

**CLINICAL EVIDENTIAL AND METHODOLOGICAL
CHALLENGES OF EARLY ASSESSMENTS OF
NEW HEALTH TECHNOLOGIES**

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Pascale J. S. Dequen BSc MSc
Department of Health Sciences
University of Leicester

2016

Abstract

CLINICAL EVIDENTIAL AND METHODOLOGICAL CHALLENGES OF EARLY ASSESSMENTS OF NEW HEALTH TECHNOLOGIES

Pascale J. S. Dequen

This thesis explores the challenges of assessing the relative effectiveness of new health technologies earlier in their clinical development and the potential implications on health technology assessment (HTA), including health policy decision-making on the basis of economic decision models. Public appeal for rapid access to new medicines has increased pressures on regulators and payers to approve and market products often before appropriate measures of effectiveness are available. First, this thesis identifies the key evidential and methodological issues posed by early or accelerated regulatory approval, as well as any parallels found in the literature for conditional reimbursement and coverage with evidence. A review of international HTA and pharmacoeconomic methods guidelines is performed to draw on cross-country experience in dealing with evidentiary issues in evidence synthesis and cost-effectiveness (Chapter 2). A summary of methods used in HTA relevant to this thesis is provided in Chapter 3. Using three examples from different therapeutic areas, I explore the impact on HTA outcomes of i) subgroup and comparator selection (Chapter 4), ii) specific search strategies to identify indirect evidence for network meta-analysis (Chapter 5), and iii) bias adjustment techniques to include observational data in evidence synthesis (Chapter 6). Each chapter evaluates how the uncertainty in relative clinical estimates influences cost-effectiveness results. Using a simulation approach, Chapter 7 extends the example in Chapter 4—ticagrelor for acute coronary syndromes—to model evolving evidence within the context of HTA. The pivotal trial data is replicated and truncated at different time points, both in terms of follow-up and calendar time, to assess relative treatment effects and costs under different scenarios of ‘early’ HTA. This thesis illustrates how on-going regulatory changes impact clinical evidence considerations in HTA and how existing HTA methods can be adapted to allow for earlier product assessments and ensure timely access to new health technologies.

(300 words)

Acknowledgements

I would like to thank everyone who has accompanied me over the last four and a half years for their invaluable support, encouragement and advice. I extend special thanks to Prof. Keith Abrams and Prof. Nicola Cooper for all their time and imparted knowledge, it has been a real honour to learn from you both.

Since joining the University of Leicester, I have had the immense pleasure to work alongside esteemed colleagues and thank the Biostatistics Research Group for being such a supportive, friendly and intellectually stimulating team. In particular, thanks to Prof. Alex Sutton for welcoming me into the fold and inspiring me to start this PhD, and thanks to my office ‘mates’ over the years: Dr. Mark Rutherford for acting as my honorary third supervisor and being such a good friend, Dr. Michael Crowther for always pointing out my mistakes and sharing this fellowship experience with me, and Ms. Rhiannon Owen for all your technical expertise, for making every conference attendance a trip to be remembered, and for always believing—“you can do it!”.

I am grateful to the NIHR for giving me the opportunity to conduct this research and would like to acknowledge the members of my advisory panel: Prof. Tony Ades, Prof. Karl Claxton, Ms. Aline Gauthier, Dr. Elisabeth George, Dr. Wim Goettsch, Prof. Dyfrig Hughes, Dr. Clare McGrath, Dr. Zoe Philips, Prof. Adrian Towse and Dr. Allan Wailoo. I am also very grateful to the many colleagues at other institutions who have contributed to the work presented in this thesis: Dr. Hayley Jones, Dr. Ansgar Lange, Prof. Matthew Stevenson, and Dr. Rebecca Turner.

I would also like to thank my family, especially my parents, for supporting me throughout my entire education and for putting up with me being a ‘student’ a little while longer. Special thanks to my dear friends and fellow LSE alumni, Catherine Bennett and Yiling Jiang, for all your helpful feedback, for taking the time to proof-read this thesis and to troubleshoot my models, and for always being only an email or phone call away.

A last thank you to John, for everything.

Contents

Abstract.....	i
Acknowledgements.....	ii
List of Tables	vii
List of Figures.....	ix
Abbreviations.....	x
Chapter 1 Introduction.....	1
1.1 Aims of the thesis.....	1
1.2 Adaptive licensing.....	2
1.3 Health Technology Assessment	3
1.3.1 The National Institute for Health and Care Excellence	3
1.4 Managed entry agreements	4
1.5 Layout of the thesis	6
Chapter 2 Review of the literature on early drug evaluations and current international HTA landscape	8
2.1 Background	8
2.2 Literature review of early drug evaluations	8
2.2.1 Objectives	8
2.2.2 Methods	9
2.2.3 Results.....	10
2.2.4 Discussion.....	15
2.3 Review of pharmacoeconomic and HTA guidelines	16
2.3.1 Objectives	16
2.3.2 Methods	17
2.3.3 Search results	18
2.4 Discussion	29
2.4.1 Caveats and limitations.....	29
Chapter 3 Methods in HTA	31
3.1 Background	31
3.2 Bayesian methods.....	31

3.2.1	MCMC and Gibbs sampling	33
3.3	Meta-analysis	35
3.3.1	Fixed effect model	36
3.3.2	Random effects model	37
3.3.3	Extensions and limitations	40
3.4	Network meta-analysis	40
3.4.1	Assumptions	41
3.4.2	Extensions and limitations	42
3.5	Economic evaluation	43
3.5.1	Decision-analytical models	44
3.5.2	Markov models	46
3.5.3	Handling heterogeneity and uncertainty	47
3.5.4	Extensions and limitations	48
3.6	Discussion	50
Chapter 4 Choosing a relevant comparator: evaluating ticagrelor for ACS in Germany and the UK		52
4.1	Background	52
4.1.1	Ticagrelor in acute coronary syndromes	52
4.2	Objectives	54
4.3	International comparison of ticagrelor assessments	55
4.3.1	Methods	55
4.3.2	Results	55
4.4	Cost-effectiveness analysis	59
4.4.1	Methods	59
4.4.2	Results	74
4.5	Discussion	77
4.5.1	Caveats and model limitations	78
Chapter 5 Searching for indirect evidence: NMA of interventions for VTE prevention		79
5.1	Background	79
5.1.1	Apixaban for VTE prevention	79
5.2	Objectives	81
5.3	Systematic literature review	82
5.3.1	Methods	82
5.3.2	Results	88

5.4	Network meta-analysis	92
5.4.1	Methods	92
5.4.2	Results.....	93
5.5	Cost-effectiveness analysis	97
5.5.1	Methods	97
5.5.2	Results.....	101
5.6	Discussion	103
5.6.1	Caveats and model limitations.....	104
Chapter 6 Mitigating bias in observational data: meta-analysis of vertebral augmentation procedures.....		107
6.1	Background	107
6.1.1	Surgical augmentation procedures for VCFs.....	107
6.2	Objectives.....	109
6.3	Adjusted meta-analysis	110
6.3.1	Mortality data.....	110
6.3.2	Methods	114
6.3.3	Results.....	116
6.4	Cost-effectiveness analysis	119
6.4.1	Methods	119
6.4.2	Results.....	126
6.5	Discussion	151
6.5.1	Caveats and model limitations.....	151
Chapter 7 Simulating early clinical evidence: cost-effectiveness of ticagrelor over time		153
7.1	Background	153
7.1.1	Ticagrelor in acute coronary syndromes.....	153
7.2	Objectives.....	155
7.3	Simulation study.....	156
7.3.1	Methods	156
7.3.2	Results.....	162
7.4	Cost-effectiveness analysis	167
7.4.1	Methods	167
7.4.2	Results.....	168
7.5	Discussion	171

7.5.1 Caveats and model limitations.....	172
Chapter 8 Conclusions and further work	175
8.1 Summary of thesis.....	175
8.2 Caveats and limitations	178
8.3 Further work.....	180
8.3.1 Recommendations for future HTA practice.....	181
8.4 Conclusion.....	182
Appendix A.....	183
Appendix B.....	188
Appendix C.....	204
Appendix D.....	224
Appendix E.....	241
Appendix F	251
References.....	260

List of Tables

Table 1 Search strategy for OVID (May 7, 2013)	9
Table 2 Overview of HTA guidelines identified for review in Australia, Canada, France, the Netherlands, Sweden, and in the UK	20
Table 3 Summary of clinical evidence and methods in HTA guidelines in Australia, Canada, France, the Netherlands, Sweden, and the UK ¹²	26
Table 4 Summary of international HTA reports for ticagrelor	58
Table 5 Summary of clinical variables in one-year decision-tree	64
Table 6 Transition matrix for Markov model	65
Table 7 Summary of clinical variables in Markov model	66
Table 8 Conversion of results from Biondi-Zoccai <i>et al.</i> [2011] into relative risks for health economic modelling	68
Table 9 Event rates for ticagrelor and prasugrel in PLATO-INVASIVE subgroup analysis (1).....	68
Table 10 Hazard ratios for ticagrelor vs. prasugrel from IQWiG ITC for health economic modelling	69
Table 11 Event rates for ticagrelor and prasugrel in PLATO-STEMI subgroup analysis (2).....	69
Table 12 Quality-of-life values in one-year decision tree	70
Table 13 Quality-of-life values in Markov model	71
Table 14 Annual drug costs	72
Table 15 Mean health state costs per treatment for the decision tree	73
Table 16 Health state costs for the Markov model	73
Table 17 Adverse events costs	74
Table 18 Cost-effectiveness results (PSA means) for clopidogrel, prasugrel, and ticagrelor from reconstructed economic model.....	75
Table 19 Breadth-first search strategy	83
Table 20 Breadth-first search strategy for VTE prevention	84
Table 21 First order search strategy without rivaroxaban (cf. Table 19 search 1i)	85
Table 22 First order search strategy rivaroxaban only (cf. Table 19 search 1ii)	88

Table 23 Goodness-of-fit statistics for fixed and random effects NMA models by outcome and by network order	95
Table 24 Drug acquisition costs.....	100
Table 25 Cost-effectiveness results for apixaban, dabigatran etexilate, and rivaroxaban vs. enoxaparin 40mg for the base case (ITC), first and second network orders	102
Table 26 Overall mortality at different time points from RCTs	111
Table 27 Overall mortality at different time points from long-term claims databases.	113
Table 28 Fixed effect model results for PVP vs. OPM for the observational and randomised data extracted at 12 months.....	117
Table 29 Bias allowance model results for PVP vs. OPM according to σ^2	118
Table 30 Annual risks of vertebral fracture following an initial vertebral fracture based on age and T-Score on entry to the model	121
Table 31 Annual risks of hip fracture following an initial vertebral fracture based on age and T-Score on entry to the model.....	121
Table 32 Cost inputs in the economic model.....	125
Table 33 Cost-effectiveness results (PSA means) for PVP, BKP, and OPM.....	126
Table 34 Results from the AstraZeneca Weibull regression for the PLATO study [2010a]	158
Table 35 Results from Wiviott <i>et al.</i> for the TRITON-TIMI 38 study [2007]	162
Table 36 Summary HRs and 95%CIs for the base case simulation.....	163
Table 37 Summary percentage bias and MSE from the base case simulation	164
Table 38 Results from the Weibull regression for the PLATO-INVASIVE study	165
Table 39 Summary HRs and 95%CIs for the subgroup simulation.....	167
Table 40 ICERs for ticagrelor vs. clopidogrel from the reconstructed economic model for the base case simulation.....	169
Table 41 Cost-effectiveness probabilities for ticagrelor for the base case simulation .	170
Table 42 ICERs for ticagrelor vs. prasugrel from the reconstructed economic model for the subgroup analysis simulation.....	171

List of Figures

Figure 1 Economic model structure diagram.....	60
Figure 2 (a) Cost-effectiveness plane and (b) CEAC of ticagrelor vs. prasugrel in INVASIVE population (subgroup analysis 1)	76
Figure 3 (a) Cost-effectiveness plane and (b) CEAC of ticagrelor vs. prasugrel in STEMI (PCI) population (subgroup analysis 2)	76
Figure 4 Study selection flow diagram	90
Figure 5 Network of studies including only ‘closed loops’ based on search/network orders (* indicates multiple dosages).....	91
Figure 6 Odds ratios for all VTE/all-cause death, total DVT, and any bleeds from fixed effects NMA models	94
Figure 7 First, second and third order NMA inconsistency plots	96
Figure 8 Economic model structure diagram.....	98
Figure 9 Cost–effectiveness planes by network order	103
Figure 10 Power transform prior model results for PVP vs. OPM according to α	118
Figure 11 Economic model structure diagram.....	120
Figure 12 Cost-effectiveness acceptability curves according to α and σ^2 for scenario 2 assuming a mortality benefit up to 5 years	128
Figure 13 Cost-effectiveness acceptability curves for different values of α and σ^2 for scenario 2 assuming a mortality benefit up to 5 years	128
Figure 14 Cost–effectiveness plane for different values of σ^2 for scenario 2	150
Figure 15	157
Figure 16 Coverage across simulation scenarios for shape, scale, and treatment effect parameters	164
Figure 17a) MI	166
Figure 18 Cost-effectiveness plane for ticagrelor vs. clopidogrel for all simulation scenario	169
Figure 19 Cost-effectiveness acceptability curves for ticagrelor and clopidogrel for all simulation scenarios.....	170

Abbreviations

ACS	acute coronary syndromes
AIDS	acquired immune deficiency syndrome
AHRQ	Agency for Healthcare Research and Quality
AHTAPoL	Agency for Health Technology Assessment in Poland
AMCP	Academy of Managed Care Pharmacy
ANAES	Agence nationale d'accréditation et d'évaluation en santé (France)
ASA	acetylsalicylic acid, i.e. aspirin
AWR	approval with research
BMS	Bristol-Myers Squibb
CABG	coronary artery bypass graft
CADTH	Canadian Agency for Drugs and Technologies in Health (Canada)
CEA	cost-effectiveness analysis
CEAC	cost-effectiveness acceptability curve
CED	coverage with evidence development
CHMP	Committee for Medicinal Products for Human Use
CI	confidence interval
CMA	conditional market authorisation
CMIMG	Cochrane Comparing Multiple Interventions Group
CRD	Centre for Reviews and Dissemination
CrI	credible interval
CV	cardiovascular
CVZ	College van Zorgverzekeringen (Netherlands)
DIC	deviance information criterion

DSU	decision support unit
DVT	deep vein thrombosis
ECG	electrocardiogram
EE	economic evaluation
EFPIA	European Federation of Pharmaceutical Industries and Associations
EMA	European Medicines Agency
EQ-5D	EuroQol-5 dimensions scale
ERG	evidence review group
EThOS	UK E-Theses Online Service
EUnetHTA	European network for health technology assessment
FDA	U.S. Food and Drug Administration
GPRD	General Practice Research Database
HAS	Haute Autorité de Santé (France)
HIQA	Health Information and Quality Authority (Ireland)
HIV	human immunodeficiency virus
HLPF	High Level Pharmaceutical Forum
HoC	House of Commons
HRQoL	health-related quality of life
HTA	health technology assessment
HTAi	Health Technology Assessment international
ITC	indirect treatment comparison
ICER	incremental cost-effectiveness ratio
IMI	Innovative Medicines Initiative
INAHTA	International Network of Agencies for Health Technology Assessment
IPD	individual patient data
IQWiG	Institute for Quality and Efficiency in Healthcare (Germany)

ISPOR	International Society for Pharmaceutical Outcomes Research
KCE	Federaal Kenniscentrum/Centre fédéral d'expertise (Belgium)
LFN	Pharmaceutical Benefits Board (Sweden)
LMWH	low-molecular-weight heparin
MCMC	Markov Chain Monte Carlo
MEA	managed entry agreement
MI	myocardial infarction
MIMS	Monthly Index of Medical Specialities
MINAP	Myocardial Ischaemia National Audit Project
MTC	mixed treatment comparison
NHS	National Health Service
NICE	National Institute of Health and Care Excellence
NIHR	National Institute for Health Research
NMA	network meta-analysis
NPC	National Pharmaceutical Council
NSCLC	non-small cell lung cancer
NSTEMI	non-ST-segment-elevation myocardial infarction
OIR	only in research
ONS	Office of National Statistics
OR	odds ratio
OS	overall survival
PBAC	Pharmaceutical Benefits Advisory Committee (Australia)
PCI	percutaneous coronary intervention
PCORI	Patient-Centered Outcomes Research Institute
PE	pulmonary embolism
PFS	progression free survival

PHARMAC	Pharmaceutical Management Agency (New Zealand)
PICO	Patient, Intervention, Comparator/Control, Outcome
PPRS	Pharmaceutical Price Regulation Scheme
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PTS	post-thrombotic syndrome
QALYs	quality-adjusted life years
RCT	randomised clinical trial
REA	relative effectiveness assessment
RR	relative risk
SBU	Swedish Agency for Health Technology Assessment and Assessment of Social Services (Sweden)
SMC	Scottish Medicines Consortium
STEMI	ST-segment-elevation myocardial infarction
THR	total hip replacement
TKR	total knee replacement
TLV	Dental and Pharmaceutical Benefits Agency (Sweden)
UK	United Kingdom
VBP	value based pricing
VTE	venous thromboembolism
VTE	venous thromboembolism or venous thromboembolic event
ZIN	Zorginstituut Nederland (Netherlands)

Chapter 1

Introduction

1.1 Aims of the thesis

Public appeal for rapid access to new medicines has increased pressure on regulators such as the European Medicines Agency (EMA) to approve drugs earlier in their development cycle. Early approval allows a drug to be marketed before a robust clinical benefit has been demonstrated and measures of effectiveness are available, i.e. before the completion of Phase III studies. Since the early 1980s, the USA Food and Drug Administration (FDA) has established several special expedited development and review pathways in order to improve access to promising treatments and incentivise research and development investments [Kesselheim 2015]. Notably, in 1992, the FDA instituted the accelerated approval regulations to allow for earlier approval of drugs that fill an unmet medical need based on a surrogate or intermediate clinical endpoint [FDA 2014]. Orphan drug and breakthrough designations were also introduced in Europe and the USA, in an effort to facilitate the assessment and subsequent marketing of new and innovative medicine providing or predicted to provide noticeable clinical advances [Kesselheim 2015]. Recently the EMA launched an adaptive pathways approach to promote the early and progressive access for patients to new medicines [EMA 2016]. This practice of ‘adaptive licensing’ and ‘staggered’ or ‘managed’ approval relies on the iterative assessment of clinical data as it becomes available; early findings form the basis of a conditional approval given a presumed positive benefit-risk profile whilst confirmatory trials are required post-market to substantiate a full license [Eichler 2012, 2015, EMA 2016].

However, early drug approval must be accompanied by timely reimbursement decisions in order to ensure patients reap the intended benefits. In this context of change, the role of health technology assessment (HTA) agencies has been challenged and HTA will need to adapt to meet the new evidential and methodological requirements of conditional decision-making [Towse 2010]. Understanding how facilitated regulatory pathways will impact clinical evidence considerations and evaluating what methods will be needed to conduct HTA in this new environment is critical and very timely to ensure appropriate

evidence-based healthcare decisions. The primary aim of this thesis is to assess existing methods to address early evidentiary issues for HTA practice in view of the changing regulatory and reimbursement landscape.

1.2 Adaptive licensing

In 2014, the EMA launched an adaptive licensing pilot project to explore how adaptive pathways might be used within the current legislative and regulatory frameworks to optimise product development and potentially accelerate patients' access to new medicines [EMA 2016]. Adaptive licensing is defined as a “*prospectively planned, flexible approach to regulation of drugs and biologics*”; it refers to a new policy paradigm based on the staged approval of drugs with “*indications, coverage, and therapeutic value [...] revisited at several points along the clinical development pathway*” [Eichler 2015: p235]. Adaptive pathways will often consider an early approval for a restricted patient population based on lower evidence requirements and/or surrogate outcomes, followed by iterative phases of evidence gathering to later expand marketing authorisation to a wider indication [Miyamoto 2011, EMA 2014]

Expedited and adaptive regulatory pathways build on existing schemes that have been introduced by licensing bodies over the last 20 years such as FDA's and Health Canada's early-access initiatives, and EMA's 'approval under exceptional circumstances' and 'conditional marketing authorisations' (CMA) [Baird 2014, European Commission 2015]. Under CMA guidelines, the products considered for conditional approval target seriously debilitating or life-threatening diseases, as well as medicinal products to be used in emergency situations and orphan indications [EMA 2006]. Similar initiatives give preference to new—often first-in-class—drugs in areas of high unmet need, for example HIV, tuberculosis, and oncology.

