyping and bioinformatics, and the Wellcome Trust for their support. GDL was supported in part by a Neurology Entry/Exit Scholarship from the Guarantors of Brain. This study and DNA collection was funded by the Wellcome Trust".[42]

Other information: Databases

54. State whether databases for the analysed data are or will become publicly available and if so, how they can be accessed.

Explanation

If databases are available, study authors should provide sufficient details that a reader would easily be able to locate these resources. This item is from the GRIPS statement[63] (item 24); further guidance and rationale for this item is detailed in the GRIPS explanation and elaboration paper.[64]

Example

"The complete data set of genotypes and clinical variables, as well as the full genotype quality-control data, is available to registered PharmGKB users at www.pharmgkb.org (full data set accession number, PA162355460)".[65]

References

1. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147(8):573-7.
2. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg* 2014;12(12):1500-24.
3. Leschziner GD, Jorgensen AL, Andrew T, Williamson PR, Marson AG, Coffey AJ, et al. The association between polymorphisms in RLIP76 and drug response in epilepsy. *Pharmacogemonics* 2007;8(12):1715-22.
4. Hawcutt DB, Francis B, Carr DF, Jorgensen AL, Yin P, Wallin N, et al. Susceptibility to corticosteroid-induced adrenal suppression: a genome-wide association study. *Lancet Respir Med* 2018;6(6):442-50.
5. Jorgensen AL, Williamson PR. Methodological quality of pharmacogenetic studies: issues of concern. *Stat Med* 2008;27(30):6547-69.
6. Carr DF, Chaponda M, Cornejo Castro EM, Jorgensen AL, Khoo S, Van Oosterhout JJ, et al. CYP2B6 c. 983T> C polymorphism is associated with nevirapine hypersensitivity in Malawian and Ugandan HIV populations. *J Antimicrob Chemother* 2014;69(12):3329-34.
7. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, Von Elm E, et al. STrengthening the REporting of Genetic Association studies (STREGA): an extension of the STROBE statement. *Hum Genet* 2009;125(2):131-51.
8. Bourgeois S, Jorgensen A, Zhang EJ, Hanson A, Gillman MS, Bumpstead S, et al. A multi-factorial analysis of response to warfarin in a UK prospective cohort. *Genome Med* 2016;8(1):2. Doi:10.1186/s13073-015-0255-y.
9. Wang JY, Tsai CH, Lee YL, Lee LN, Hsu CL, Chang HC, et al. Gender-dimorphic impact of PXR genotype and haplotype on hepatotoxicity during antituberculosis treatment. *Medicine (Baltimore)* 2015;94(24):e982. Doi:10.1097/MD.0000000000000982.

10. Jorgensen AL, Al-Zubiedi S, Zhang JE, Keniry A, Hanson A, Hughes DA, et al. Genetic and environmental factors determining clinical outcomes and cost of warfarin therapy: a prospective study. *Pharmacogenet Genomics* 2009;19(10):800-12.
11. Carr DF, Francis B, Jorgensen AL, Zhang E, Chinoy H, Heckbert SR, et al. Genomewide association study of statin-induced myopathy in patients recruited using the UK Clinical Practice Research Datalink. *Clin Pharmacol Ther* 2019;106(6):1353-61.
12. Finkelstein J, Friedman C, Hripcsak G, Cabrera M. Pharmacogenetic polymorphism as an independent risk factor for frequent hospitalizations in older adults with polypharmacy: a pilot study. *Pharmgenomics Pers Med* 2016;9:107-16.
13. Cho H-J, Koh W-J, Ryu Y-J, Ki C-S, Nam M-H, Kim J-W, et al. Genetic polymorphisms of NAT2 and CYP2E1 associated with antituberculosis drug-induced hepatotoxicity in Korean patients with pulmonary tuberculosis. *Tuberculosis (Edinb)* 2007;87(6):551-6.
14. Chatterjee S, Lyle N, Mandal A, Kundu S. GSTT1 and GSTM1 gene deletions are not associated with hepatotoxicity caused by antitubercular drugs. *J Clin Pharm Ther* 2010;35(4):465-70.
15. Out M, Becker ML, van Schaik RH, Lehert P, Stehouwer CD, Kooy A. A gene variant near ATM affects the response to metformin and metformin plasma levels: a post hoc analysis of an RCT. *Pharmacogenomics* 2018;19(8):715-26.
16. Zhang JE, Jorgensen AL, Alfirevic A, Williamson PR, Toh CH, Park BK, et al. Effects of CYP4F2 genetic polymorphisms and haplotypes on clinical outcomes in patients initiated on warfarin therapy. *Pharmacogenet Genomics* 2009;19(10):781-9.
17. COMET Initiative. *Core Outcome Measures in Effectiveness Trials* [cited 2020 10 July]. Available from: http://www.comet-initiative.org/.
18. Kim S-H, Kim S-H, Yoon HJ, Shin DH, Park SS, Kim Y-S, et al. GSTT1 and GSTM1 null mutations and adverse reactions induced by antituberculosis drugs in Koreans. *Tuberculosis (Edinb)* 2010;90(1):39-43.
19. McDonagh E, Whirl-Carrillo M, Altman R, Klein T. Enabling the curation of your pharmacogenetic study. *Clin Pharmacol Ther* 2015;97(2):116-9.
20. Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S. Guidelines for human gene nomenclature. *Genomics* 2002;79(4):464-70.
21. Wain HM, Lush M, Ducluzeau F, Povey S. Genew: the human gene nomenclature database. *Nucleic Acids Res* 2002;30(1):169-71.
22. Thorn CF, Whirl-Carrillo M, Hachad H, Johnson JA, McDonagh EM, Ratain MJ, et al. Essential characteristics of pharmacogenomics study publications. *Clin Pharmacol Ther* 2019;105(1):86-91.
23. Dandara C, Lombard Z, Du Plooy I, McLellan T, Norris SA, Ramsay M. Genetic variants in CYP (-1A2,-2C9,-2C19,-3A4 and-3A5), VKORC1 and ABCB1 genes in a black South African population: a window into diversity. *Pharmacogenomics* 2011;12(12):1663-70.
24. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308-11.
25. Antonarakis SE, Group NW. Recommendations for a nomenclature system for human gene mutations. *Hum Mutat* 1998;11(1):1-3.
26. Dunnen JTd, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000;15(1):7-12.
27. Carr D, O'meara H, Jorgensen A, Campbell J, Hobbs M, McCann G, et al. SLCO1B1 genetic variant associated with statin-induced myopathy: a proof-of-concept study using the clinical practice research datalink. *Clin Pharmacol Ther* 2013;94(6):695-701.
28. Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE, et al. The Pharmacogene Variation (PharmVar) Consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database. *Clin Pharmacol Ther* 2018;103(3):399-401.

29. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Twist GP, Klein TE, Miller NA, et al. The evolution of PharmVar. *Clin Pharmacol Ther* 2019;105(1):29-32.
30. Ramsey TL, Liu Q, Brennan MD. Replication of SULT4A1-1 as a pharmacogenetic marker of olanzapine response and evidence of lower weight gain in the high response group. *Pharmacogenomics* 2014;15(7):933-9.
31. Farzan N, Vijverberg SJ, Hernandez-Pacheco N, Bel EH, Berce V, Bønnelykke K, et al. 17q21 variant increases the risk of exacerbations in asthmatic children despite inhaled corticosteroids use. *Allergy* 2018;73(10):2083-8.
32. Hawcutt DB, Ghani A, Sutton L, Jorgensen A, Zhang E, Murray M, et al. Pharmacogenetics of warfarin in a paediatric population: time in therapeutic range, initial and stable dosing and adverse effects. *Pharmacogenomics J* 2014;14(6):542-8.
33. Wendt FR, Rahikainen A-L, King JL, Sajantila A, Budowle B. A genome-wide association study of tramadol metabolism from post-mortem samples. *Pharmacogenomics J* 2020;20(1):94-103.
34. He L, Gao L, Shi Z, Li Y, Zhu L, Li S, et al. Involvement of cytochrome P450 1A1 and glutathione S-transferase P1 polymorphisms and promoter hypermethylation in the progression of anti-tuberculosis drug-induced liver injury: a case–control study. *PloS One* 2015;10(3):e0119481. Doi:10.1371/journal.pone.0119481.
35. Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL, Tavendale R, et al. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet* 2011;43(2):117-20.
36. Laugsand EA, Fladvad T, Skorpen F, Maltoni M, Kaasa S, Fayers P, et al. Clinical and genetic factors associated with nausea and vomiting in cancer patients receiving opioids. *Eur J Cancer* 2011;47(11):1682-91.
37. Pirmohamed M, Park K, Williamson P, Smyth R, Hawcutt D, Smyth A, et al. Protocol: MAGIC study. Molecular genetics of adverse drug reactions in paediatric patients [unpublished]. 2011.
38. McDonough CW, Gong Y, Padmanabhan S, Burkley B, Langaee TY, Melander O, et al. Pharmacogenomic association of nonsynonymous SNPs in SIGLEC12, A1BG, and the selectin region and cardiovascular outcomes. *Hypertension* 2013;62(1):48-54.
39. Thiesen S, Yin P, Jorgensen AL, Zhang JE, Manzo V, McEvoy ML, et al. TPMT, COMT and ACYP2 genetic variants in paediatric cancer patients with cisplatin-induced ototoxicity. *Pharmacogenet Genomics* 2017;27(6):213-22.
40. Giannoudis A, Wang L, Jorgensen AL, Xinarianos G, Davies A, Pushpakom S, et al. The hOCT1 SNPs M420del and M408V alter imatinib uptake and M420del modifies clinical outcome in imatinib-treated chronic myeloid leukemia. *Blood* 2013;121(4):628-37.
41. Musumba C, Jorgensen A, Sutton L, Van Eker D, Zhang E, O'Hara N, et al. CYP2C19* 17 gain-of-function polymorphism is associated with peptic ulcer disease. *Clin Pharmacol Ther* 2013;93(2):195-203.
42. Leschziner G, Jorgensen AL, Pirmohamed M, Williamson PR, Marson AG, Coffey AJ, et al. Clinical factors and ABCB1 polymorphisms in prediction of antiepileptic drug response: a prospective cohort study. *Lancet Neurol* 2006;5(8):668-76.
43. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68(4):978-89.
44. Marchini J, Band G. *SNPTEST* [cited 2020 23 September]. Available from: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.
45. Mefford J, Witte JS. The covariate's dilemma. *PLoS Genet* 2012;8(11):e1003096. Doi:10.1371/journal.pgen.1003096.
46. Pirinen M, Donnelly P, Spencer CC. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat Genet* 2012;44(8):848.