An additional rationale for adaptive licensing is the early involvement of stakeholders [EMA 2016]. The EMA and FDA have established parallel scientific advice between manufacturers, regulators and HTA agencies to promote early dialogue, align evidence and product development requirements, and—hopefully—shorten timelines [Baird 2014].

1.3 Health Technology Assessment

HTA is a multi-disciplinary and evidence-based practice which aims to compare the effectiveness, appropriateness and cost of competing health interventions within a healthcare system [Drummond 2008, Sorenson 2008, Goodman 2014]. HTA organisations and networks have been set up to evaluate the use of new technologies and inform government and third-party decision-making authorities at a regional, national, and international level. HTA is now recognised as a legitimate policy tool and widely used to support coverage and reimbursement decisions worldwide. Furthermore, HTA practice has given rise to a very dynamic field of study producing cutting-edge research on data collection and evidence synthesis, decision-analysis modelling, quality-of-life measurements, etc. [Goodman 1999, Hailey 2003, Draborg 2005].

HTA agencies are often seen as an additional ‘gatekeeper’ or ‘hurdle’ to market access, but in recent years, they have also experienced pressures by patients, health professionals, and industry alike, to make earlier recommendations on promising new health interventions and work with stakeholders to develop new pathways for evidence generation in the ‘real-world’. Indeed, efforts by regulators to facilitate market authorisation are in vain if there is no corresponding willingness by HTA agencies and payers to reimburse the ‘accelerated’ products. For this reason, HTA requirements and methods must adapt to cope with early and evolving clinical evidence whilst decision-makers learn to face greater uncertainty and higher risks to address an apparent public health need [European Commission 2015].

Chapter 2 reviews the HTA guidelines from agencies worldwide and presents an overview of HTA practice in six countries: Australia, Canada, France, the Netherlands, Sweden, and the UK. I specifically focus on HTA practice in England using examples from technology assessments undertaken by the National Institute for Health and Care Excellence (NICE).

1.3.1 The National Institute for Health and Care Excellence

NICE is responsible for appraising new technologies and ensuring patient access to safe and effective medicines at a price that guarantees the best value for money for the National Health Service (NHS) England and Wales [NICE 2016]. Decisions based on NICE’s guidance are officially “*England-only*”; however, agreements are in place with Wales, Scotland and Northern Ireland for NICE to advise and provide certain products

and services as part of their remit to improve health and social care [NICE 2016]. The work of NICE usually begins once a technology, often a new pharmaceutical product, has received market authorisation, and precedes informal price negotiations between the Department of Health and manufacturers. In reality these processes are often intertwined and overlap in time. For example, early on in the product development, NICE Scientific Advice offers guidance to manufacturers to support future HTA submissions. The Technology Appraisal Committee within NICE assesses the cost-effectiveness and budget impact of new interventions, and the key determinants of the final price of a medicine at launch. However, on-going changes to the regulation of pharmaceuticals in Europe and the UK are shifting the role of HTA and subsequently the evidence burden and methodological needs to assess new products. Even the most forward thinking bodies, such as NICE, will need to change their practice.

1.4 Managed entry agreements

Eichler *et al.* point out that payers have responded to new regulatory pathways with analogous conditional reimbursement schemes including managed entry agreements (MEAs), coverage with evidence development (CED) and approaches “*to flexibly develop needed real-world effectiveness and value information*” [2015: p235]. MEA schemes, also referred to as risk-sharing agreements or patient access schemes, have been used in the UK and in other countries; however, at this time no European country systematically considers this approach for new products [Grimm 2016]. Ferrario and Kanavos describe a full-model MEA as follows:

“The concept of managed entry of new medicines goes from horizon scanning for new compounds which are likely to enter the market within the next 1 to 3 years, to forecasting use and expenditure of the new medicine, to HTA assessment, to pricing and reimbursement, to the development of MEAs and continues with post marketing studies and surveillance” [2013: p26]

The broader term MEA actually encapsulates most forms of conditional coverage and performance-linked reimbursement, and much like adaptive pathways, MEAs are progressive payment approaches that allow patient access under certain conditions [Goodman 2014]. Grimm *et al.* finds that most schemes will centre around two main dimensions: i) price adjustments and ii) further research [2016]. For example, CED grants restricted coverage to a new health technology on the basis that further data is collected

in parallel to provide more compelling and robust evidence of effectiveness, safety, or economic impact [Hutton 2007, Goodman 2014]. CED seeks to reduce the uncertainty around HTA outcomes of interest, as well as to gather ‘real-world’ evidence, without jeopardising access to patients most likely to benefit from the use of a technology; however, CED also includes coverage ‘only in research’ (OIR) which limits access to patients participating in a clinical trial. In 2012, Claxton et al. published a comprehensive report on the use of health technologies in the context of an appropriately designed programme of evidence development, focusing on the role of NICE and the type of assessments needed to inform an OIR or ‘approval with research’ (AWR) recommendation [2012]. Checklists were devised by authors to guide decision-makers and attempt to estimate the potential added benefits and opportunity costs of additional research associated with such conditional recommendations. Key considerations raised in NICE technology appraisals that resulted in OIR/AWR recommendations were also reviewed by Longworth et al. up to January 2010 [2013]. Longworth et al. found that OIR/AWR recommendations were generally used for procedures and devices (rather than pharmaceuticals) by NICE and that more common reasons cited to justify the need for additional evidence were uncertainty in the relative effectiveness of new product, or uncertainty about the long-term effects and adverse events, and often resulting lack of cost-effectiveness information [2013].

In the UK, the first risk-sharing scheme was introduced in 2002 to ensure the availability of the first line disease modifying drug therapies to patient with multiple sclerosis, despite a negative recommendation by NICE. Manufacturers agreed to share the financial risk with the Department of Health and Pickin *et al.* discuss the lessons learned from this scheme [2009]. Since then, a number of patient access schemes have also been announced for new drugs not found to be cost-effective under current NICE thresholds. On the other hand, performance-linked reimbursement links payment or price rebates to the achievement of predetermined health outcomes targets [Goodman 2014]. Goodman highlights further advantages of MEAs across stakeholder groups:

“They can enable access for certain types of patients for whom existing evidence suggests net health benefit, provide some financial compensation for generating better evidence sooner than in the absence of reimbursement, enable refinement of clinical technique and services delivery, and build

expertise and experience among physicians and other providers.” [2014: pX-20]

Regardless of the form MEAs take, they enable earlier access to new products at the point of entry; however, they are associated with a number of practical challenges and require an assessment of risk based on limited and uncertain evidence [Trueman 2010, Walker 2012]. In order for HTA to optimise the early use of a health technology and inform health-care payers’ decisions, it should be able to initially assess its relative effectiveness and cost-effectiveness, even under conditions of uncertainty, and on the basis of an unconventional evidence base.

1.5 Layout of the thesis

The thesis is organised as follows. Issues associated with immature and incomplete trial data for early drug evaluations were identified by reviewing the relevant literature and were discussed in Chapter 2. Chapter 2 also provides an overview of current guidelines for HTA methods and highlights the gaps in international practice, particularly in dealing with different sources of clinical evidence and evidence synthesis. Chapter 3 introduces the concept of Bayesian methods in HTA and briefly describes three common statistical and modelling techniques used in HTA and referred to throughout this thesis: i) meta-analysis, ii) network meta-analysis (NMA), and iii) economic evaluation. Three case studies were conducted in different disease areas in order to investigate the impact of a number of evidentiary issues previously identified in Chapter 2 and potential methodological solutions on the relative effectiveness and cost-effectiveness of health technologies. Chapter 4 investigates the impact of selecting different subgroups and comparators of interest on HTA outcomes using the cross-country example of ticagrelor in acute coronary syndromes (ACS). Chapter 5 evaluates how specific search strategies can be used to identify indirect evidence for NMA and whether network size can influence the relative effectiveness assessment (REA) of new products in the context of HTA. This retrospective analysis was undertaken for a recently approved anticoagulant—apixaban—in the prevention of venous thromboembolism after joint surgery. Chapter 6 explores the use of non-randomised evidence to inform Bayesian meta-analysis models to estimate the mortality benefit of vertebral augmentation procedures following an osteoporotic vertebral compression fracture (VCF). Lastly, in Chapter 7, data simulation was used to recreate individual patient data (IPD) for a pivotal Phase III trial and evaluate a drug based on an evolving evidence base at different time-points throughout its

lifecycle. The simulation study was an extension of the ticagrelor case study presented in Chapter 4. I conclude the thesis in Chapter 8 with a discussion, summarising the findings and limitations from each chapter, suggesting recommendations for early HTA guidance and outlining further work that could be undertaken.

Chapter 2

Review of the literature on early drug evaluations and current international HTA landscape

2.1 Background

Early drug evaluations raise a number of concerns for regulators, HTA agencies, and payers. Traditional market authorisation and HTA require the completion of at least one Phase III trial, but an earlier approval and assessment of product value may be based on a combination of non-randomised data, adaptive and Phase II trial results, and/or interim data from on-going Phase III studies. Given this heterogeneous evidence base, a number of methodological issues are anticipated to be associated with smaller sample sizes, short-term follow-ups, and surrogate endpoints. Moreover, early technology appraisals will need to estimate the relative effectiveness of a new health intervention conditional on immature and incomplete clinical data. This may lead to additional challenges with regards to evaluating indicated population and subgroups of interest, selecting appropriate comparators, performing evidence synthesis, or dealing with greater bias and uncertainty.

In this chapter, I review the literature on early drug evaluations and explore what HTA and pharmaceutical guidelines recommend as best practice to assess the relative effectiveness of new medicines. Since the majority of CED decisions have been limited to approved clinical trials and no standard practice is sanctioned in the UK [Carlson 2010]; I consider how guidance on HTA methods can overlap with methods to address early evidentiary issues and highlight gaps in current practice.

2.2 Literature review of early drug evaluations

2.2.1 Objectives

In order to understand on-going trends and emerging issues related to the early assessment of pharmaceuticals across stakeholder groups, I review the recent literature on early drug evaluations focusing particularly on methodological discussions. The specific section objectives are:

- i. to identify all relevant evidence on early drug evaluations;
- ii. to summarise the evidential and methodological issues raised in the literature for the early assessment of new medicines.

2.2.2 Methods

A systematic search of the literature was undertaken on May 7, 2013. Table 1 details the search strategy ran in OVID; a combination of free-text and MeSH/Emtree terms were used. The final strategy was refined over a number of iterative preliminary searches and duplicates were removed prior to citations being exported to Endnote X7. The search was carried out on the OVID platform to access the following sources:

- Medline® In-Process & Other Non-Indexed Citations;
- Medline® 1946 to Present;
- OLDMedline® 1946 to 1965;
- EMBASE 1974 to 2012 August 27.

Table 1 Search strategy for OVID (May 7, 2013)

#	Search terms	Hits
1	(licens* or authorisation\$ or authorization\$ or regist* or regulat* or access or entr* or legislat* or approval\$ or endorse* or market* or evaluat* or assess* or control\$).ab,kw,ti.	13,647,978
2	(early or earlier or adapti* or pragmatic or condition* or manag* or dynamic or continue\$ or stagger* or accelerate\$ or provisional or rolling or progress* or incremental or (post adj1 market) or post-market* or (life adj1 cycle) or life-cycle).ab,kw,ti.	8,799,618
3	1 ADJ 2	36,665
4	(pharmac* or intervention\$ or drug\$ or medicin* or treatment\$ or therap*).ab,ti.	11,628,592
5	3 AND 4	13,865
6	exp evidence based practice/	682,306
7	(eviden* or method* or data*).ab,ti.	12,986,839
8	6 OR 7	13,265,285
9	(issue\$ or problem* or challeng* or difficult* or bias* or concern*).ab,ti.	4,359,963
10	5 AND 8 AND 9	2,136
11	(addresses or autobiography or bibliography or biography or case reports or clinical trial or clinical trial phase i or clinical trial phase ii or clinical trial phase iii or clinical trial phase iv or controlled clinical trial or conference abstract or duplicate publication or erratum or in vitro or interactive tutorial or interview or multicenter study or patient education handout or personal narratives or portraits or randomized controlled trial or video audio media or webcasts).pt,sh.	5,062,284
12	10 NOT 11	1,578

#	Search terms	Hits
13	limit 12 to yr="1998 -Current"	1,342
14	remove duplicates from 13	847

Clinical trials were excluded from the search, thus specific study selection criteria could not be defined according to a PICO (Population, Intervention, Comparator/Control, Outcome) format. Abstracts were reviewed and full-text publications retrieved for citations with any mention of early assessments of drugs or devices, irrespective of wording. Citation searching and more restricted sub-topic searches were also employed to complement the literature review. Following consultation with information specialists, conference proceedings and additional web sources were surveyed including the Agency for Healthcare Research and Quality (AHRQ), EThOS (Electronic Theses Online System), the NIHR Horizon Scanning Centre, Opengrey.eu, and WorldWideScience.org. An *ad hoc* data extraction was performed concentrating on identifying evidential issues associated with early drug evaluations and potential methodological developments to help appraise new health technologies.

2.2.3 Results

Out of a total 847 search hits, 37 non-duplicate citations were selected based on abstracts and a final 13 full-text publications were included for review. Since adaptive licensing and managed entry schemes remain a topical and evolving area in health care policy, a search update was performed on February 1, 2016. However, even after broadening search filters, only an additional 7 publications since 2013 met the inclusion thresholds. Of the 20 studies reviewed, 11 were discussion papers presenting the perspectives or experiences of expert groups and key opinion leaders [Rawson 2000, Carroll 2008, Eichler 2008, 2012, 2015, Tolley 2010, Henshall 2011, Schneeweiss 2011, Wonder 2012, de Jong 2013, Jönsson 2013]. For example, Henshall *et al.* [2011] summarise discussions from the Health Technology Assessment international (HTAi) Policy Forum on the interactions between regulatory, HTA, and coverage processes at an early stage; whilst Carroll *et al.* [2008] describe issues raised from the PSI (Statisticians in the Pharmaceutical Industry) Discussion Group on Conditional Approval held in November 2006. Another 7 studies conducted retrospective analyses and case studies looking chiefly at the impact of CMAs on approval rates in the short and long-term, safety risks, research and development costs, and time to launch [Poole 2009, Arnardottir 2011, Davis 2011, Miyamoto 2011, Hoekman 2015, Liberti 2015, Scannell 2015]. Hoekman *et al.* [2015]

and Liberti *et al.* [2015] also used structured interviews and surveys to gather responses on facilitated regulatory pathways (i.e. adaptive licensing or CMA) from a variety of stakeholder groups including pharmaceutical companies, regulatory and HTA agencies, patient groups and others. One of the remaining publications was only available as a conference poster presentation [Kanniche 2015] and Chen *et al.* provided a statistical approach to model the progression-free survival (PFS) and overall survival (OS) relationship in the context of accelerated approvals of new cancer drugs in the USA [2012].

The majority of the evidence focused on endorsing new regulatory initiatives, primarily by the EMA and FDA, highlighting the rationale for such programmes as breakthrough or accelerated approvals in the USA, and adaptive licensing or CMAs in Europe. Most of the reasons put forward to justify the pressures on regulators to fast-track new products and prioritise disease areas of high unmet medical need have previously been highlighted in section 1.2 and barriers to the implementation of such initiatives are discussed in Chapter 8. However, a number of key discussion points were extracted from the review namely about the evidentiary trade-offs required by early assessments that raised concerns with regards to future trial designs, the use of observational data, as well as subgroup analyses and surrogate endpoints.

2.2.3.1 Trial designs

Existing regulatory pathways that offer companies an ‘exceptional’ opportunity to obtain accelerated approval are subject to similar conditions. First, promising results should be demonstrated in a Phase II trial, ideally randomised and double-blind, or in a planned interim Phase III trial analysis, with a high probability of a clinically meaningful outcome—at least 90% power [Carroll 2008]. That is, even at an early stage, a positive risk–benefit ratio must be established. Second, the provision of supplemental data post-authorisation to confirm preliminary findings is usually mandatory and should be planned *a priori* in view of a full application. Phase II trials can be single-arm studies, unblinded and/or uncontrolled; thus, confirmatory randomised trials remain the norm to gain a full licence.

Drawing inspiration from marketing requirements for medical devices, such as the Conformité Européenne (CE mark) in Europe, Stordeur *et al.* suggest that non-inferiority trial designs or the use of concurrent observational controls could facilitate the early assessment of drugs without dispelling a rigorous scientific evaluation [2013]. Whilst,

Wonder *et al.* assume looking back at the approval processes in Australia that active comparator trials designed for superiority would be more suited to managed entry schemes [2012]. Moreover, Carroll *et al.* caution of the practical, legal and ethical issues of conducting a randomised trial following a conditional approval, especially if positive and clinically meaningful results have been shown in Phase II or prior-completion of Phase III [2008].

2.2.3.2 *Non-randomised evidence*

In addition to the consideration of different trial designs, several studies mentioned the use of observational and historical data to fill in the gaps in early evidence [Eichler 2008, Tolley 2010, Chen 2011, Schneeweiss 2011, Stordeur 2013]. Despite, methodological challenges, Tolley submits that observational data and pragmatic RCTs have “(potentially) greater external validity” than traditional RCTs for decision making related to the market access of new pharmaceuticals [2010]. Chapter 6 section 6.3.1.2 summarises the inherent limitations associated with non-randomised evidence, namely confounding and other biases. Schneeweiss *et al.* provide an interesting discussion of these issues specifically in the context of post-launch evidence generation and utilisation [2011]. The authors introduce the notion of multi-level confounding and describe numerous patient-, physician- and health system-level factors that can not only influence the uptake and exposure of a new drug, but also influence the outcomes of interest [Schneeweiss 2011]. For this reason and other issues that could lead to biased conclusions from observational evidence alone, Eichler *et al.* emphasise that observational studies should complement RCTs [2008, 2015].

Bias modelling has been used to account for the presence of internal and external bias and to increase the validity of estimates when using observational data [Höfler 2007, Thompson 2011]. Adjusting for differences across studies in terms of internal bias allows the synthesis of otherwise incompatible studies. Internal bias, also known as a lack of rigour, implies variability in the use of randomisation, adequacy of allocation concealment, degree of blinding, and/or attrition levels [Turner 2009]. Furthermore, accounting for external bias implies modelling the relevance and generalisability of studies against a proposed research question. In many instances, the objectives set-out in Phase III studies differ from the ones previously hypothesised in observational studies, making multiple-bias modelling of both internal and external bias key to address issues of heterogeneity in the sources of evidence for early drug evaluations. Indeed, an

approach which allows both relevance, especially for Phase III RCTs, and rigour, for non-randomised evidence, maybe the most appropriate. Chen *et al.* proposed a synthesized approach to incorporate prior information on OS into a joint test statistic to assess immature OS data from RCTs, as well as, to validate the surrogacy relationship between OS and PFS in a trial [2011]. Bayesian methods described in Chapter 3 section 3.2 also offer a framework for including observational data as prior beliefs into an evidence synthesis of treatment effect. Chapter 6 illustrates this methodology using long-term claims data to inform a meta-analysis of mortality risks using the example of percutaneous vertebroplasty to treat osteoporotic VCFs.

2.2.3.3 Subgroup analyses

Drugs that have been ‘fast tracked’ in the recent past have all addressed an unmet medical need such as treatments for orphan indications or serious and life-threatening diseases. However, in order to obtain accelerated approval, manufacturers may seek restricted indications for a narrower patient population based on Phase II trial enrolment or on a subgroup analysis revealing the most encouraging benefit-risk balance. In fact, Eichler *et al.* recommend circumscribing the treatment-eligible population for drugs under adaptive licensing, as well as prohibiting off-label use, to ensure only patients willing to accept the increased risk and greater uncertainty are targeted [2012].

Nonetheless, defining the most suitable therapeutic indication for a new product should be based on sufficient evidence of efficacy and safety even under accelerated approval conditions. Using the case of drotrecogin alfa (activated) for the treatment of severe sepsis, Poole *et al.* exemplify the dangers of licensing a drug based solely on subgroup analyses [2009]. In 2001, based on the results from the same Phase III trial, the FDA approved drotrecogin alfa in patients with a high risk of death whilst the EMA granted its market authorisation under ‘exceptional circumstances’ to patients with multiple organ failure [Poole 2009]. More than one confirmatory study was requested and Poole *et al.* criticise the pathways taken by the regulatory agencies to approve drotrecogin alfa especially in light of the confusing and sometimes conflicting findings from the post-approval studies compared to earlier results for both subgroups initially identified. Assmann *et al.* also warn of the misuse of subgroup analyses suggesting these are often prone to over-interpretation, in particular, if they are not pre-specified in the trial’s statistical analysis plan [2000]. Poole *et al.* further caution that subgroups analyses and subsequent trial designs are predominantly driven by manufacturers’ interests and should

not become a ‘loop-hole’ for financial gain rather than a legitimate restriction of patient access [2009]. Using the example of ticagrelor in acute coronary syndromes, Chapter 4 looks at the impact of selecting two different patient subgroups in the context of HTA.

2.2.3.4 *Surrogate outcomes*

Although it may not be explicitly cited in the current legislation, conditional approval based on surrogate endpoints or well-accepted biomarkers may also be granted by agencies on a case-by-case basis. Miyamoto *et al.* state that: “*the use of surrogate endpoints to achieve drug approval is a pressing issue in more than one continent*” [2011]. A surrogate endpoint is a marker or measure of effect thought to be a valid predictor of clinical benefit—e.g. cholesterol level or blood pressure to predict heart disease—but is not itself a real clinical endpoint. A correlation between an intermediate and final outcome is not always sufficient to guarantee a surrogate endpoint adequately captures the effect of the treatment on a patient- or payer- relevant outcome. In oncology, PFS is considered “*reasonably likely*” to predict clinical benefit and is increasingly used as a surrogate of OS in Phase III confirmatory trials [Chen 2011]. Similarly, HIV viral load has also been widely accepted as a surrogate measure of HIV disease progression and death [Carroll 2008]. Since 1992, the FDA has granted 97 accelerated approvals for New Drug Applications and Biologic License Applications based on surrogate outcomes, 47 of these were cancer drugs based on tumour load or PFS, and 29 HIV therapies based on viral load or CD4 count¹ [FDA 2015, Miyamoto 2011].

However, Cortazar *et al.* warn that interim analyses of PFS could be misleading as they may result: “*in a trial being stopped before accrual is complete, provide an overestimate of the treatment effect, or be underpowered to detect a survival difference*” [2012: p1711]. Davis *et al.* describe a flagship example of these pitfalls with the accelerated approval of gefitinib (Iressa®) for the third-line treatment of non-small-cell lung cancer (NSCLC) [2011]. In 2003, gefitinib was approved based mainly on tumour response rates and the promise of conducting a post-marketing confirmatory RCT in 1,700 third-line patients with survival as the primary clinical endpoint [Davis 2011]. However, in December 2004, AstraZeneca reported that their trial showed no survival benefit for gefitinib compared to placebo in their indicated population. The FDA did not withdraw Iressa® from the market

¹CD4 count is a laboratory test to measure the number of CD4 T lymphocytes in a sample of blood and is an indicator of a HIV patient’s immune system’s strength.

but restricted its distribution to patients who had benefited from the drug or who were enrolled in a clinical trial; gefitinib was never approved in Europe for NSCLC [Davis 2011]. Although PFS may be an acceptable endpoint for early assessment, OS should always be measured as part of a full submission to support a drug's listing and ensure no survival decrement in the long-term [Cortazar 2012, Wonder 2012].

2.2.4 Discussion

Evidence from the literature was largely case-specific relating to either particular disease areas, such as oncology or HIV, or specific medicines. I chose to expand on four issues of early drug evaluations; but other topics were discussed by authors included in the review. The interaction between conditional approvals and HTA was a major discussion point, in particular the key role of early dialogue across agencies and jurisdictions [Henshall 2011, Eichler 2012, de Jong 2013, Husereau 2014, Kaaniche 2015]. Retèl *et al.* highlighted that HTA has a tendency to take a 'ceteris paribus' approach to drug evaluation; but given the potential new remit of HTA earlier in the product development cycle, assessors will have to take into account both changing parameters and environment [2008]. Since 2006, 17 medicines have been approved by the EMA under the CMA programme; however, Kaaniche *et al.* point out that only 8 of these medicines have been assessed by NICE and only 1 received a positive recommendation [2015]. In Europe, building on parallel and joint scientific advice processes has been promoted as a solution to coordinate the design of pre- and post-market evaluations [Henshall 2011].

The review summarised some of the evidentiary trade-offs required when assessing new pharmaceuticals under accelerated or conditional regulatory pathways. The guiding principle underlying these new pathways is that the benefits to patients' health of immediate availability outweigh the risks of collecting additional data [Kaaniche 2015]. Eichler *et al.* emphasise that the success of any adaptive licensing pathway depends on the willingness across all stakeholders involved—patients, practitioners, regulators, and payers—to accept a greater level of uncertainty [2012]. However, greater risk tolerance and decision uncertainty surrounding the licensing of new drugs doesn't necessarily imply a compromised assessment of value; Eichler *et al.* state:

“Greater willingness by patients, practitioners, and regulators to accept uncertainty is not to be equated with lack of scientific or methodological rigor. For example, an open-label, noninferiority study with soft end points

may be no more convincing under AL [adaptive licensing] than it would under the conventional licensing paradigm, whereas an increased nominal level of statistical significance or use of an unvalidated surrogate marker might be acceptable in some circumstances.” [2012: p428]

Moreover, a number of statistical approaches and methodological developments to address some of the issues aforementioned have been identified. For example, NMA maximises the use of data available by synthesising direct and indirect evidence to estimate the relative effectiveness of treatments compared, c.f. Chapter 3 [Hoaglin 2011]. The use of indirect and mixed treatment comparisons (I/MTC) has grown rapidly and is now recommended by several HTA agencies including NICE, particularly if it adds information not obtained from head-to-head comparisons, see section 2.3.3.1.4 [Tolley 2010, Schneeweiss 2011, NICE 2013a]. Beyond the basic approach of I/MTC, methodologies have been developed to extend the scope of NMA to address other issues such as bias adjustment and small numbers [Dias 2010, Siebert 2011]. Similarly, bias adjustment, hierarchical and multi-level models have been used in meta-analysis to correct for study characteristics and account for within- and between-study variance. As will be discussed below in section 2.3.3.1.5, in the instance of dealing with short-term follow-up or truncated data, extrapolation is now commonly used to model both health benefits and economic outcomes [Eichler 2012, 2015, Tolley 2010]. However, limited data inevitably leads to greater uncertainty in predicted outcomes. Sensitivity and scenario analyses can be used in HTA to assess a range of results, but these do not provide an aggregate value of effectiveness or cost-effectiveness. Model averaging has been used previously in biology and ecology to account for model uncertainty; however, its application to health economics is limited [Wintle 2003].

2.3 Review of pharmacoeconomic and HTA guidelines

2.3.1 Objectives

At present, little guidance is available on methods for the early assessment of health interventions and lessons from new MEA schemes remain to be learned. Therefore, by exploring how well-established HTA agencies describe and prescribe methods for clinical data collection and data synthesis, I attempt to determine whether a methodological gap exists between current HTA practice and how HTA may need to adapt to meet the new evidentiary challenges of early assessment. In this section, I aim to identify and critically

compare international HTA methods guidelines to better understand the differences in evidence requirements and methods recommended by international HTA agencies.

Recent reviews of HTA guidelines show that many HTA agencies have embraced recent methodological developments in comparative effectiveness research and have shown willingness to adopt more complex HTA methods [HLPF 2008, Kleijnen 2012]. However, discrepancies in the REA of health interventions remain across countries. Exploring the sources of international HTA heterogeneity can help standardise assessment processes, provide a starting point for extending early dialogue with regulators, stimulate methods research, and ensure optimal and efficient health care decision-making even at an earlier date.

The specific section objectives are:

- i. to identify all relevant HTA and pharmacoeconomic guidelines; and
- ii. to examine the current evidence requirements and methods recommended for HTA.

2.3.2 Methods

A comprehensive literature and website review was conducted in October 2011. The systematic literature search was performed in Medline®, Medline® In-Process, EMBASE, and the Cochrane Library. Search strategies used free-text and indexed terms such as “Technology Assessment, Biomedical” and restricted the number of hits by date (post 2006) and by excluding non-relevant publication types (e.g. clinical trials, case reports, and editorial/letters). The search strategies performed on the OVID platform and the Cochrane Library are presented in Table A1 and A2 in the Appendix. Country-specific website searches were used to identify publicly available methodological guidelines, as well as submission templates and technical support documents. Website searches were extended to national and sub-national HTA agencies (i.e. provincial) and recognised HTA organisations within a country. The International Society for Pharmaceutical Outcomes Research (ISPOR), the International Network of Agencies for Health Technology Assessment (INAHTA) and the HTAi websites were searched for additional material, including relevant conference abstracts and released presentations from the latest

conferences². A hand-search of reference lists from key papers and guidelines was also carried out to identify any additional citations.

HTA and/or pharmacoeconomic guidelines for the assessment of pharmaceuticals published by HTA agencies and/or HTA organisations were included for review alongside any supplementary or technical guidance available on their websites. Clinical practice guidelines and documentation regarding solely the assessment of diagnostic, medical, or surgical interventions were excluded. Information on scope and comparator(s), data collection methods (for RCTs and non-randomised evidence), sources of clinical evidence (i.e. efficacy, effectiveness and safety), and qualitative and quantitative evidence synthesis was captured in a data extraction form (c.f. Table A3 in the Appendix). In particular, guidance on the conduct and reporting of (network) meta-analysis was extracted from included guidelines and supporting documentation.

2.3.3 Search results

Out of 1,777 search hits, the literature review only included 7 full-text publications, 1 HTA methods manual and 6 pharmacoeconomic guidelines. Web-based searching was found to be a more suitable method to identify relevant literature. In total, 18 HTA guidelines, 2 formulary submission guidelines, 29 pharmacoeconomic guidelines and 5 clinical evaluation guidelines were identified. Guidelines were published by HTA agencies or HTA organisations, including commissioned expert groups, in 38 different countries across the world.

To restrict the workload and avoid repetition, a detailed assessment of the HTA guidelines published by six HTA agencies was carried out³. The six agencies were selected because they were considered the most well-established and offered the most comprehensive coverage of HTA methodology and ‘best practice’ worldwide: the Canadian Agency for Drugs and Technologies in Health (CADTH, Canada), the Zorginstituut Nederland (ZIN formerly CVZ⁴, Netherlands), the Haute Autorité de santé (HAS formerly ANAES⁵, France), the National Institute for Health and Care Excellence (NICE, England and

²HTAi 8th Annual Meeting, Rio de Janeiro, Brazil (25th-29th June 2011); ISPOR 3rd Latin America Conference, Mexico City, Mexico (8th-10th September 2011); ISPOR 14th Annual European Congress, Madrid, Spain (5th-8th November 2011).

³A database of all the licensing, reimbursement and HTA bodies worldwide was compiled to keep track of all the guidelines published internationally, c.f. CDA1 in the Appendix.

⁴College voor zorgverzekeringen/Dutch Health Care Insurance Board

⁵Agence nationale d'accréditation et d'évaluation en santé

Wales), Pharmaceutical Benefits Advisory Committee (PBAC, Australia) and the Swedish Council on Health Technology Assessment (SBU, Sweden).

Table 2 provides an overview of the guidelines and supplemental material included for review by CADTH, ZIN (CVZ), HAS, NICE, PBAC, and SBU. The majority of the guidelines were published or translated in English including the *Guidelines for Pharmacoeconomic Research in the Netherlands* by ZIN [2006], *General method for assessing health technologies* by HAS [2007], and the *General guidelines for economic evaluations* published by the Swedish Pharmaceutical Benefits Board (TLV⁶ formerly LFN⁷) [2003]. Guidelines vary widely in scope and size; for example the ZIN HTA directives are intentionally succinct but supported by external references, whilst the Australian guidelines provide the most detailed account of processes and methodologies required for submission [2008]. Additional technical support and help guides are provided by CADTH, HAS and NICE.

I report on five criteria for HTA that echo the issues previously identified by the literature review on early drug evaluations: i) population and subgroups, ii) relevant comparator(s), iii) clinical data sources, iv) methods of data synthesis, and v) methods of data translation for economic evaluations.

⁶Tandvårds-och läkemedelsförmånsverket (Dental and Pharmaceutical Benefits Agency)

⁷Läkemedelsförmånsnämnden (Pharmaceutical Benefits Board)

Table 2 Overview of HTA guidelines identified for review in Australia, Canada, France, the Netherlands, Sweden, and in the UK⁸

	CADTH	ZIN (CVZ)	HAS	NICE	PBAC	SBU/TLV⁹
Country	Canada	Netherlands	France	England and Wales	Australia	Sweden
Guidelines	Guidelines for Authors of CADTH Health Technology Assessment Reports [2003] HTA Guidelines for the Economic Evaluation of Health Technologies Canada [2006]	Guidelines for Pharmacoeconomic Research in the Netherlands (updated version) [2006]	General method for assessing health technologies [2007] Economic evaluation at the Haute Autorité de Santé : Principles and methods ¹⁰ [2010]	Guide to the methods of technology appraisal [2008a]	Guidelines for preparing submissions to the PBAC [2008]	Evaluation of methods in health services - A handbook ¹¹ [2010] General guidelines for economic evaluations from the Pharmaceutical Benefits Board [2003]
Supplemental material	Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis [2009] USER GUIDE - Indirect Treatment Comparison [2009]		Guidance on literature analysis and grading recommendations ¹² [2000]	Decision Support Unit Technical Support Documents		

⁸Limited to NHS England and Wales.

⁹Guidance provided by TLV was also included if provided additional information

¹⁰Translated title, original title: “L’évaluation économique à la Haute Autorité de Santé Principes et méthodes”

¹¹Translated title, original title: “Utvärdering av metoder i hälso-och sjukvården. En handbok”

¹²Translated title, original title: “Guide d’analyse de la littérature et gradation des recommandations”

2.3.3.1.1 Population and subgroups

HTA agencies generally consider the target population for assessment as that defined by the therapeutic indications specified in the product registration or marketing authorisation, as well as the indications sought for reimbursement. CADTH, PBAC, and TLV describe the proposed indication for HTA as all or the largest proportion of patients treated with the drug, i.e. main indication. HAS suggests that the entire population covered by an intervention, either directly or indirectly, should be considered; whilst the SBU highlights the need for clear inclusion and exclusion criteria when determining the target population. However, some restrictions of indication are acknowledged such as circumstances of use and patient characteristics. All agencies, except ZIN, strongly recommend the use of subgroup analyses to investigate any variability in the target population. However, echoing reservations about subgroup selection highlighted in section 2.2.3.3, assessors clearly stated a preference for a stratified analysis of pre-specified subgroups. Only NICE [2008a] and PBAC [2008] acknowledge the use of IPD for the estimation of subgroup-specific parameters.

2.3.3.1.2 Relevant comparator(s)

In accordance with findings from the European network for HTA (EUnetHTA), a similar description of the comparator criteria is used by all HTA agencies [Kleijen 2012]; however, the selection of comparators across countries may differ due to contextual factors such as local clinical practice. Often referred to as ‘usual’ or ‘routine’ care, the comparator of choice is defined as the most common or most widely used treatment in clinical practice for the condition in that jurisdiction. Alternative wording includes therapies that prescribers would most likely replace with the proposed drug [PBAC 2008] or recommended therapies by experts at the time of the evaluation [HAS 2007]. All six agencies stipulated that relevant comparator(s) could be licenced or unlicensed therapies, including off-label prescriptions, if these were used in clinical practice. These could also be medicinal, non-medicinal, or ‘do-nothing’—i.e. best supportive care. However, PBAC prefers the use of a pharmacological comparator when assessing medicines and recommends standard medical management only if no drugs are currently licensed for the proposed indication. CADTH is the only agency to further require in their economic evaluation guideline that the lowest cost available alternative that is often used for the same indication also be considered for comparison.

2.3.3.1.3 Clinical data sources

CADTH, HAS, NICE, PBAC, and SBU consider systematic literature reviews in line with a pre-specified scope and protocol to identify randomised and non-randomised clinical evidence. The use of valid and replicable methods including the reporting of search strategies and restrictions, inclusion/exclusion criteria, and the numbers of abstracts/publications identified is discussed in all guidelines except that published by ZIN. CADTH, PBAC, and SBU use search and selection criteria in line with the PICO—patient(s), intervention(s), comparator(s), and outcome(s)—method; and CADTH, HAS, and SBU recognise the use of a ‘pre-analysis’, expressly preliminary searching, to optimise the retrieval of clinical evidence.

Hierarchies of evidence are discussed in all guidelines to a varying extent, most agencies expressed a ‘top-down’ approach to clinical evidence with a strong preference for systematic reviews of high quality (double-blind) RCTs. On the basis of the commonly cited GRADE evidence grading system [Guyatt 2008], uncontrolled and non-randomised evidence was often cited as a ‘second-best’ data source and to be considered only if RCT(s) were not available. NICE’s language on levels of evidence is more elusive stating that: *“in the absence of valid RCT evidence, evidence from studies least open to bias will be considered preferentially with reference to the inherent limitations of the specific design”* [2008a]. Moreover, PBAC [2008] and the Academy of Managed Care Pharmacy (AMCP) [2010] in the USA explicitly acknowledged the value of observational data in the assessment of comparative harms as non-RCTs can provide better ‘safety signals’ in the real-world. The AMCP, as well as Polish guidelines [2009], also state the need for a distinct effectiveness data collection, recommending the review of pragmatic trials, patient registries, and observational studies and databases.

The sources of information most widely sanctioned are bibliographic databases such as EMBASE, Medline®, PubMed, and the Cochrane Library. Searching trial registries is reported by CADTH, PBAC, and SBU; and reviewing regulatory texts and market authorisation submissions is suggested by HAS and PBAC. PBAC and SBU also consider a hand-search of reference-list as part of the literature review as a complementary data collection method. All HTA agencies except SBU include unpublished data in their evaluations, but the TLV which also performs rapid assessments in Sweden for reimbursement decisions does accept commercial-in-confidence evidence.

A quality appraisal of studies was required by all agencies, but only NICE and SBU recognise that the validity of the review is increased if at least two reviewers screen and appraise search results. However, to standardise HTA reporting, data extraction template tables were provided by all agencies as an appendix to their guidelines, except for ZIN.

2.3.3.1.4 Methods of data synthesis

Meta-analysis is mentioned by all agencies; however, ZIN and HAS only acknowledge the use of quantitative data synthesis as a potentially useful tool, whilst PBAC provides a comprehensive description and specific methodological guidance for the meta-analysis of different clinical outcomes, and NICE via its Decision Support Unit (DSU) provides additional technical support for HTA authors on general linear modelling for pair-wise meta-analysis, which can be extended naturally to an NMA setting. When conducting and reporting a meta-analysis, the following key features are highlighted in the NICE and PBAC guidelines: fixed/random effects model, heterogeneity, meta-regression to identify potential treatment effect modifiers, publication bias and sensitivity analysis to explore the impact of excluding trials, and consistency of evidence. CADTH and SBU comment on the assessment of heterogeneity and the use of forest plots to illustrate results, but unlike NICE and PBAC, they provide no guidance as to how these should be performed. PBAC also discusses differences in pooling dichotomous, continuous, ordinal/categorical, and time-to-event data, the inclusion of cross-over trials in a meta-analysis, and alternative statistical methods for combining data. However, only the NICE DSU technical support document describes how models should be compared and goodness of fit assessed.