47. Zaitlen N, Lindström S, Pasaniuc B, Cornelis M, Genovese G, Pollack S, et al. Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet* 2012;8(11):e1003032. Doi:10.1371/journal.pgen.1003032.
48. Zaitlen N, Paşaniuc B, Patterson N, Pollack S, Voight B, Groop L, et al. Analysis of case–control association studies with known risk variants. *Bioinformatics* 2012;28(13):1729-37.
49. Zhang H, Chatterjee N, Rader D, Chen J. Adjustment of nonconfounding covariates in case-control genetic association studies. *Ann Appl Stat* 2018;12(1):200-21.
50. Jorgensen AL, Hughes DA, Hanson A, Van Eker D, Toh CH, Pirmohamed M, et al. Adherence and variability in warfarin dose requirements: assessment in a prospective cohort. *Pharmacogenomics* 2013;14(2):151-63.
51. Chen R, Wang J, Tang S, Zhang Y, Lv X, Wu S, et al. Association of polymorphisms in drug transporter genes (SLCO1B1 and SLC10A1) and anti-tuberculosis drug-induced hepatotoxicity in a Chinese cohort. *Tuberculosis (Edinb)* 2015;95(1):68-74.
52. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet* 2002;359(9314):1309-10.
53. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *Eur J Human Genet* 2004;12(5):395-9.
54. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol* 2003;56(4):297-303.
55. Costa GN, Magno LA, Santana CV, Konstantinovas C, Saito ST, Machado M, et al. Genetic interaction between NAT2, GSTM1, GSTT1, CYP2E1, and environmental factors is associated with adverse reactions to anti-tuberculosis drugs. *Mol Diagn Ther* 2012;16(4):241-50.
56. Fernandes DCRO, Santos NPC, Moraes MR, Braga ACO, Silva CA, Ribeiro-dos-Santos A, et al. Association of the CYP2B6 gene with anti-tuberculosis drug-induced hepatotoxicity in a Brazilian Amazon population. *Int J Infect Dis* 2015;33:28-31.
57. Kim SH, Kim SH, Yoon HJ, Shin DH, Park SS, Kim YS, et al. TNF-α genetic polymorphism− 308G/A and antituberculosis drug-induced hepatitis. *Liver Int* 2012;32(5):809-14.
58. Ebid A-HI, Ehab M, Ismail A, Soror S, Mahmoud MA. The influence of SLC22A1 rs622342 and ABCC8 rs757110 genetic variants on the efficacy of metformin and glimepiride combination therapy in Egyptian patients with type 2 diabetes. *J Drug Assess* 2019;8(1):115-21.
59. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002;3(4):299-309.
60. Ng C-S, Hasnat A, Al Maruf A, Ahmed MU, Pirmohamed M, Day CP, et al. N-acetyltransferase 2 (NAT2) genotype as a risk factor for development of drug-induced liver injury relating to antituberculosis drug treatment in a mixed-ethnicity patient group. *Eur J Clin Pharmacol* 2014;70(9):1079-86.
61. Higashi MK, Veenstra DL, Kondo LM, Wittkowsky AK, Srinouanprachanh SL, Farin FM, et al. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. *JAMA* 2002;287(13):1690-8.
62. Kim K-M, Kim H, Lim SH, Cheong SH, Choi E-J, Kang H, et al. Effect of genetic polymorphisms of OPRM1, ABCB1, CYP3A4/5 on postoperative fentanyl consumption in Koren gynecological patients. *Int J Clin Pharmacol Ther* 2013;51(5):383-92.
63. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ. Strengthening the reporting of Genetic RIsk Prediction Studies: the GRIPS Statement. *PLoS Med* 2011;8(3):e1000420. Doi:10.1371/journal.pmed.1000420.