If there are no head-to-head RCTs comparing the intervention and relevant comparators, CADTH, HAS, NICE, and PBAC support the use of ITC in their guidelines to estimate relative effectiveness. However, based on interviews with HTA agencies, Kleijnen *et al.* reports that all jurisdictions in Europe, Australia, Canada, and New Zealand—except for Turkey—may use indirect comparisons in the case when no direct comparisons are available in a rapid assessment [2012]. PBAC provides a detailed account of ITC methods, and CADTH and NICE supplement their HTA guidelines with separate ITC guidance. On the other hand, HAS addresses the use of ITC as an appendix to their economic evaluation guidelines whilst ZIN and SBU do not discuss indirect evidence.

Different methods are presented in the guidelines including adjusted indirect comparisons using the Bucher *et al.* approach, MTC and/or NMA [1997]. NICE and PBAC consider the need to broaden search criteria to capture all relevant RCTs for ITC, in particular, PBAC stresses the importance of identifying all trials with a common comparator. Generally, a network or ‘master list’ of studies is presented and statistical support for underlying assumptions such as consistency of evidence is required. The assessment of heterogeneity, potential bias, and sensitivity analyses are discussed by CADTH, NICE, and PBAC. PBAC guidelines also provide detailed descriptions on how to extract data and present results for different clinical outcomes [2008]. Data analysis and synthesis for adverse effects is only discussed by CADTH and PBAC; but PBAC is the only HTA agency to further encourage an extended assessment of comparative harms beyond the direct randomised trials.

Indirect comparisons are not explicitly listed in the hierarchy of evidence reported by HTA agencies; except for PBAC which states that the second step in the absence of direct evidence is to present an ITC based on two or more sets of randomised trials. Controversially, the third step in the hierarchy is to present “*a comparison across non-randomised studies, including comparisons across single arms extracted from randomised trials that do not involve a common reference arm*” [PBAC 2008]. This approach is not supported by NICE guidelines that emphasise that trial randomisation must be preserved when pooling indirect or mixed evidence. If no valid randomised evidence is available, NICE and SBU both suggest a formal or informal qualitative synthesis of the data including the critical appraisal of individual studies—low and medium quality—and the tabular presentation of their results.

2.3.3.1.5 Methods of data translation for economic evaluations

All agencies support the use of cost-effectiveness analysis in HTA, although this was not a requirement set out by the HAS guidelines at the time for the reimbursement of pharmaceuticals in France. Clinical model inputs are informed by identified RCTs and/or quantitative analysis in all six countries. Methods of translating clinical data for economic modelling purposes are context-driven and mentioned by CADTH, ZIN (CVZ), HAS, and PBAC. The most common translational issue raised by agencies is to ensure efficacy measures adequately reflect the effectiveness of treatment in current practice in the country, i.e. taking into account ‘real-world’ factors such as patients’ characteristics and adherence. These contextual issues are flagged by a number of agencies; however, only

PBAC requires a pre-modelling study be undertaken by HTA authors to address the applicability and extrapolation issues arising from the parameterisation of clinical inputs.

Another consideration for the translation of efficacy measures for cost-effectiveness modelling was the surrogate criteria. Despite recognising that pharmacoeconomic results should be expressed in terms of final outcomes, many HTA agencies made allowances for the use of surrogate and intermediary endpoints in the REA of health technologies. Thus, extrapolation was commonly recommended not only to model health benefits beyond the time horizon of the trial(s) but also to predict final endpoints from surrogate measures [KCE 2008, Avksentieva 2010, Chaikledkaew 2014]. When discussing surrogate measures, several agencies only endorsed their use if their validity had previously been demonstrated or if sufficient explanation was provided to justify the robustness of a predictive relationship to final outcomes of interest [AOTM/AHTAPol 2009, AMCP 2010, PHARMAC 2015].

PBAC [2008] and NICE [2008a] provided detailed guidance on how to extrapolate short-term follow-up data such as evaluating different scenarios to test assumptions, using observational data to inform the expected impact of an intervention in the long-run, exploring alternative methods of extrapolation and conducting sensitivity analysis.

It should also be noted that the inclusion of adverse events in the economic evaluation is only discussed by CADTH, ZIN (CVZ), and SBU; whilst CADTH, NICE and PBAC guidelines report on how to quantify baseline risk of events for the modelled population of interest. Table 3 summarises the main HTA components covered by HTA and pharmacoeconomic guidelines in Australia, Canada, France, the Netherlands, Sweden, and in the UK¹³; and highlights key differences and gaps in guidance provided.

¹³Limited to NHS England and Wales.

Table 3 Summary of clinical evidence and methods in HTA guidelines in Australia, Canada, France, the Netherlands, Sweden, and the UK¹²

	Where the following HTA components clearly defined and appropriately discussed by authors?	CADTH HTA guidelines [2003]	CADTH EE guidelines [2006]	CVZ/ZIN EE guideline [2006]	HAS HTA guideline [2007]	HAS EE guideline [2010]	NICE Methods of TA [2008]	PBAC HTA guidelines [2008a]	SBU HTA guideline [2010]	TLV EE guideline [2003]
Scope of assessment	Research question(s) and target population	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Relevant subgroup(s)	✓	✓			✓	✓	✓	✓	✓
	Relevant comparator(s)	✓	✓	✓		✓	✓	✓	✓	✓
Evidence identification	Systematic literature review	✓	✓		✓	✓	✓	✓	✓	
	Additional clinical data collection		✓	✓		✓	✓	✓	✓	
	Safety data collection	✓	✓					✓	✓	
	Hierarchy of evidence	✓	✓	✓	✓	✓	✓	✓	✓	
	Quality assessment of evidence	✓	✓		✓	✓	✓	✓	✓	
Evidence synthesis	Meta-analysis	✓	✓			✓	✓	✓	✓	
	Network meta-analysis (I/MTC)	✓	✓			✓	✓	✓		
	Qualitative data synthesis						✓	✓	✓	
	Limitations: heterogeneity, bias, consistency, etc.	✓	✓			✓	✓	✓		
Modelling	Translational issues	✓	✓	✓		✓	✓	✓		✓
	Assessment of baseline risk		✓				✓	✓		
	Clinical effectiveness model inputs	✓	✓	✓		✓	✓	✓	✓	✓
	Safety model inputs	✓	✓	✓				✓		

■ Gaps in guidelines; EE: economic evaluation

2.3.3.2 Discussion

This review distinguishes between contextual, evidential, and methodological differences in HTA practice across agencies. Variation in evidence requirements and methods recommended in the guidelines can result from differences in the mandate of HTA agencies and the structure of the assessment exercise or from differences in methodological approaches used. Kleijnen *et al.* highlight a number of these systemic

differences such as reimbursement criteria, type of assessment (i.e. single/rapid or full), initiation of the assessment, and purpose of the assessment [2012].

Clinical evidence requirements for REA are overwhelmingly aligned across selected HTA agencies, particularly with regards to the hierarchy of research evidence, with a clear preference for adequately measured and synthesised effectiveness data. The use of systematic literature review and meta-analysis for identifying and pooling treatment effects is now widely accepted and documented in HTA guidelines. Network meta-analysis is also discussed by four out of the six HTA agencies compared and several other agencies worldwide. The importance of acknowledging translational issues in clinical input data in cost-effectiveness models is also considered by most agencies.

An overview of HTA guidelines published by agencies in the 32 countries not selected for in-depth data extraction reveals global similarities in the definition of target population and comparators with detailed selection rules also provided in Colombia, Israel, Poland, and Russia. The whole population of interest is commonly defined by a product's indication but no clear criteria are set out by HTA agencies to identify relevant subgroups or to make use of IPD. In addition, HTA agencies in Poland, Russia, and Thailand specifically request comparators such as the “*most effective*” or the “*cheapest drug*” [Teerawattananon 2008, Avksentieva 2010]. There appears to be a wide use of systematic literature reviews to identify clinical evidence and HTA agencies in Ireland (HIQA) and New Zealand (PHARMAC) have published their own guides on clinical information retrieval. In addition, meta-analysis is a well-accepted methodology among both emerging and established HTA agencies.

One important evidentiary limitation of current HTA guidelines is the restricted use of indirect evidence. Despite recent work demonstrating the potential value of mixed or multiple treatment comparisons [Ades 2006, Griffin 2006, Cooper 2011], the majority of HTA agencies consider ITC and NMA as useful tools only when no head-to-head comparisons are available. Very few agencies discuss methods and potential issues associated with NMA; the use of indirect evidence is also often limited to the context of pharmacoeconomic evaluations, such as in the French HAS guidelines, and in countries like Brazil, Belgium, Colombia, Germany, Finland, and South Africa. Moreover, established agencies NICE and PBAC disagree on the appropriateness of comparisons involving single arms extracted from randomised trials that do not involve a common comparator. PBAC recommends, as a third step in the hierarchy of evidence when direct

and indirect comparisons are not possible, comparing across non-randomised studies, including comparisons across single arms extracted from randomised trials [2008]. However, NICE finds this methodological approach “*not acceptable*” and suggests this data should be treated as observational and appropriate steps taken to adjust for possible bias and increased uncertainty [2008a].

Another limitation is the lack of guidance provided on the most adequate data collection and synthesis methods required to assess relative harms. The safety profile of new pharmaceuticals is a key feature of REA that is not well captured by current HTA guidelines. CADTH and PBAC recognise the need for an assessment of comparative harms based on wider sources of evidence than pre-marketing clinical trials; but only PBAC encourages the use of different search techniques to identify pharmacovigilance studies and the reporting of cases, as well as, the pooling of adverse event results using random effects models [2008]. Amongst other countries, only the Polish Agency for HTA (AHTAPol) discusses the most appropriate sources of evidence for safety analysis within HTA including case series, patient registers, and periodic reports collected by pharmaceutical manufacturers and regulatory agencies [AOTM/AHTAPol 2009]. More guidance on the identification and synthesis of harms including any necessary adjustments to using non-RCT and sparse data, especially in NMA, is required [Warren 2012]. Recommendations as to the inclusion of adverse events for economic modelling purposes are provided by CADTH, CVZ (now ZIN) and PBAC; whilst key guidance on how to account for baseline risks is given by CADTH, NICE and PBAC. Such recommendations are crucial to ensure REA is appropriately tailored to population-based cost-effectiveness analysis [Welton 2012].

This review demonstrates that similarities in REA methodologies in the context of HTA are greater than differences across selected countries. A number of methodological discrepancies remain with regards to new evidence synthesis methods and the identification, synthesis, and inclusion of observational and safety data in HTA. Highlighted gaps in current HTA guidelines should be addressed to remove unnecessary methodological differences across jurisdictions and provide a core set of evidentiary standards.

2.4 Discussion

The two reviews presented in this chapter provide two different pictures of the evidential and methodological challenges facing regulators and HTA agencies worldwide. The advent of early and accelerated drug evaluations, particularly in Europe and the USA, has given rise to a number of concerns. Most notably, the scientific community appears to be ‘nervous’ as to the risks involved in trading early access for less evidence. Davis *et al.* [2011] and Poole *et al.* [2009] warn of the potential for gaming and risky decision-making, suggesting new expedited regulatory pathways are actually designed in favour of manufacturers rather than patients. However, on the other hand, proponents of adaptive licensing like Eichler *et al.* do not find that less evidence necessarily equates to worse evidence [2012, 2015].

The literature review also identified a number of methodological developments that could address some of the evidentiary issues associated with immature and incomplete trial data. As pointed out in the international HTA guidelines review, some of these methods are already recommended as ‘best practice’ by HTA agencies when RCT data is unavailable or insufficient, including the use of surrogate outcomes, NMA and extrapolation.

However, surveying the current HTA landscape also highlighted a number of methodological gaps that may widen as assessors face new evidential challenges in the context of ‘early’ HTA. It also showed the differences across jurisdictions in the uptake of new methods, for example NMA, and the pronounced penchant by HTA agencies to continue to rely on high quality randomised evidence to assess new technologies

For the second part of my thesis, I considered three challenges to present and future HTA practice: i) the selection of relevant patient subgroups and comparators for appraisal, ii) the use of specific search strategies to identify indirect evidence for NMA; and iii) bias adjustment techniques to include observational data in evidence synthesis. I chose three examples to explore these issues and their impact on HTA outcomes and these are presented in Chapter 4, 5, and 6, respectively.

2.4.1 Caveats and limitations

Unfortunately, the literature review of early drug evaluations did not identify as many relevant publications as initially hoped. This could be due in part to the search techniques used and information sources considered, but also to the evolving nature of the evidence

base. For example, the EMA's adaptive pathways pilot project was only launched in 2014 and is due to report later this year.

In addition, I did not update the international review of HTA and pharmacoeconomic guidelines since it was conducted in 2011. I have informally tracked any new documentation made available by CADTH, ZIN, HAS, PBAC, SBU, and NICE; but I am only aware of one updated guide to the methods of technology appraisals published by NICE in 2013 [2013a]. Further work should not only focus on HTA guidelines but actually examine if and how recent appraisals have assessed early clinical evidence. Such research could provide valuable insight on what evidential issues and methodological challenges HTA agencies presently face, as well as how adaptive HTA practice has been in recent years, irrespective of recommended 'best practice'.

Chapter 3

Methods in HTA

3.1 Background

As described in Chapter 1, HTA is a multidisciplinary process that requires the use of multiple research methods from a wide range of fields including epidemiology, medical statistics, and health economics. As White *et al.* fittingly point out:

“The ability of [HTA] to answer questions about the effectiveness and cost-effectiveness of new technologies relies on the availability of appropriate methodologies including statistics.” [2000: p(iii)]

This chapter provides a brief introduction to Bayesian methods and presents three statistical and modelling techniques commonly used in HTA and referred to throughout this thesis:

- i. meta-analysis;
- ii. network meta-analysis (NMA);
- iii. and economic evaluation.

3.2 Bayesian methods

Bayesian methods have evolved from the Bayes’ Rule or Theorem—first published by its eponym in 1763—which is a mathematical equation for computing conditional probabilities [Spiegelhalter 1999]. This equation formulates how the prior plausibility of a hypothesis is taken into account, expressed as a probability distribution, and modified by new information such as evidence from a study. Under a traditional ‘Frequentist’ statistical framework, prior knowledge from other studies may be informally used in the design of a trial; in a Bayesian framework this knowledge or beliefs are formally specified in the prior distribution. For example and using the notation by Spiegelhalter *et al.*, let θ denote some unknown parameter such as a treatment effect or event rate of interest for a new health intervention and $p(\theta)$ the probability of each possible value of θ [2000]. Then $p(\gamma|\theta)$ expresses the conditional probability of some observed evidence γ for all possible

values θ . This implies that the likelihood of γ , such as the observed results from a clinical trial, depends on θ . The posterior distribution $p(\theta|\gamma)$ combines the prior distribution $p(\theta)$ with the likelihood of γ and provides the probabilities of events for different values of θ [Spiegelhalter 2000]. The Bayes Theorem can be written as:

$$p(\theta|\gamma) \propto p(\gamma|\theta) * p(\theta) \quad (1)$$

A prior distribution can be informed from external evidence coming from a meta-analysis of previous RCTs evaluating the intervention of interest; or more controversially, it can rely on elicited expert opinion. In order to minimise subjective judgement, a very uncertain distribution—also known as a ‘vague’ or ‘non-informative’ prior—can be used to encompass all feasible values of θ and represent the lack of external evidence available [Welton 2012]. Moreover, as more observed evidence becomes available, the prior will be overpowered by the likelihood and its influence on the posterior probabilities becomes negligible.

Posterior distributions can be summarised by direct probability statements that could not otherwise be made with a conventional statistical approach. A Bayesian credible interval (e.g. 95% CrI)—analogous to a confidence interval—can be interpreted as, given the prior distribution, the model and the data, there is 95% chance that the ‘true’ value of the event of interest lies in the 95% range [Spiegelhalter 2000]. Posterior probabilities are easily interpretable particularly in the context of clinical research, i.e. the probability that an outcome for a given treatment exceeds a certain threshold or lies within a certain interval. In the context of HTA, Spiegelhalter *et al.* provide the following definition of Bayesian methods:

“The explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation, and reporting of a health technology assessment.”
[1999: p508]

Moreover, authors describe four important uses for Bayesian methods: i) designing randomised trials, ii) pooling results from published trials, iii) simultaneously handling sub-studies and estimating effects on many subgroups, and iv) applying methods to non-randomised evidence [Spiegelhalter 1999, 2000, 2004]. Bayesian methods to synthesise evidence including meta-analysis and NMA models are of particular relevance to HTA and to this thesis. In addition, a Bayesian framework allows the combination of all

available evidence, from multiple sources and different study types, for example the meta-analysis of randomised and observational data presented in Chapter 6.

In recent years, Bayesian methods have become more attractive and routine in HTA. Welton *et al.* describe them as more efficient and useful [2012], whilst Spiegelhalter *et al.* claim they are more flexible and ethical than traditional methods [1999]. Eddy *et al.* simply state:

“Bayesian methods provide an attractive approach to the assessment of health technologies because they correspond to the way we think about assessment problems intuitively.” [1990: p32]

Bayesian methods also lend themselves to making predictions and the form in which conclusions are drawn naturally input into decision making. The development of Markov Chain Monte Carlo (MCMC) simulation and the availability of the WinBUGS software¹⁴ have substantially facilitated the computational approach and fitting of very complex Bayesian models. I have presented the methods for a single parameter θ , but Bayes' theorem can extend to multi-parameter models to address more realistic healthcare questions about multiple interventions, multiple outcomes, multiple subgroups, and the meta-analysis of multiple studies [Welton 2012]. For further details about the theoretical and practical considerations of Bayesian methods for HTA, Spiegelhalter *et al.* published a comprehensive review of Bayesian methods in HTA sponsored by the NHS R&D HTA programme [2000].

3.2.1 MCMC and Gibbs sampling

MCMC methods use simulation draws for each parameter repeatedly so as to eventually sample from the posterior distribution [Gelman 1996, Welton 2012]. Gelman *et al.* describe:

“The essential idea of iterative simulation is to draw values from a random variable $[\theta]$ from a sequence of distributions that converge, as iterations continue, to the desired target distribution of $[\theta]$.” [1992: p457]

For illustrative purposes and using the notation from Welton *et al.*, imagine the values of several parameters $\theta_1, \theta_2, \dots, \theta_m$ are initially drawn as $\theta_1(1), \theta_2(1), \dots, \theta_m(1)$; for each

¹⁴Latest stable version WinBUGS 1.4.3 is freely available from:
<http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>

new iteration, values are sampled based on the values of the previous iteration [2012]. The Markov process describes a stochastic process by which future probabilities depend only on the most recent values sampled, a *Markov chain* refers to the sequence of values generated by a Markov Process. In other words, a Markov chain is a consecutive set of random draws of θ that are each slightly dependent on the previous one, such that for t iterations, the values of the parameters are $\theta_1(t), \theta_2(t), \dots, \theta_m(t)$ and for $\theta_1(t)$ the value of the next draw $\theta_1(t + 1)$ is solely dependent on the present value $\theta_1(t)$ and not on any past draws [Lam 2008, Welton 2012]. The *Monte Carlo* in MCMC denotes the simulation approach used to estimate the parameters of interest.

A number of sampling methods exist but one simple approach commonly associated with Bayesian inference and used by WinBUGS is the Gibbs sampler. Described as a “*Markovian updating scheme*” [Gelfand 1990], Gibbs sampling is an MCMC algorithm for obtaining a sequence of values which are approximated from a joint probability distribution of two or more random parameters. Each parameter is taken in turn and the Gibbs sampler draws their values one at a time from their posterior distribution, respectively, conditional on the known information and the values of all the other parameters being fixed at their present value [Welton 2012]. This sampling process is repeated for a number of iterations to ‘update’ the values for each parameter given their full conditional distributions. Gibbs sampling is considered to be a more practical method for estimating parameters of interest, especially in a HTA context due to the often hierarchical nature of many of the models used, as it is often easier to express and sample from these full conditional distributions than the joint distribution.

Initial samples may not effectively represent the ‘target distribution’ for θ , but it has been shown that over a number of iterations and regardless of the starting point, the chain of simulated values will eventually converge to estimate a ‘stationary’ joint posterior distribution. For this reason, the first draws are usually discarded as a ‘burn-in’ sample and once convergence is reached, summary measures are calculated from a large number of further iterations [Welton 2012]. As Lam explains “*this is to make [the] draws closer to the stationary [posterior] distribution and less dependent on the starting point*” [2008]. However, the number of iterations required to achieve convergence is unclear; convergence can be assessed through graphical exploration in WinBUGS. For example, history plots for simulated values can provide a visual indication of convergence. Any obvious patterns or systematic structure in the plots for a given parameter suggests slow

convergence [Welton 2012]. Density plots and trace plots can also be used to map the iteration number against the value of the draw of the parameter at each iteration and provide an indication of how well the Markov chain is moving around the parameter space, also known as ‘mixing’ [Lam 2008]. Another diagnostic for assessing convergence is calculating the autocorrelation statistics which measure the correlation between draws over a specified ‘lag’, i.e. number of iterations apart. Autocorrelation can also be plotted against different lag times; a relatively high autocorrelation across a large lag implies a high degree of correlation between draws and slow convergence. More details about MCMC performance and working with WinBUGS can be found in Welton *et al.* [2012] and the BUGS book by Lunn *et al.* [2012].

3.3 Meta-analysis

Systematic literature reviews and meta-analyses are the most widely used methodologies to assess the clinical effectiveness of health interventions; Stephens *et al.* found that over 50% of the HTA agencies they surveyed considered both methods as the “*starting point and primary methodology*” for the synthesis of evidence in HTA [2012]. Systematic reviews are a cornerstone of evidence-based medicine and frequently used to collect relevant information not only on the effectiveness of a technology, but also on adverse events, quality of life, and economic evaluations. In order to answer a specific research question, an exhaustive search of the literature is performed and all the studies meeting pre-defined eligibility criteria are appraised and summarised. The CRD’s guidance for undertaking reviews in health care provides a detailed account of the rigorous methodology that should underpin a systematic literature review including the review protocol, evidence search, study selection, data extraction, quality assessment, and data synthesis [2009]. Each step should be explicitly reported in order to ensure the transparency and reproducibility of the methods, as well as, to uphold the validity of findings.

A meta-analysis is a statistical technique for combining data from independent but similar studies [Crombie 2009]. It provides a consolidated and quantitative summary of the empirical evidence extracted from studies identified by systematic review. The benefits of meta-analysis include the ability to formally digest a “*large and often complex, sometimes apparently conflicting, body of literature*” by increasing the power of small or inconclusive studies by pooling their results [Haidich 2010, Ioannidis 1999]. Borrowing strength across studies can improve the precision of an estimated treatment effect and

allow us to detect a small, yet meaningful difference, which may not have otherwise been demonstrated by an individual study. The studies most often considered for synthesis are RCTs; but methodologies are being developed to extend the principles of meta-analysis to other study designs [White 2000].

In medical research, the basic principle of meta-analysis is that an overall treatment effect—e.g. for a new treatment B vs. a control intervention A— is estimated as a weighted average of the observed treatment effects from single RCTs for the same pairwise comparison [Higgins 2011]. The summary treatment effect size from each study may be a mean difference if the data are continuous or a ratio measure if the data are dichotomous, such as response rates or time to event outcomes [Bartolucci 2000, Deeks 2001]. Treatment effects expressed as risk ratios, odds ratios (OR), or hazard ratios (HRs) are conventionally modelled on the log scale. Further details about how to interpret and summarise different types of data and effect measures can be found in the CRD guidance and the Cochrane Handbook for Systematic Reviews of Interventions [CRD 2009, Higgins 2011].

When combining treatment effects, the weights—given to each study result in the meta-analysis—determine the contribution or influence of each trial on the overall ‘pooled’ effect. If all the weights are the same then the pooled effect is equal to the mean treatment effect from all the included trials [Higgins 2011]. However, observed effects are usually weighted according to the inverse of their variance (standard error squared). Treatment effects from larger studies will tend to have smaller variances—i.e. larger inverse variances—which give more weight to larger RCTs with bigger sample sizes in the meta-analysis [Bartolucci 2000]. Two meta-analysis models are most commonly used and are presented in this section: a fixed effect model and a random effects model.

3.3.1 Fixed effect model

Using the notation from Higgins *et al.* [2011], the weighted average is calculated as:

$$pooled\ estimate = \frac{sum\ of\ (estimate * weight)}{sum\ of\ weights} = \frac{\sum(Y_i * W_i)}{\sum(W_i)} \quad (2)$$

where Y_i is the treatment effect in the i^{th} of k studies, and W_i is the weight given to the i^{th} study with $W_i = \frac{1}{se_i^2}$ and se_i^2 is the within-study variance of the i^{th} study. The fixed effect model assumes that each treatment effect combined in the meta-analysis estimates the same underlying overall effect d :

$$Y_i \sim \text{Normal}(d, se_i^2) \quad i = 1, \dots, k \quad (3)$$

This model considers the variability in the estimated treatment effects across studies is solely due to sampling error and does not make allowances for variations between studies [CRD 2009, Welton 2012]. Within a Bayesian framework, prior beliefs are combined with the meta-analysis. The only unknown parameter in the model is the common treatment effect d and requires a prior distribution. Generally a vague or non-informative prior is specified, Welton *et al.* describe this as “a Normal distribution (centre at no effect for comparative outcomes) with large variance (relative to that of the outcome in question)” [2012]. However, if prior knowledge on the ‘true’ underlying treatment effect is available, for example from observational data, this information can be included in a Bayesian analysis, see Chapter 6.

Meta-analytic statistics are routinely summarised graphically in a forest plot (see Figure 7 in section 5.4.2 in Chapter 5). Forest plots illustrate both individual study data and overall pooled results and provide a simple representation of the evidence base [Bartolucci 2000]. Point estimates from each included study are often plotted as smaller or larger markers according to sample size and their precision (e.g. 95% CI) as horizontal lines of varying width.

3.3.2 Random effects model

The random effects model relaxes the assumption of a common underlying treatment effect across all studies; it assumes that each observed effect size from individual RCTs is estimating its own unknown underlying effect, which are assumed to come from a common population mean [Sutton 2001]. The latter assumption is thought to be more realistic as it allows for differences across studies, such as dissimilarities in patient populations, as well as potential biases due to trial design, to influence treatment effect. When the causes of variability cannot be identified and explicitly included in the analysis—e.g. as known covariates—a random effect is modelled to account for the variability in treatment effects. Each study results are weighed according to their own variance and the between-study variance [Sutton 2001]:

$$Y_i \sim \text{Normal}(\delta_i, se_i^2) \quad i = 1, \dots, k \quad (4)$$

$$\delta_i \sim \text{Normal}(d, \tau^2) \quad i = 1, \dots, k \quad (5)$$

where Y_i is the treatment effect in the i^{th} of k studies with δ_i is the unique underlying effect in the i^{th} study and se_i^2 is the within-study variance of the i^{th} study. Each underlying effect δ_i is drawn from a normal distribution with mean d and variance τ^2 [Welton 2012]. δ_i the underlying treatment effect for each study is considered to be a random sample from a population distribution of effect sizes, assuming that all studies are exchangeable [Spiegelhalter 2000]. τ^2 is the between-study variance, i.e. the heterogeneity parameter, if τ^2 equals 0, then a fixed effect model is obtained. As with the fixed effect model, a prior distribution is needed on the overall pooled effect d and a similar normal non-informative prior can be used. However, for the between-study variance parameter τ^2 , Welton *et al.* find a vague prior such as a uniform distribution on the standard deviation scale, e.g. $\tau \sim \text{Uniform}(0, 10)$, to be more suitable to cover all plausible values [2012]. Note as the CRD's guidance points out:

“Where there is little between-study variability, the within-study variance will dominate and the random-effects weighting will tend towards that of the fixed-effect weighting. If there is substantial between-study variability, this dominates the weighting factor and within-study variability contributes little to the analysis.” [2009: p55]

If there are few studies included in the analysis, alternative prior distributions can be used for τ^2 and a sensitivity analysis is often undertaken as there may then be some concern about the prior distribution having a considerable influence on the results. The advantage of the random effects model is that it takes into account both the within-study variability and the between-study heterogeneity [Sutton 2001]. Therefore, it can identify sources of heterogeneity across studies but it cannot explain these sources.

3.3.2.1 Heterogeneity

Heterogeneity simply describes the variation in observed treatment effects across studies but is an important issue in meta-analysis that should be explored as it may challenge the interpretation of the results. Statistical heterogeneity describes any underlying differences among the trials included in the meta-analysis that cannot be explained by chance alone. Some variation is to be expected due to random error; but systematic discrepancies in patients recruited (e.g. baseline disease severity, co-morbidities), interventions given (e.g. dose schedule and delivery), outcomes reported, or study design characteristics (e.g. blinding and concealment of allocation) can influence treatment effect size and direction

and lead to heterogeneity between studies [Crombie 2009, Deeks 2011]. The CRD submits the following:

“Exploring statistical heterogeneity in a meta-analysis aims to tease out the factors contributing to differences, such that sources of heterogeneity can be accounted for and taken into consideration when interpreting results and drawing conclusions.” [2009: p66]

The sources of heterogeneity, whether due to clinical or methodological diversity, can be difficult to identify but testing whether there is significant heterogeneity across the studies to cause concern is best practice. In addition, the choice between a fixed and random effects model is often based on the degree of heterogeneity in a meta-analysis.

Poor overlap between the studies' confidence intervals on a forest plot can provide a visual indication of heterogeneity, but the most widely used tests for statistical heterogeneity are Cochran's Q -test and Higgins's I^2 statistic [CRD 2009, Sedgwick 2015]. The formula for calculating the Q statistic is a function of the difference between individual study's treatment effect and the summary mean effect of all the studies combined [Bartolucci 2000, Higgins 2002]. A Q -test close to 0 suggests little or no difference between the single study and pooled treatment effects and thus indicates a small amount of heterogeneity that may be clinically unimportant [Deeks 2011]. The I^2 test attempts to assess the impact of heterogeneity on the meta-analysis by quantifying the inconsistency across studies; it compares the Q statistic to its expected value assuming homogeneity—i.e. degrees of freedom [Bartolucci 2000, Higgins 2003]. The I^2 statistic can be easily interpreted as the percentage of the variability in treatment effect sizes that is due to heterogeneity rather than chance [Deeks 2011]. Further details about how to calculate these test statistics can be found in the Cochrane Handbook for Systematic Reviews of Interventions [Higgins 2011]. Higgins *et al.* provide an insightful commentary on the handling of statistical heterogeneity in systematic reviews [2009]. The authors recommend caution when investigating heterogeneity, in particular if relying too heavily on statistical tests to diagnose heterogeneity as these tests tend to be underpowered when pooling a small number of studies, or if trying to explain the sources of heterogeneity by identifying treatment modifiers *post hoc* [Higgins 2009].

3.3.3 Extensions and limitations

Beyond a simple narrative review of the evidence, a meta-analysis provides a quantifiable summary of multiple studies. Its application has now been widely embraced in health care evaluation and is commonly used to inform HTA and clinical guidelines [Higgins 2011]. Indeed, the underlying motivation for combining data from the literature is to reach a definitive conclusion about a specific health intervention when multiple studies evaluating this intervention have been conducted [Bartolucci 2000]. However, meta-analyses also have the potential to mislead decision-makers, particularly if variation across studies and other biases are not carefully considered [Deeks 2011]. Meta-regression is a tool developed to handle the sources of heterogeneity particularly with pre-specified covariates [Thompson 2002]; more complex meta-analysis models have also been developed to combine cluster-randomised trials and crossover trials, as well as, to pool IPD from RCTs [Bartolucci 2000, Sutton 2001]. In addition, methodological research into the different types of biases has flourished, especially the problem of publication bias, providing both ways to identify and adjust for it in meta-analysis [Song 2000, Sutton 2000, Rothstein 2006]. Lastly, although examples are rare, the same meta-analytical techniques as described in 3.4.1 and 3.4.2 can be applied to utility and cost data in order to inform other parameters in an economic model [Welton 2012].

3.4 Network meta-analysis

Pairwise meta-analysis can be generalised to evaluate multiple treatment comparisons and to synthesise evidence across several studies comparing different health interventions. NMA combines both the direct and indirect evidence from randomised studies forming a connected network of evidence to produce an ‘internally coherent’ set of effect estimates for each treatment of interest relative to every other [Caldwell 2014]. When no head-to-head comparison is available, interventions can be compared ‘indirectly’ through a common comparator. For example, if a trial compares interventions A vs. C and another compares B vs. C, despite not having a direct comparison, an ‘indirect’ estimate of A vs. B can be obtained from the relative effects of AC and BC:

$$\theta_{AB}^{indirect} = \theta_{AC}^{direct} - \theta_{BC}^{direct} \quad (6)$$

where θ denotes the ‘true’ underlying treatment effect estimate and C is the common comparator (e.g. placebo or active treatment comparator) [Welton 2012]. If direct

evidence on A vs. B becomes available, both direct and indirect can be pooled into an MTC. The term NMA refers to both indirect and mixed treatment comparisons (I/MTC). A network diagram can be used to illustrate the evidence base for NMA (c.f. Figure 5 in Chapter 5); each ‘node’ represents an intervention and the lines connecting the nodes represent the RCTs comparing each pair of competing interventions [Hoaglin 2011]. When direct and indirect evidence is available for a given comparison, a ‘closed-loop’ is formed in the network. Jansen *et al.* [2011] and Sutton *et al.* [2008] describe a number of different network shapes based on various ‘paths’ between nodes and ‘anchor’ treatments; moreover, networks can vary substantially in size, as demonstrated in the NMA for apixaban to prevent venous thromboembolism presented in Chapter 5.

Both fixed and random effects models can be extended to NMA. In the same way as a meta-analysis, a fixed effect approach assumes a ‘true’ underlying effect size for each treatment comparison and any difference between estimates from included studies is attributable to chance alone. On the other hand, a random effects model allows for variation in the ‘true’ treatment effects across trials. The WinBUGS software can also be used to perform Bayesian NMAs and has the advantage of being able to provide rankings of the different treatments, as well as allowing for differences in the conduct and reporting of the studies in a NMA due to its flexibility.

3.4.1 Assumptions

A number of basic assumptions underpin NMA, namely the similarity and consistency assumptions that follow Bayesian or frequentist statistics. First, the RCTs included in the network have to be sufficiently similar to be combined in an NMA. An NMA should be based on a systematic literature review and rigorous selection criteria; however, similarly to a traditional pairwise meta-analysis, it may be challenging to determine if included trials are ‘similar enough’. Randomisation only holds within an individual RCT, thus covariates that could influence relative treatment effects should be comparable across studies or adjusted for using meta-regression [Jansen 2011].

The consistency assumption only applies to the closed loops of evidence in the network; it entails that there is no discrepancy between direct and indirect estimates for any given pairwise comparison. Using the notation above, consistency across direct and indirect evidence implies that $\theta_{AB}^{direct} = \theta_{AB}^{indirect}$. If either or both of these assumptions are violated, confounding may bias the results of the NMA and the theory of transitivity may

no longer hold. Transitivity simply states that if intervention A is better than C and C is better than B based on direct RCT results, it follows that A is better than B.

3.4.1.1 Heterogeneity and inconsistency

Variation in the distribution of treatment effect modifiers across studies can bias comparisons resulting in between-study heterogeneity and inconsistency [Jansen 2011, 2013]. In the context of NMA, heterogeneity describes systematic differences between effect estimates within the same pairwise comparison, and inconsistency refers to differences between treatment effects between direct and indirect evidence, and different routes of indirect evidence [Salanti 2014].

The similarity assumption and the presence of heterogeneity within a network of evidence can be assessed using the same diagnostic tools described in section 3.3.2.1. for pairwise meta-analysis. Initially, patient and trial characteristics should be compared for homogeneity and forest plots for each treatment pairing can also be visually inspected. The Q statistic and I^2 index are also equally applicable to NMA; whilst meta-regression and subgroup analyses can be used to identify treatment effect modifiers [Cooper 2009, Achana 2013].

Donegan *et al.* provide an exhaustive list of the different methods available to explore inconsistency in a network of evidence, including comparing outcome measurements in the referent group, node-splitting, multidimensional scaling, the back transformation and graph-theoretical methods, and two-stage approach [2013]. The authors also summarise each method and provide key references.

3.4.2 Extensions and limitations

The popularity of NMA has increased in recent years, particularly for decision-making, as it enables the simultaneous comparison of multiple competing treatments in a single statistical model [Cooper 2011]. Caldwell *et al.* note that:

“NMA has matured and models are available for all types of underlying data and summary effect measures and can be readily implemented in both frequentist and Bayesian frameworks with pre-written programmes available in widely used softwares” [2014: p1]

Within a Bayesian framework, NMA has the added advantage of being able to calculate the probabilities of each treatment within a network being the ‘best’ for a specific

outcome of interest and to rank interventions from ‘best’ to ‘worst’ [Caldwell 2014]. Unlike performing a series of pairwise meta-analyses, NMA allows for the concurrent REA of multiple treatments compared to all other options, but still provide recommendations for the use of individual health technologies. To this extent, Salanti *et al.* [2014] and Tan *et al.* [2013] have proposed new tabular and graphical formats to present the results from NMA beyond a traditional forest plot.

As pointed out in Chapter 2, NMA is now commonly acknowledged by several HTA agencies as a valid form of evidence synthesis; however, there is no consensus as to the use of indirect comparisons when head-to-head evidence is available. For example, NICE still advises that data from head-to-head RCTs should be presented in the reference-case analysis and an NMA should only be conducted alongside pairwise meta-analyses for each treatment comparisons of interest [2013a]. Jansen *et al.* argue that both direct and indirect evidence contribute to the total body of evidence and justify the use of NMA as followed:

“The results from indirect evidence combined with the direct evidence may strengthen the assessment between treatments directly evaluated. Even when the results of the direct evidence are conclusive, combining them with the results of indirect estimates in a [MTC] may yield a more refined and precise estimate of the interventions directly compared and broaden inference to the population sampled because it links and maximizes existing information within the network of treatment comparisons” [2011: p418].

Nonetheless, most HTA agencies do agree that NMA should be limited to RCTs and any naïve pooling of single treatment arms from different studies or observational data should be regarded as biased and associated with increased uncertainty [Lu 2004].

A number of statistical developments have extended the application of meta-regression methods and hierarchical modelling to NMA [Owen 2015]; the use of observational data as well as multiple and competing risk outcomes have also been explored in MTC to further optimise the evidence base available for health-care decision-making [Ades 2010, Schmitz 2013, Achana 2014].

3.5 Economic evaluation

Goodman highlights that “[the] studies of costs and related economic implications comprise a major group of methods used in HTA” [2014]. Indeed, decision analytic cost-

effectiveness modelling has become an integral part of HTA and a prevalent policy tool when deciding how best to allocate scarce resources within a healthcare system [Newmann 2005].

In essence, an economic evaluation can be defined as the comparison of alternative options in terms of their costs and consequences [Drummond 2005a, Briggs 2006]. In the context of HTA, cost-effectiveness analysis relates this comparison to evaluating alternative health technologies in terms of their health care costs (e.g. staff time, GP visits and hospitalisation, tests, and drug acquisition costs) and their health benefit (or detriment) in terms of a measured treatment effect (e.g. episode-free days, cases avoided, life years gained). In actual fact, cost-utility analysis is more frequently used whereby health ‘consequences’ are not only quantified but ‘valued’ using a more generic measure of health, such as a quality-adjusted life year (QALY). In cost-utility analysis, health consequences are adjusted by health state preference scores or utility weights, valued relative to one another, to better inform resource allocation decisions across treatments and disease areas [Drummond 2005b].