64. Janssens ACJ, Ioannidis JP, Bedrosian S, Boffetta P, Dolan SM, Dowling N, et al. Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration. *Eur J Epidemiol* 2011;26(4):313-37.
65. International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New Engl J Med* 2009;360(8):753-64.

Glossary of terms

Term	Definition
Adverse effect	Undesired pharmacological outcome resulting from an intervention, such as medication or surgery
Allele	A variant form of a gene. For example, at a SNP where individuals may either have a 'C' nucleotide base or a 'T' nucleotide base, the possible alleles for this SNP are C and T.
Amino acid	Organic molecules which serve as the building blocks of proteins
Assay	Used in genetic research to perform genotyping
Candidate gene study	A study in which researchers investigate a small set of SNPs. SNPs are selected on the basis of how likely they are to be associated with the outcome of interest.
Chromosome	Thread-like structures that are made up of tightly coiled DNA. Human cells contain two sets of 23 chromosomes, with one set inherited from each parent.
DNA	DNA stands for deoxyribonucleic acid. DNA is a long molecule that contains an organism's genetic information.
Enzyme	Substance produced by an organism that increases the rate of biochemical reactions, without itself being altered in the process
Excretion	Process by which metabolic waste is removed from an organism
Gamete	Male or female reproductive cell that contains half the genetic material of the organism
Gene	Small section of DNA that contains the information required to build a specific molecule (usually a protein)
Gene expression	Process by which the information from a gene is used in the synthesis of a functional product (usually a protein)
Genome	The complete set of genetic information for an organism
Genome-wide association study (GWAS)	A genetic association study in which researchers genotype hundreds of thousands of SNPs to provide coverage of the entire human genome
Genotype	Noun: set of two alleles (one on each chromosome) observed at a particular locus Verb: process of identifying which two alleles an individual possesses at a particular locus
Heterozygous/heterozygote	An individual with different alleles on each chromosome at the locus for a particular SNP has the heterozygous genotype, or is a heterozygote.
Homozygous/homozygote	An individual with the same allele on each chromosome at the locus for a particular SNP has the homozygous genotype, or is a homozygote.
Intergenic region	Sections of DNA between genes. Intergenic DNA makes up a large proportion of the human genome, but mostly has no recognised function.
In vitro	Study of biological properties that occurs outside a living organism
Linkage disequilibrium	Extent to which the frequency of a combination of alleles at two loci differs to the frequency that would be expected if the loci were independent and associated randomly
Locus/loci	Specific physical location on a chromosome. The plural of locus is loci.
Metabolism	The complete set of biochemical processes that occur within an organism to sustain life
Mutant-type allele	The allele that is least commonly observed at a particular locus for a given population

Term	Definition
Natural selection	Process whereby heritable traits which increase an organism's chances of survival and successful reproduction become more common in a population over time
Nucleotide	Structural component of DNA. Each nucleotide consists of a nitrogenous base (adanine [A], cytosine [C], guanine [G], or thymine [T]), a sugar molecule, and a phosphate group.
Pharmacodynamics	Study of the physiological and biochemical effects of drug exposure on the body
Pharmacogenetic test	A test that is used to determine an individuals genotype at one or more loci. Results may be used to inform choice of treatment regimen.
Pharmacokinetics	Study of how the body affects an administered drug. Pharmacokinetics explores what happens to a drug from the moment it is administered up until the point that it is completely eliminated from the body.
Phenotype	Observable, physical properties of an organism including appearance, development, and behaviour, for example, blue eye colour, or A+ blood type. An individual's phenotype is determined by both genotype and environmental factors.
Platform (genotyping)	Technology which facilitates the process of genotyping
Primer	Short, single-stranded DNA sequence that is complementary to a known section of DNA. Primers are required for the replication of DNA during the genotyping process.
Promoter region	Sequence of DNA that initiates transcription of the genes that follow it. In other words, promoter regions are required to turn certain genes 'on' and 'off'.
Protein	Large molecules consisting of one or more chains of amino acids. All organisms produce proteins, which have a variety of functions, such as catalysing reactions, transporting oxygen, and defending organisms from infection.
Single nucleotide polymorphism (SNP)	A locus where the type of nucleotide base (A, T, C, or G) present can differ between individuals
Stratified medicine	An approach based on identifying groups of patients with distinct mechanisms of disease, or responses to treatment, which allows treatment of an individual to be tailored according to their specific characteristics
Trait	A feature or characteristic of an organism that can be determined by genes, the environment, or both, for example, eye colour or blood type
Wild-type allele	The allele that is most commonly observed at a particular locus for a given population