The following sections briefly describe the most common model structures in healthcare economic evaluation. Note that, I consider cost-utility analysis to be a variation to the general cost-effectiveness approach and don’t formally distinguish the two in subsequent analyses presented in this thesis.

3.5.1 Decision-analytical models

Drummond *et al.* state that:

“Decision analytical modelling provides a framework for decision-making under conditions of uncertainty” [2005b].

Decision models provide an analytical structure to evaluate alternative healthcare programmes and interventions by including and translating all relevant evidence to a specific decision problem into estimates of cost and effects [Drummond 2005b, Briggs 2006]. By applying a decision rule, analysis results can identify the ‘best’ option. Decision-analytical models also provide a flexible framework to assess uncertainty related to the economic evaluation.

Perhaps the simplest and most widely used decision model structure is the decision tree. A decision tree is schematically represented as a series of possible pathways or ‘branches’ all originating from the same starting point, also known as the decision node. The decision

node is often illustrated as a square box. The first part of the model structure presented in Figure 1 of Chapter 4 section 4.4.1.2 shows this decision node for an initial ACS event and the ensuing decision tree ‘branches’. Each branch or pathway characterises a particular event a patient may experience, such as a myocardial infarction, stroke, or death. Probabilities are assigned to each ‘branch’ to account for the likelihood of each event occurring. Moving along from left to right, the probabilities of any subsequent event will be conditional on that of the previous event [Drummond 2005b]. For example, the probability of a treatment being successful is conditional on the probability of a patient experiencing an adverse event in the first place. The combination of different ‘branches’ form mutually exclusive and exhaustive pathways a given patient might follow.

Costs and utilities are assigned to each pathway in the decision tree and by ‘rolling back’ the model, effectively multiplying the conditional probabilities for each event along the series of branches by the associated costs and expected values of effect, the cost-effectiveness of each pathway is estimated. If the likelihood of any event is treatment-dependent, a decision tree can inform the ‘optimal’ choice between two or more interventions compared.

Decision rules are often centred on the calculation of an incremental cost-effectiveness ratio (ICER), i.e. the additional cost per extra unit of health benefit (e.g. a QALY). A cost-effectiveness plane is a two-dimensional representation of costs and effects, formed of four quadrants on which the horizontal axis represents the difference in effect between two interventions and the vertical axis the difference in cost [Gray 2001]. The cost-effectiveness of a health technology can be visualised on a cost-effectiveness plane depending in which quadrant of the plane incremental benefits and costs fall on. A willingness to pay or cost-effectiveness threshold can be used as a decision rule and is illustrated on the cost-effectiveness plane by delimiting an area under an acceptable ICER, e.g. £20,000 per additional QALY [Gray 2001].

Decision trees have two important limitations. First, decisions are assumed to be modelled over an instantaneous and discrete period [Drummond 2005b]; that is a time variable is not explicitly incorporated in a decision model. However, most elements of an economic evaluation are time-dependent. For example, survival or quality of life changes as patients gets older and should be adjusted for if modelling over a lifetime horizon (i.e. until death). Similarly, the ‘best practice’ of discounting future costs and health benefits relies on appropriately modelling a time component. Second, decision-analytical models can

become very complex especially if modelling long-term complicated disease pathways, such as those of chronic diseases, where competing risks need to be taken into account as well as individualised treatment sequences. In such instances, the decision tree can grow exponentially in size and complexity to model adverse events, relapses or recurrence, remission, until eventually death.

Markov models are another model structure commonly used in economic evaluations and address some of the limitations of decision-analytical models.

3.5.2 Markov models

In lieu of ‘branches’ in a decision tree, a Markov model defines a finite set of health ‘states’ that a patient can be in at any given point in time [Gray 2001, Drummond 2005b]. Transition probabilities model the movements of patients from one state to another; and the probability of a patient remaining in a given state is also assessed over a series of ‘cycles’ or time intervals. Thus, time-dependency is built-in the Markov model structure. Patients can start in any health state, remain in or move to another other health state over time—allowing for forward and backward progression—until they reach death, also known as the absorbing state [Gray 2001]. Similarly to a decision-analytical model, each state is associated with costs and utilities; but expected costs and values are weighted by the time a patient spends in that state. The most common method to calculate the probability of a patient being in a given state at each cycle is the cohort method. If I assume a cohort of 1000 patients entering the model at time 0, for each cycle I can calculate the proportion of patients in the cohort in all the different health states modelled. For any given cycle, the proportion of patients ‘being in a state’ depends on the proportions of patients in other states in the last cycle and the transition probabilities. Running the analysis over many cycles creates a “profile” of how many patients are in each state and move between states over time [Briggs 1998], this is also known as a Markov trace. The time horizon of analysis is split into equal cycles, but cycles can range in length from a month to a year based on the nature of the disease and/or treatment modelled [Gray 2001].

Overall expected costs for the cohort of patients can then be estimated by summing the costs across all health states according to the proportion of patients in each one at each cycle, over the total number of cycles. In the same way, Briggs and Sculpher explain:

“By weighting the quality of life of the state by the length of time in the state and the number of patients from the cohort in the state, an estimate of the number of QALYs experienced by the cohort is obtained for each cycle.”

[1998: p406]

If transition probabilities, costs or utilities are treatment-dependent, competing interventions can be evaluated in different ‘arms’ of the model by comparing the overall expected costs and values for each intervention resulting from different Markov traces. Appendix B, C, and D provide the Markov traces for the cost-effectiveness models included in Chapter 4, 5, and 6, respectively.

Markov models may not always be suitable to address a decision problem, often a combination of both a decision tree and Markov process is used (c.f. example of ticagrelor in ACS described in section 4.4.1.2 in Chapter 4). In addition, one limitation of Markov models is that they are ‘memoryless’ [Drummond 2005b]. That is, the underlying ‘Markov assumption’ implies that the probability of a given transition in a Markov model is independent of earlier transitions. However, this assumption may not hold for certain diseases and can complicate the modelling of certain time-dependencies.

3.5.3 Handling heterogeneity and uncertainty

Economic models provide a flexible framework to adjust or account for different types of uncertainty and heterogeneity. As described in section 3.3.2.1, heterogeneity can be caused by systematic differences in patient baseline characteristics, such as sex, age or disease severity. In models, this inherent variability can be evaluated by running subgroup analyses or by defining model parameters as a function of other parameters [Gray 2001]. For example, transition probabilities can be conditioned on gender, age or disease severity. Moreover, a number of methods have been developed and ‘best practice’ recommendations put in place to assess and deal with uncertainty in economic evaluations [Briggs 2000].

There are several different types or sources of uncertainty associated with cost-effectiveness analysis models. Parameter uncertainty refers to the data requirements and model inputs used; whilst methodological or ‘structural’ uncertainty relates to the modelling approach and assumptions made [Gray 2001]. Additional uncertainty may be present from the use of extrapolation techniques as well as the desire to generalise results to other settings [Sculpher 2004, Drummond 2005b].

Sensitivity analysis is the simplest way to test parameter uncertainty and gauge its impact on the model results. Briggs *et al.* [2006] describe in details the different methods of sensitivity analysis, but in short, particular model variables are varied over a plausible range of parameter estimates and compared to a ‘reference’ or ‘base’ case using the ‘best estimate’ [Drummond 2005b]. Beyond this deterministic approach to representing uncertainty, probabilistic sensitivity analyses (PSA) have now become standard practice to explore the impact of joint parameter uncertainty on the results of a cost-effectiveness analysis [Gray 2001]. It allows for the assessment of uncertainty across some or all the parameters in the model at the same time. First, parameter values are sampled from a distribution, then the uncertainty is propagated through the model using simulation techniques [Drummond 2005b]. Monte Carlo simulation is most frequently employed to randomly draw from each of the input parameter distributions over a large number of iterations (e.g. 1,000 runs). The costs, benefits and thus ICERs can be averaged over all iterations to obtain probabilistic means and confidence intervals. The parameter uncertainty can also be represented graphically by plotting the simulations on the cost-effectiveness plane. Cost-effectiveness acceptability curves (CEAC) can also be used to evaluate the probability of a health intervention being cost-effective at different willingness to pay threshold values, using the PSA results to calculate the proportion of ‘cost-effective’ ICERs over many simulations [Fenwick 2001].

Scenario analyses are recommended to address methodological uncertainty [Drummond 2005b]. Methodological uncertainty exists as a result of choice of modelling methods or structural assumptions that underpin a decision model and can lead to different prediction or very different results [Edlin 2015]. For this reason, sensitivity analysis is not straightforward for structural assumptions and this type of uncertainty can only be explored by considering different model ‘scenarios’. Often a ‘best case’ scenario and ‘worst case’ scenario can be defined to test more optimistic or conservative model assumptions, respectively, and be compared to the base case.

3.5.4 Extensions and limitations

As Drummond *et al.* remark:

“[...]all models are a simplification of reality, and the ultimate objective in selecting an appropriate structure for a decision model is to make the model

no more complex than it has to be to address the policy questions appropriately” [2005b: p300]

In this sense, economic models are never perfect and only provide an abstract representation of ‘real-life’ to address a specific decision problem. In addition, decision models are very sensitive to the analysis perspective taken, i.e. that of a patient, health care system, or society as a whole. Different approaches will influence what costs and consequences are considered relevant to an economic evaluation [Barton 2014]. In my thesis, all case studies take an NHS and Personal Social Services (PPS) perspective as is recommended by NICE for HTA submission in England and Wales [NICE 2013a].

Moreover, extensive research has been undertaken in the field of cost-effectiveness and cost-utility analysis to address issues and concerns regarding specific elements of economic evaluation. There is a growing literature on sources and methods to include clinical input parameters, methods for eliciting preferences and utilities, the valuation of willingness to pay-thresholds, and how economic models can be used in value of information studies to inform future research [Spiegelhalter 2003, Drummond 2005b, Shiroiwa 2010, Thorlund 2014]. For example, Dias *et al.* considered methods to include evidence synthesis in probabilistic cost-effectiveness analysis [2013]; Ara and Brazier examined how utility values should be obtained to populate a model [2010]; and Claxton *et al.* evaluated methods for the estimation of NICE’s cost-effectiveness threshold [2013].

It should also be noted that cost-effectiveness and cost-utility analysis are the most common types of economic evaluation, but cost-minimisation or cost-benefit analysis are also used [Drummond 2005b]. In particular, cost-benefit analysis has gained attention—good and bad—for monetising health benefits and thus overtly including a patient or population’s willingness to pay for extra units of health [Robinson 1993]. Its advantage is that by converting all outcomes in a common denomination, i.e. monetary value, it facilitates the assessment of interventions across sectors, e.g. education, social services, and healthcare. However, it is often critiqued on equity grounds as the willingness to pay for health is strongly linked to an individual’s or a system’s ability to pay for it [Drummond 2005b]. Likewise, other modelling frameworks such as patient-level simulation and discrete event simulation are growing in popularity in the context of health economic evaluations as alternative to decision and cohort Markov models [Caro 2005, 2010].

3.6 Discussion

The quantitative synthesis of clinical data is a key and often necessary step to the REA of medical interventions both pre- and post-market launch [White 2000]. Meta-analysis is widely used to combine results from multiple clinical studies and considered best practice by many regulatory and HTA bodies worldwide. The potential advantages, as well as, standard methodology for conducting meta-analysis are well-established in the scientific community with acknowledged guidelines by the Cochrane Collaboration and the Centre for Reviews and Dissemination [CRD 2009, Higgins 2011]. The application of meta-analytical techniques to networks of studies has also grown considerably since the early 2000s with the increasing recognition of NMA to synthesise evidence from both direct and indirect treatment comparisons [Cooper 2011, Lumley 2002, Lu 2004]. In recent years, the ISPOR¹⁵ Indirect Treatment Comparisons Good Research Practices Task Force has been key in promoting ‘best practice’ for NMA and a number of HTA agencies have produced their own guidelines for I/MTC; for example, NICE and the NICE DSU in the UK continue to provide up-to-date methodological guidance on new statistical developments for NMA [Hoaglin 2011, Jansen 2011, NICE 2013a].

Alongside methodological developments in evidence synthesis, economic evaluations and particularly cost-effectiveness/utility analyses have grown to become a central feature of HTA practice worldwide. In fact, Ades *et al.* point out:

“Prior to the 1990s, most economic studies were undertaken primarily for publication and were not aimed at the requirements of specific decision makers. More recently, health systems internationally have begun to use CE research as a formal input into decisions about which interventions and programmes should be funded from collective resources” [2006: p1]

A number of guidelines have been published to critically appraise cost-effectiveness analysis and ensure the quality, validity and transparency of economic evaluations. Most notably, Drummond *et al.* [1996] developed a checklist for economic evaluations and Philips *et al.* [2004] reviewed ‘best practice’ standards for decision-analytical modelling in the context of HTA. As discussed in Chapter 2, a number of HTA agencies have also published pharmacoeconomic guidelines for authors of HTA. Moreover, a great deal of the cost-effectiveness research has been dedicated to address specific issues associated to

¹⁵International Society For Pharmacoeconomics and Outcomes Research

HTA. For this reason, this thesis considers all case studies in terms of both of relative effectiveness and cost-effectiveness.

Chapter 4

Choosing a relevant comparator: evaluating ticagrelor for ACS in Germany and the UK

4.1 Background

As mentioned in Chapter 2, many HTA agencies around the world have published comprehensive guidelines to help authors adhere to the ‘best principles’ and methods of HTA. Chapter 2 also exposed some discrepancies between these international guidelines including in the definition and selection of relevant comparator technologies for REA. In this chapter, I consider how choosing comparators for HTA can impact HTA outcomes using the cross-country example of ticagrelor for ACS.

4.1.1 Ticagrelor in acute coronary syndromes

Ticagrelor (Brilique™, AstraZeneca) is a novel platelet aggregation inhibitor to be administered orally, at a dose of 90mg twice daily, following onset of symptoms and/or heart surgery. The indicated population for ticagrelor is heterogeneous encompassing the broad diagnostic range for ACS in both primary and secondary care settings [Bassand 2007]. Patients presenting with chest pains—the leading symptom of reduced blood flow to the heart—are categorised, based on electrocardiogram (ECG) at admission and levels of cardiac enzymes, into two groups eligible for ticagrelor:

- ST-segment-elevation myocardial infarction (STEMI) with intent to treat by primary percutaneous coronary intervention (PCI) or
- non-ST-segment-elevation myocardial infarction (NSTEMI)

Medical management and/or revascularisation procedures such as PCI and coronary artery bypass grafts (CABG) are used to prevent thrombotic cardiovascular (CV) events such as myocardial infarctions (MI), ischemic attacks or strokes, in patients with ACS. Dual anti-platelet with acetylsalicylic acid (ASA, i.e. aspirin), unless contraindicated, is the UK standard of care for the medical management of ACS. Clopidogrel (Plavix™, Sanofi) is the most commonly used antithrombin therapy; updated NICE clinical

guidelines recommend its use for the early management of unstable angina and NSTEMI [2010a, 2013b]. For ACS patients who are to undergo PCI, prasugrel (Efient™, Eli Lilly) in combination with ASA is also available as an alternative treatment, although its use is limited in the UK [NICE 2009a, 2014].

In October 2011, NICE published a technology appraisal guidance—TA236 Ticagrelor for the treatment of ACS—recommending its use alongside low-dose aspirin for adult patients with ACS in the UK [2011a]. NICE also recommends initiating treatment with ticagrelor in hospitalised patients suspected of having unstable angina defined by NICE TA236 as “*ST or T wave changes on electrocardiogram suggestive of ischaemia and one risk factor for cardiovascular disease¹⁶*” [2011a]. Clopidogrel was identified as the main comparator of interest for the appraisal of ticagrelor; but prasugrel was also considered in a narrow group of ACS patients recommended for early PCI [NICE 2011a].

Both clopidogrel and prasugrel were selected by NICE as relevant comparators based on their licensed indications and previous recommendations for the UK setting [NICE 2009a, 2010, 2011a]. Since its approval by the EMA and FDA in the late 90s, clopidogrel has become the standard ‘add-on’ therapy to aspirin for a wide spectrum of cardiovascular diseases including ACS; professional bodies have also sanctioned its use worldwide (e.g. European Society of Cardiology, American Heart Association, and American College of Cardiology) [EMA 1998, FDA 1997, Zambahari 2007]. Hence the efficacy and safety of ticagrelor (with ASA) was consistently evaluated vs. clopidogrel (with ASA) in clinical trials for patients presenting with ACS regardless of revascularisation. However, concerns regarding the appropriateness of comparing ticagrelor with prasugrel were voiced by the manufacturers of both drugs during the NICE scoping consultation and later throughout the appraisal process [NICE 2010c]. In their submission, AstraZeneca highlighted the differences in patient populations and trial designs for the pivotal Phase III studies—PLATO and TRITON-TIMI 38—comparing ticagrelor and prasugrel to clopidogrel, respectively, and warned against an indirect comparison. Manufacturers’ concerns were echoed by the Evidence Review Group (ERG) who describes this comparison as

¹⁶ NICE considers the following risk factors to define treatment with ticagrelor for unstable angina: “*age 60 years or older; previous myocardial infarction or previous coronary artery bypass grafting (CABG); coronary artery disease with stenosis of 50% or more in at least two vessels; previous ischaemic stroke; previous transient ischaemic attack, carotid stenosis of at least 50%, or cerebral revascularisation; diabetes mellitus; peripheral arterial disease; or chronic renal dysfunction, defined as a creatinine clearance of less than 60 ml per minute per 1.73 m² of body-surface area.*” [2011a: p3]

“problematic” and support their decision not to perform an indirect treatment comparison (ITC) [Bagust 2011]. NICE requested in their appraisal scope that a subgroup analysis, taken from the PLATO-INVASIVE sub-study, be performed for the economic evaluation of ticagrelor. Thus, in order to address the decision problem set out by NICE for the appraisal of ticagrelor, AstraZeneca refer to a published ITC by Biondi-Zoccai *et al.* [2011] and submit a subgroup analysis to assess the cost-effectiveness of ticagrelor vs. prasugrel for an ‘invasive’ population. PLATO-INVASIVE included patients identified at randomisation with investigator intent for an early invasive strategy and more closely matched the prasugrel TRITON-TIMI 38 study population invasively managed by PCI [AstraZeneca 2010a].

4.2 Objectives

Despite evidential limitations, ticagrelor was assessed against both available alternatives in the UK—clopidogrel and prasugrel. However, HTA outcomes presented in TA236 may be very sensitive to the assumptions made with regards to the population and parameters included in the cost-effectiveness analysis of ticagrelor vs. prasugrel. Recent EUnetHTA guidelines on comparators and comparisons recommend the following:

“It is highly desirable that only comparators be used in REA for which a reasonable amount of good quality evidence is available.[...]There are situations where no good evidence for the effectiveness of the routine care is available, and in these situations no clear advice is given in national guidelines.” [2013: p13]

Using the example of ticagrelor in the treatment of ACS, I compare country-specific HTA processes and investigate the implications of different approaches to select relevant comparators on clinical and cost-effectiveness outcomes. The specific chapter objectives are:

- i. to critically compare all publically available ticagrelor HTA reports—published from national HTA agencies—with respect to decision problem outlined, comparator selection, and the use of clinical evidence in REA; and
- ii. to evaluate the impact of comparator selection on the clinical and cost-effectiveness analyses of ticagrelor from an NHS/PSS perspective.

4.3 International comparison of ticagrelor assessments

4.3.1 Methods

Using the list of international HTA agencies compiled in the Appendix CD A1, national HTA agency websites were searched for assessments of ticagrelor for ACS. Guidance, appraisals, as well as manufacturer submissions, were reviewed to compare country-specific assessments with their respective HTA guidelines and against NICE TA236 [2011a]. Where possible, documents not in English or French were translated using Google Translate. If a translation was not available or deemed insufficient for data extraction, assessments were excluded from the review. Information relevant to the selection of comparator technologies and appropriate subgroups of interest, clinical data sources and I/MTC was extracted to identify key differences in clinical data handling and data synthesis for the REA of ticagrelor.

4.3.2 Results

Web searches were conducted between January and February 2012 and identified seven assessments of ticagrelor for ACS patients. Reports from CADTH [2011], ZIN [2011], the German Institute for Quality and Efficiency in Health Care (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)) [2011], NICE [2011a], the Australian PBAC [2011], the Scottish Medicines Consortium (SMC) [2011], and the Swedish TLV [2011] were reviewed.

Table 4 summarises the key findings from the review of HTA reports in terms of their adherence to local methods guidelines; as well as, which comparators, primary data source(s) and indirect evidence were used in the country-specific REA of ticagrelor. Ticagrelor in combination with aspirin was evaluated for the prevention of atherothrombotic events in adult patients with ACS who are managed medically and those who are managed with PCI or CABG. The same target population was assessed by all agencies in accordance with ticagrelor's licensed indication and the PLATO patient population. The pivotal Phase III PLATO study comparing ticagrelor vs. clopidogrel was the main source of clinical data for all seven assessments; and when reported, the cost-effectiveness analyses were also predominantly trial-based. CADTH was the only agency to consider a regional subgroup analysis of the PLATO trial (PLATO North America) in their recommendation for ticagrelor. However, NICE requested an additional subgroup

analysis for the PLATO-INVASIVE sub-study population for the economic evaluation of ticagrelor.

Clopidogrel was defined as the most appropriate comparator by all HTA agencies given its widespread use in current clinical practice worldwide. Prasugrel was also considered as an alternative treatment option in five of the seven HTA reports; CADTH and PBAC did not recognise prasugrel as a potential comparator under its label for ACS patients undergoing primary or delayed PCI. Contextual variations such as the choice of comparator(s) were in line with selection criteria defined by local HTA guidelines, based on indicated population and current clinical practice, but these led to differences in the final decision problem outlined and the evidence base considered by each agency. In the absence of a head-to-head comparison between ticagrelor and prasugrel, the published ITC by Biondi-Zoccai *et al.* [2011] was identified by all but one HTA agency (PBAC) but only IQWiG performed an in-house ITC [2011]. The IQWiG indirect comparison used published results from the PLATO STEMI cohort (ticagrelor vs. clopidogrel) and TRITON-TIMI 38 (prasugrel vs. clopidogrel) to compare ticagrelor vs. prasugrel for the treatment of ACS patients undergoing PCI. IQWiG found the risk of bias to be low in both studies and reported endpoints, but stated that: “*because of the indirect comparison, the significance of the evidence [...] was reduced*” [2011]. In their manufacturer submission for NICE TA236, AstraZeneca argued that although the STEMI subgroups in the PLATO and TRITON-TIMI 38 trials may appear to lend themselves better to an indirect comparison—“*at first glance*”—the included patients were not similar enough to support an ITC¹⁷ [2010a]. The Final Appraisal Determination by NICE concluded the following on the submitted indirect evidence for ticagrelor:

“The manufacturer took the view that the [PLATO and TRITON] trials were not comparable and, by inference, a comparison between prasugrel and ticagrelor based on these trials was inappropriate and should be viewed with caution. [...] The ERG agreed with the manufacturer that sufficient clinical evidence is not yet available for a credible indirect comparison of ticagrelor plus aspirin compared with prasugrel plus aspirin” [2011b: p6]

¹⁷ The most important differences between PLATO and TRITON-TIMI 38 presented by AstraZeneca included the timing of PCI and the percentage of STEMI patients undergoing PCI and ‘secondary PCI’ (i.e. PCI >12 hours from onset of ACS symptoms), the loading dose for clopidogrel received in each trial’s control arm, and the assessment of MI in both trials [2010a].

The Dutch CVZ (now ZIN) concurred describing the interpretation of the Biondi-Zoccai *et al.*'s ITC as complicated, acknowledging that the conditions of “*homogeneity and conformity*” between studies had not been met and qualifying the indirect comparison as “*naïve or uncorrected*” [2011]. TLV refers to “*deficiencies*” in the indirect comparison but find it remained the best evidence available for the comparison of ticagrelor vs. prasugrel [2011]. NICE and TLV both reviewed manufacturer's cost-effectiveness model results comparing ticagrelor to prasugrel in a subset of ACS patients undergoing PCI [NICE 2011a, TLV 2011]. These comparisons were modelled using the HRs obtained by Biondi-Zoccai *et al.* and required a number of assumptions to be made to correct for differences in the ticagrelor and prasugrel patient populations of interest [2011].

Recommendations for reimbursement are also reported in Table 4 for each HTA agency. Five out of the seven HTA reports favoured the reimbursement or positive ‘listing’ of ticagrelor as a treatment option in ACS patients. In May 2011, Health Canada approved the use of ticagrelor in adult ACS patients, but later that year CADTH in a rapid response REA of clopidogrel, prasugrel and ticagrelor concluded: “*[ticagrelor's] place in therapy with respect to other antiplatelet agents [was] not clear*” [2011]. This position was reinforced by the Canadian Drug Expert Committee who recommended that ticagrelor not be listed by publicly funded drug plans based on the clinical results of the PLATO trial in North American patients, which did not justify its higher price [CADTH 2011]. In turn, IQWiG recommended that the indication for ticagrelor be restricted to unstable angina and NSTEMI by the German reimbursement authorities¹⁸ given the uncertainty around the added value for money of ticagrelor compared to clopidogrel and prasugrel, for medically managed STEMI patients and patients undergoing PCI, respectively [2011]. However, NICE guidance highlighted the favourable risk-benefit profile of ticagrelor in all ACS subgroups considered and found the ICERs vs. clopidogrel and prasugrel to be within an acceptable range for the cost-effective use of NHS resources [2011a].

¹⁸ The Federal Joint Committee (Gemeinsamersame Bundes-aus-schuss (G-BA))

Table 4 Summary of international HTA reports for ticagrelor

	NICE [2011]	CADTH [2011]	ZIN (CVZ) [2011]	IQWiG [2011]	PBAC [2011]	SMC [2011]	TLV [2011]
Manufacturers submissions' adherence to local HTA guidelines							
<i>Population</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Comparators</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Identification of clinical evidence</i>	Yes	NA ¹⁹	Yes	Yes	Yes	Yes	Yes
Relative effectiveness assessment (REA)							
<i>Target population</i>	ACS patients	ACS patients	ACS patients	ACS patients	ACS patients	ACS patients	ACS patients
<i>Subgroups of interest</i>	STEMI, NSTEMI, UA, INVASIVE	STEMI, NSTEMI, UA, North America (region)	STEMI, NSTEMI, UA	STEMI, NSTEMI, UA	STEMI, NSTEMI, UA	STEMI, NSTEMI, UA	STEMI, NSTEMI, UA
<i>Comparators selected</i>	clopidogrel, prasugrel	clopidogrel	clopidogrel, prasugrel	clopidogrel, prasugrel	clopidogrel	clopidogrel, prasugrel	clopidogrel, prasugrel
<i>Primary efficacy data source(s)</i>	PLATO trial, published ITC	PLATO trial	PLATO trial, published ITC	PLATO trial, published ITC	PLATO trial	PLATO trial, published ITC	PLATO trial, published ITC
<i>Evidence synthesis performed</i>	inappropriate	not reported	not reported	adjusted ITC	not reported	not reported	not reported
Recommendation for reimbursement							
	Recommended	Not recommended	Recommended	Restricted indication	Recommended	Recommended	Recommended

UA: unstable angina

¹⁹A full assessment was not undertaken by CADTH.

4.4 Cost-effectiveness analysis

Despite recognised methodological limitations, indirect evidence was used by three agencies—NICE, IQWiG, and TLV—to assess the cost-effectiveness of ticagrelor against all relevant comparators in their respective countries. However, discrepancies in how these agencies selected populations of interest for the comparison of ticagrelor vs. prasugrel, namely NICE and IQWiG, could have contributed to their conflicting recommendations for STEMI patients intensively managed by PCI.

Based on the *de novo* cost-effectiveness analysis for the UK put forward in the manufacturer's submission for NICE TA236 [2011a], I reconstructed an economic model to assess the impact that selecting population(s) and comparator(s) of interest may have on HTA outcomes. Holding economic model parameters constant as efficacy inputs were varied for ticagrelor vs. prasugrel allowed me to evaluate the impact of that relevant clinical evidence—as identified and interpreted by different HTA agencies—on both the relative effectiveness and cost-effectiveness of ticagrelor in the UK.

4.4.1 Methods

In this section, I describe the model structure and input parameters used to construct the cost-effectiveness analysis for ticagrelor from the NHS/PSS perspective, as per the reference case by NICE. The two approaches taken by NICE and IQWiG to compare ticagrelor vs. prasugrel are presented and evaluated within the same economic model. HTA outcomes are given as ICERs in QALYs per GBP (£).

The comparison of ticagrelor vs. clopidogrel was considered the base case and used to validate the reconstructed economic model results against the ICERs reported in the NICE TA236 manufacturer submission [2011a]. For the comparison of ticagrelor vs. prasugrel, I refer to the PLATO-INVASIVE and PLATO-STEMI subgroup analyses to describe the NICE and IQWiG approaches, respectively.

4.4.1.1 Patients

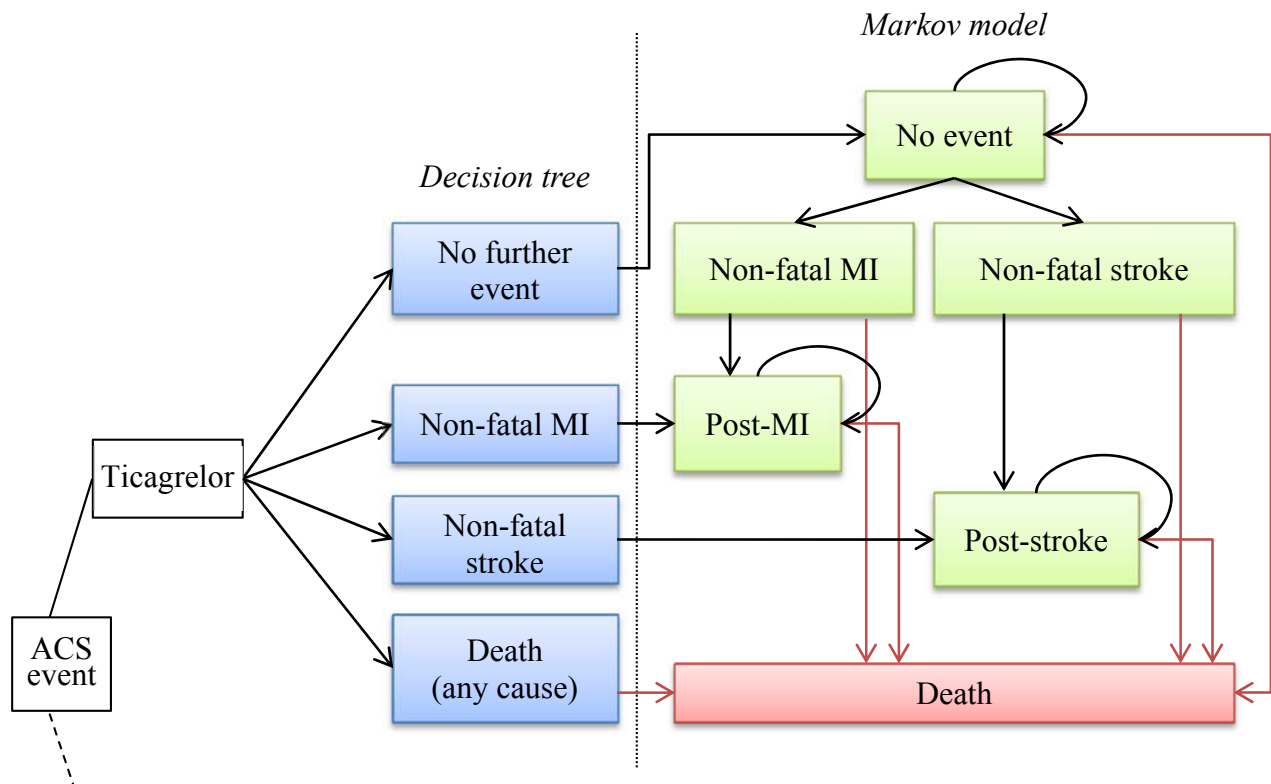
For the base case economic evaluation, patients with ACS (STEMI, NSTEMI, and unstable angina) managed medically or managed with PCI or CABG were included as per ticagrelor's marketing authorisation [EMA 2010]. The target patient population matched that of the large randomised controlled clinical trial assessing the safety and efficacy of ticagrelor (i.e. PLATO). The PLATO (PLATelet inhibition and patient

Outcomes) study compared ticagrelor (a single 180 mg loading dose, two tablets of 90mg and then 90 mg twice daily thereafter) with clopidogrel (300 or 600 mg loading dose, 75 mg thereafter) in 18,624 patients with ACS including a UK population (Wallentin 2009). Sub studies PLATO-INVASIVE (13,408 patients out of 18,624), PLATO-STEMI (8,430 patients), and PLATO-HECON (18,624 patients) were also used in the economic evaluation to inform the subgroup analyses, quality-of-life mapping and the resource use patterns for ACS patients.

4.4.1.2 Model structure

The economic model was designed in accordance with the model structure described in the manufacturer submission for TA236 [AstraZeneca 2010a]. The model was constructed in Excel 2010 and made up of two-parts: a one-year decision tree and a Markov process. The combination of decision tree and Markov model reflected both the clinical trial data and the long-term care pathways for ACS in the UK. Results from the PLATO study at 12 months were used in the decision tree; and major costs and clinical outcomes were extrapolated in the Markov model to capture patients' experiences over the remainder of their lives. Figure 2 provides a diagrammatical representation of the two-part model.

Figure 2 Economic model structure diagram



Patients admitted to hospital with ACS symptoms—thereafter referred to as initial ACS event—were treated with ticagrelor following early diagnosis of STEMI, NSTEMI or unstable angina. Patients were allocated to four health states in the first year before entering six Markov states over a time horizon of 40 years. The model used an annual cycle length after the first year of treatment.

Health states in the decision tree are depicted in blue nodes in Figure 1 and described as follows:

- No further event – includes patients who have experienced no further event in the first year following initial ACS event.
- Non-fatal MI – includes patients who have experienced a non-fatal MI in the first year following initial ACS event.
- Non-fatal stroke – includes patients who have experienced a non-fatal stroke in the first year following initial ACS event.
- Death from any cause – includes patients who died from non-vascular or vascular event, including fatal MI or stroke, in the first year following initial ACS event.

After one year, patients started in one of four of the six Markov health states based on the decision-tree. Patients in the ‘No further event’ state moved to the ‘No event’ Markov state and patients who died remained in the absorbing ‘Death’ Markov state. Patients who experienced a non-fatal MI or stroke moved to the ‘Post-MI’ and ‘Post-stroke’ Markov states, respectively. The Markov states ‘Non-fatal MI’ and ‘Non-fatal stroke’ only captured patients who experienced such an event at least one-year after initial ACS event. Health states in the Markov model are depicted in green or red nodes in Figure 2 and described as follows:

- No event – includes patients who have experienced no further event in the decision tree and patients who remained event-free at the end of each Markov model cycle. Each year, patients in this health state were at risk of a non-fatal MI, a non-fatal stroke, or death; and if such an event was experienced, patients transitioned to the ‘Non-fatal MI’, ‘Non-fatal stroke’, or ‘Death’ states, respectively.
- Non-fatal MI – includes patients who experienced a new non-fatal MI after initial one-year decision tree. This health state corresponds to the prognosis of survival in the first year post-MI. After one year, patients who survived transitioned to ‘Post-MI’ state and patients who died moved to the absorbing ‘Death’ state.

- Non-fatal stroke – includes patients who experienced a new non-fatal stroke after initial one-year decision tree. This health state corresponds to the prognosis of survival in the first year post-stroke. After one year, patients who survived transitioned to ‘Post-stroke’ state and patients who died moved to the absorbing ‘Death’ state.
- Post-MI – includes patients who have experienced a non-fatal MI in the decision tree and patients who suffered a non-fatal MI in any of the subsequent Markov cycles. This health state corresponds to the prognosis of survival in the second year and subsequent years post-MI. Each year, patients in this health state were at risk of death; and if patients died they transitioned to the absorbing ‘Death’ state.
- Post-stroke – includes patients who have experienced a non-fatal stroke in the decision tree and patients who suffered a non-fatal stroke in any of the subsequent Markov cycles. This health state corresponds to the prognosis of survival in the second year and subsequent years post-stroke. Each year, patients in this health state were at risk of death; and if patients died they transitioned to the absorbing ‘Death’ state.
- Death – includes patients who died from any cause in the decision tree and during any Markov cycle.

All health states are mutually exclusive and represent key cardiovascular events for which an ACS population is at risk both in the acute and long-term phase of the disease [AstraZeneca 2010a]. The model allows for a worse prognosis in the first year following a non-fatal MI or stroke in line with clinical trial findings [Wallentin 2009]. Patients were allocated to health states based on the first event they experienced in the year (i.e. MI or stroke). This assumption was made by AstraZeneca and implies that the non-fatal MI state also captured patients that have experienced an MI followed by a stroke, and vice-versa for the non-fatal stroke state. However, death took precedence over other non-fatal events and patients were assigned to the ‘Death’ state even if they had previously experienced an MI and/or stroke in that cycle.

The same decision tree was used for clopidogrel in the base case and prasugrel in the subgroup analysis. Based on clinical trial data, only the event rates in the decision tree were conditional to treatment in the first year following initial ACS event. The conservative approach proposed by AstraZeneca assigns the same transition probabilities to all patients in the Markov model irrespective of treatment. The only difference beyond

the one-year decision tree for the ticagrelor, clopidogrel, and prasugrel models was the number of patients in each starting Markov state.

4.4.1.3 Clinical model parameters

4.4.1.3.1 Base case analysis

This section provides a summary of the clinical data used in the economic evaluation and how transition probabilities were derived for the decision tree and Markov model. Based on the model structure described in section 4.4.1.2, the following probabilities were required to populate the one-year decision tree:

- Probability of having no further event in the first year following initial ACS event
- Probability of having non-fatal MI in first year following initial ACS event
- Probability of having non-fatal stroke in first year following initial ACS event
- Probability of dying from any cause in first year following initial ACS event

AstraZeneca calculated these probabilities using a parametric time-to-event survival model, fitted with a Weibull distribution, to transform the crude proportions of patients with each event based on count data from the PLATO study [2010a]. For the base case, a baseline risk (or event rate) was estimated for clopidogrel to which a HR was applied to obtain the transition probabilities for ticagrelor. The probability of having no further event was calculated as one minus the combined risks of the three other events occurring. All probabilities were also adjusted for age and gender to account for the differences between patient characteristics in the PLATO trial population and in ACS patients in England and Wales. A detailed explanation of the calculations performed by AstraZeneca is provided in the manufacturer submission for TA236 [2010a]. Table 5 is taken from the latter submission and lists the clinical variables used in the one-year decision tree.

Table 5 Summary of clinical variables in one-year decision-tree

Variable	Value (95% CI)	Distribution	Source
General			
Mean age	70		MINAP/GPRD study ²⁰
% Male	64.6%		
% of patients ≥ 75	42.7%		
Event rates for clopidogrel			
Dead any cause	0.0789 (0.0518-0.1202)	Weibull	Weibull regression equations based on PLATO study
Non-fatal MI	0.0628 (0.0426-0.0935)	Weibull	
Non-fatal stroke	0.0112 (0.0039-0.0347)	Weibull	
Dead vascular	0.0672 (0.0436-0.1038)	Weibull	
Hazard ratios for ticagrelor vs. clopidogrel			
Dead any cause	0.7845 (0.6880-0.8945)	LogNormal	Weibull regression equations based on PLATO study
Non-fatal MI	0.8598 (0.7546-0.9797)	LogNormal	
Non-fatal stroke	1.0894 (0.7949-1.4930)	LogNormal	
Dead vascular	0.7946 (0.6908-0.9139)	LogNormal	
Event rates for ticagrelor			
Death any cause	0.0619 (0.0543-0.0706)	N/A	Combination of clopidogrel event rates and ticagrelor HRs
Non-fatal MI	0.0540 (0.0474-0.0615)	N/A	
Non-fatal stroke	0.0122 (0.0089-0.0167)	N/A	
Dead vascular	0.0534 (0.0464-0.0614)	N/A	

Table extracted from section 6.3.6 Table 6.17 of the AstraZeneca manufacturer submission for TA236 [2010a].

The transition probabilities required to populate the Markov model are described in the transition matrix in Table 6.

²⁰MINAP/GPRD refers to the Myocardial Ischaemia National Audit Project/General Practice Research Database study [AstraZeneca 2010a].

Table 6 Transition matrix for Markov model

Transition from:	Transition to:					
	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death
No event	1 - combined risk of all other events	Probability of having non-fatal MI after one-year decision tree	Probability of having non-fatal stroke after one-year decision tree	0	0	Age/gender specific mortality rate adjusted for ACS
Non-fatal MI	0	0	0	1 – probability of ‘Death’	0	Age/gender specific mortality rate adjusted for ACS and MI in last year
Non-fatal stroke	0	0	0	0	1- probability of ‘Death’	Age/gender specific mortality rate adjusted for ACS and stroke in last year
Post-MI	0	0	0	0	0	Age/gender specific mortality rate adjusted for ACS with no MI in past year
Post-stroke	0	0	0	0	0	Age/gender specific mortality rate adjusted for ACS with no stroke in past year
Death	0	0	0	0	0	1 (Absorbing state)

Matrix adapted from Section 6.3.2 Table 6.11 of the AstraZeneca manufacturer submission for TA236 [2010a].

With the exception of the probabilities of having a non-fatal MI or non-fatal stroke after the initial one-year decision tree; all the transition probabilities in the Markov model were estimated by applying a relative risk to the probability of death. In the AstraZeneca manufacturer submission, the probability of death or mortality rate was taken from the standard life tables for the UK published by the Office for National Statistics [2010a]. In order for all things to remain equal, the same life tables from the years 2007-2009 were used in my reconstructed model [ONS 2013]. Moreover a weighted average mortality rate was calculated based on the percentage of male and female patients with ACS in the UK (cf. Table B2 in the Appendix B). Standardised mortality ratios were estimated from the literature and applied to age and gender specific mortality rates to reflect the increased risk of death for ACS patients, patients with recurrent MI or stroke, and patients in the post-MI and post-stroke health states. A detailed account of how the standardised mortality ratios were identified by AstraZeneca is provided in the manufacturer submission for TA236 [2010a].

Table 7 lists the clinical variables used in the Markov model. Since no treatment effects were extrapolated beyond the 12 months clinical trial period, Markov transition probabilities were modelled irrespective of intervention.

Table 7 Summary of clinical variables in Markov model

Variable	Value (95% CI)	Distribution	Source
Event rates			
Non-fatal MI	0.0315 (0.0257-0.0385)	Beta	MINAP/GPRD study
Non-fatal stroke	0.0102 (0.0072-0.0145)	Beta	
Hazard ratios relative to standard life tables			
No event	2.2121 (0.1817-4.2425)	LogNormal	CG94 [2010] and Allen <i>et al.</i> [2006]
Non-fatal MI	5.8446 (3.7176-7.9717)	LogNormal	
Post-MI	2.2121 (0.1817-4.2425)	LogNormal	
Non-fatal stroke	7.4286 (6.50-8.50)	LogNormal	Dennis <i>et al.</i> [1993]
Post-stroke	2.0715 (1.30-3.32)	LogNormal	

Table extracted from section 6.3.6 Table 6.17 of the AstraZeneca manufacturer submission for TA236 [2010a].

4.4.1.3.2 Subgroup analysis 1: PLATO-INVASIVE

A first subgroup analysis was performed based on the PLATO-INVASIVE sub-study and the ITC published by Biondi-Zoccai *et al.* [2011]. The PLATO-INVASIVE sub-study was briefly described in section 4.4.1.1 and compared ticagrelor vs. clopidogrel in invasively managed ACS patients [Canon 2010]. A comparison of ticagrelor vs. prasugrel

in an invasive subgroup of ACS patients was requested in the NICE scope for TA236 [2011a]. A similar method to that used in the base case was applied to the PLATO-INVASIVE subgroup, i.e. a Weibull regression to calculate the baseline event rates for clopidogrel in the one-year decision tree. AstraZeneca made the following assumption to calculate HRs for ticagrelor vs. clopidogrel:

“Based on the fact that there was no statistically significant interaction between the primary endpoint and final diagnosis ($p=0.41$) the hazard ratio for the overall population was used to generate the event rate for ticagrelor [in the PLATO-INVASIVE subgroup].” [2010: p220]

Biondi-Zoccai *et al.* reported the ORs for prasugrel vs. ticagrelor for MI, stroke, and death; as well as for the adverse events: major bleeding, minor bleeding, and definite/probable stent thrombosis [2010]. AstraZeneca converted the ORs based on Biondi-Zoccai *et al.*'s adjusted ITC into relative risks (RRs) to calculate the event rates for prasugrel using the conversion formula in the *Cochrane Handbook for Systematic Reviews of Interventions* [Higgins 2011]:

$$RR = OR / [1 - (\text{Control risk} * (1-OR))] \quad (1)$$

Table 8 summarises the ORs extracted from Biondi-Zoccai *et al.* [2011] and the calculations made by AstraZeneca to obtain the relative risks used in the economic model [2010a]. Table 9 presents the event rates for ticagrelor and prasugrel estimated from the relative risks. Events rates for ‘No event’ were calculated as one minus the combined risks of the three other events occurring. Parameters in the Markov model remained unchanged for the PLATO-INVASIVE subgroup analysis.

Table 8 Conversion of results from Biondi-Zoccai *et al.* [2011] into relative risks for health economic modelling

Outcome	Values from Biondi-Zoccai <i>et al.</i> [2011]			Inversed value (so that <1 favours ticagrelor)			Control risk (prasugrel)	Converted to relative risk (using Cochrane Handbook)		
	Odds ratio	95% lower	95% upper	Odds ratio	95% lower	95% upper		Relative risk	95% lower	95% upper
Primary endpoint	0.987	0.861	1.133	1.01	0.88	1.16	0.099	1.01	0.89	1.14
MI	0.893	0.75	1.062	1.12	0.94	1.33	0.073	1.11	0.95	1.30
Stroke	0.856	0.55	1.331	1.17	0.75	1.82	0.03	1.17	0.75	1.80
All-cause mortality	1.218	0.959	1.546	0.82	0.65	1.04	0.03	0.83	0.65	1.04
Stent thrombosis	0.635	0.433	0.932	1.57	1.07	2.31	0.011	1.56	1.07	2.28
Major bleeding	1.431	1.103	1.858	0.70	0.54	0.91	0.025	0.70	0.54	0.91
Minor bleeding	1.073	0.794	1.451	0.93	0.69	1.26	0.02	0.93	0.69	1.25

Table replicated from Section 5.7.6 Table 5.17 of the AstraZeneca manufacturer submission for TA236 [2010a].

Table 9 Event rates for ticagrelor and prasugrel in PLATO-INVASIVE subgroup analysis (1)

	PLATO-Invasive	
Event rates	Ticagrelor	Prasugrel
Death any cause	3.9%	4.7%
MI	5.3%	4.8%
Stroke	1.2%	1.0%
No event	89.6%	89.5%
Adverse events rates	Ticagrelor	Prasugrel
Major bleeding	7.9%	11.3%
Minor bleeding	3.8%	4.1%
Stent thrombosis	2.2%	1.5%

4.4.1.3.3 Subgroup analysis 2: PLATO-STEMI

IQWiG performed an ITC to combine evidence from the PLATO and TRION-TIMI 38 studies for STEMI patients undergoing PCI. Clinical endpoints estimated indirectly included total mortality, cardiovascular mortality, myocardial infarction, and stroke; data on adverse events and discontinuation rates were also synthesised. Details of the ITC methods used by IQWiG were not provided in the ticagrelor benefit assessment report [2011]. However, the HRs obtained from the ITC were reported as used as such in my analysis. Table 10 presents the ITC results alongside the original data extracted from the PLATO (STEMI group) and TRITON-TIMI 38 studies for comparison.

Table 10 Hazard ratios for ticagrelor vs. prasugrel from IQWiG ITC for health economic modelling

	Hazard ratio [95% CI] Ticagrelor vs. Clopidogrel	Hazard ratio [95% CI] Prasugrel vs. Clopidogrel	Hazard ratio [95% CI] Ticagrelor vs. Prasugrel
Death any cause	0.69 [0.53, 0.90]	0.69 [0.41; 1.16]	1.00 [0.56, 1.79]
MI	0.68 [0.53, 0.89]	0.70 [0.53, 0.92]	0.97 [0.67, 1.42]
Stroke	1.51 [0.90; 2.53]	1.05 [0.48, 2.30]	1.44 [0.56, 3.68]

Table extracted from Section 2.4.3 Table 12 of the IQWiG Benefit Assessment for ticagrelor [2011].

Using a similar approach than that in section 4.4.1.3.2 for subgroup analysis 1, reported HRs were used to calculate event rates for prasugrel for the one-year decision tree. The adverse events rates for ticagrelor were extracted from the PLATO-STEMI sub-study; and the adverse events rates for prasugrel remained the same as in subgroup analysis 1 (cf. Biondi-Zoccai 2011). Table 11 presents the event rates for ticagrelor and prasugrel for the PLATO-STEMI subgroup. Events rates for ‘No event’ were calculated as one minus the combined risks of the three other events occurring. Parameters in the Markov model remained unchanged for the PLATO-STEMI subgroup analysis.

Table 11 Event rates for ticagrelor and prasugrel in PLATO-STEMI subgroup analysis (2)

	PLATO-STEMI	
Event rates	Ticagrelor	Prasugrel
Death any cause	5.7%	5.7%
MI	3.5%	3.6%
Stroke	1.2%	0.8%
No event	89.6%	89.9%
Adverse events rates	Ticagrelor	Prasugrel
Major bleeding	9.0%	11.3%
Minor bleeding	4.9%	4.1%
Stent thrombosis	2.6%	1.5%

4.4.1.4 Utility valuation

Patients with ACS can experience a number of vascular events such as MI or unstable angina that impact their quality-of-life in the short-term including pain, discomfort, and hospitalisation. Revascularisation may be required for STEMI patients and is associated

with longer hospital stays and recovery times. In the long-term, patients' quality of life may suffer from recommended lifestyle changes to help prevent the recurrence of an ACS event. PLATO-HECON was a pre-specified PLATO Health Economics and Quality of Life sub-study designed to collect data on health-related quality of life (HRQoL) and resource use patterns for enrolled ACS patients. Utility scores converted from EQ-5D questionnaires and accrued over the 12 months PLATO trial follow-up were used in the cost-effectiveness analysis. These scores were obtained for both the ticagrelor and clopidogrel study arms and for the four nodes in the decision tree. Table 12 summarises the utility scores adjusted for age by health state and treatment used in the one-year decision-tree.

Table 12 Quality-of-life values in one-year decision tree

Health state	Utility score	Standard error	Source
No Event (ticagrelor)	0.840	0.003	PLATO HECON sub- study
Non-fatal MI (ticagrelor)	0.786	0.014	
Non-fatal Stroke (ticagrelor)	0.709	0.062	
Death Any Cause (ticagrelor)	0.211	0.021	
No Event (clopidogrel)	0.844	0.003	
Non-fatal MI (clopidogrel)	0.774	0.014	
Non-fatal Stroke (clopidogrel)	0.695	0.032	
Death Any Cause (clopidogrel)	0.220	0.019	

Table adapted from Section 6.4.9 Table 6.29 of the AstraZeneca manufacturer submission for TA236 [2010a].

AstraZeneca assumed an average utility score for both treatment groups adjusted for age for the 'No event', 'Non-fatal MI', and 'Non-fatal stroke' Markov health states [2010a]. The expected utility for a patient in the 'Post-MI' Markov state was estimated based on elicited values from Lacey *et al.* suggesting that HRQoL improved one-year post MI [2003]. However, based on a review of the literature, no such improvement in HRQoL was found in patients who had a stroke in the past and therefore the utility value for the 'Post-stroke' Markov state was assumed to be the same as that for 'Non-fatal stroke'. Table 13 summarises the utility scores adjusted for age by health state in the Markov model.

Table 13 Quality-of-life values in Markov model

Health state	Utility score	Standard error	Source
No event	0.842	0.002	PLATO HECON sub-study
Non-fatal MI	0.779	0.010	
Post MI	0.821	0.038	As above plus Lacey <i>et al.</i> [2003]
Non-fatal Stroke	0.703	0.010	PLATO HECON sub-study
Post Stroke	0.703	0.038	As above plus assumption
Death	0.000	N/A	N/A

Table adapted from Section 6.4.9 Table 6.29 of the AstraZeneca manufacturer submission for TA236 [2010].

Baseline HRQoL was calculated based on the health state in which a patient ended the one-year decision tree and started the Markov model in. However, AstraZeneca corrected this baseline quality-of-life using a one-off age decrement adjustment from Kind *et al.* to account for the older UK ACS population (mean age 70.4) compared to the PLATO population (mean age 62.2) [1998, AstraZeneca 2010]. An annual age decrement of 0.004 was subsequently used in the Markov model to reflect the relative loss in HRQoL as patients get older.

In addition, although adverse events were not modelled as specific health states, utility decrements were used in the economic model to reflect the negative impact on HRQoL of adverse events associated with antiplatelet therapy. Bleeding is one of the most important safety issues reported for ACS medications. Both major and minor bleeds were assigned a utility decrement of 0.1426 and 0.0033, respectively, based on suggested values in the literature and recent NICE appraisals [AstraZeneca 2010a]. Stent thrombosis was also identified as a key adverse event by Biondi-Zoccai *et al.* and in the PLATO clopidogrel treatment arm [2011]. A utility decrement of 0.06 was included for stent thrombosis based on Garg *et al.* estimate of a net annual disutility for revascularisation [2008].

In the reconstructed model, I applied a Beta distribution to utility scores for which a standard error was provided to estimate the 95% confidence interval and allow for sampling during the PSA. A half-cycle correction and a discount rate of 3.5% were applied yearly to QALYs.

4.4.1.5 Resource use and costs

In their manufacturer submission, AstraZeneca used a within-trial costing analysis from the PLATO-HECON sub-study to input relevant health service costs for the first year of the economic model. PLATO-HECON captured the resource utilisation of all patients in the PLATO study within the 12 months trial period including hospitalisations, interventions, investigations, and bleeding-related health care consumption [AstraZeneca 2010a]. AstraZeneca derived costs for each health state associated with ticagrelor and clopidogrel in the one-year decision tree from PLATO-HECON datasets [2010a]. Health state costs were marginally updated in an amendment to their original submission revising resource use for hospitalised patients in the trial; the amended costs were used in my reconstructed model [2011a]. Since no head-to-head trial data was available for prasugrel, it was assumed that the health state costs for the subgroup analyses PLATO-INVASIVE and PLATO-STEMI were the same for prasugrel as for ticagrelor. All costs were reported in GBP (£) and inflated to 2008/2009, in accordance with the manufacturer submission for TA236 [AstraZeneca 2010a].

AstraZeneca sourced drug costs from the NHS Drug Tariff for England and Wales (November 2010) and the Monthly Index of Medical Specialities (MIMS) (October 2010) for clopidogrel and prasugrel, respectively; and priced treatment with ticagrelor at £713.70 annually including a loading dose and maintenance dose for 12 months following initial ACS event [2010]. Table 14 presents the annual drug costs for ticagrelor, clopidogrel, prasugrel, and aspirin. Annual costs were calculated based on indicated treatment regimens and concomitant use of aspirin alongside all antiplatelet drugs.

Table 14 Annual drug costs

	Ticagrelor	Clopidogrel	Prasugrel	Aspirin (ASA)	Source
Drug costs for 12 months treatment duration	£713.70	£42.10	£628.47	£10.78	AstraZeneca, Drug Tariff, MIMS

Table 15 lists the mean health state costs per treatment for the one-year decision tree; upper and lower quartiles for health state costs were also reported [AstraZeneca 2010a].

Table 15 Mean health state costs per treatment for the decision tree

Health states	Ticagrelor			Clopidogrel			Source
	Mean	Low	High	Mean	Low	High	
No event	£8,573	£6,307	£10,053	£8,676	£6,378	£10,514	PLATO- HECON costing analysis
Non-fatal MI	£16,767	£12,258	£19,871	£16,563	£12,221	£19,486	
Non-fatal stroke	£15,455	£11,372	£18,414	£17,576	£13,000	£20,896	
Death (any cause)	£11,926	£8,697	£13,847	£14,078	£10,305	£16,489	

Costs inflated to 2008/2009, in accordance with the manufacturer submission for TA236 [AstraZeneca 2010a]

The costs for Markov health states and adverse events were inputted separately and assumed to be the same for all three treatment options. Table 16 summarises the health states costs for the Markov model and Table 17 the adverse events costs. Mean values as well as lower and upper cost ‘boundaries’ were estimated for each Markov health state and adverse event based on the wide range of values found in the literature. No costs were assigned to ‘Death’ in the Markov model.

Table 16 Health state costs for the Markov model

Health States	Mean	Lower	Upper	Source
No event	£217	£163	£1,793	Robinson <i>et al.</i> [2002]
Non-fatal MI	£5,003	£1,721	£5,762	
Post-MI	£285	£217	£2,002	
Non-fatal stroke	£13,084	£12,571	£13,604	Youman <i>et al.</i> [2003]
Post-stroke	£3,632	£3,317	£3,956	

Costs inflated to 2008/2009, in accordance with the manufacturer submission for TA236 [AstraZeneca 2010a]

Table 17 Adverse events costs

Adverse events	Mean	Lower	Upper	Source
Major bleed	£1,260	£960	£1,440	MINAP, NHS reference costs
Minor bleed	£420	£960	£1,440	
Stent thrombosis	£2,821	£2,192	£3,390	

Costs inflated to 2008/2009, in accordance with the manufacturer submission for TA236 [AstraZeneca 2010a]

In the reconstructed model, I applied a Uniform distribution to costs for which upper and lower bounds were provided to allow for sampling during the PSA. A half-cycle correction and discount rate of 3.5% were applied yearly to costs.

4.4.2 Results

A deterministic and probabilistic sensitivity analyses were implemented in the reconstructed economic model in Excel 2010. The PSA was run for 1000 iterations sampling values from distributions, where applicable. The Markov trace for the reconstructed and the AstraZeneca *de novo* models were compared for the base case, PLATO-INVASIVE and PLATO-STEMI subgroup analyses, to ensure the robustness of my cost-effectiveness analysis against the original submission. The Markov traces for the reconstructed model are listed in the Appendix B3 to B7; these provide the proportion of patients in each Markov health state at each model cycle, as well as total costs, life-years gained and QALYs (uncorrected and corrected for half-cycle). Table 18 summarises the ICERs for ticagrelor vs. clopidogrel in the overall ACS population (base case), ticagrelor vs. prasugrel in invasively managed ACS patients (subgroup analysis 1 – NICE approach), and ticagrelor vs. prasugrel for STEMI patients undergoing PCI (subgroup analysis 2 – IQWiG approach). Note that since results from the ITC were used in the STEMI subgroup analysis, a three-way comparison was possible and results are presented in Table 18.

Table 18 Cost-effectiveness results (PSA means) for clopidogrel, prasugrel, and ticagrelor from reconstructed economic model

	Total costs	Total LYs	Total QALYs	ICER (£/QALYs)
Base case (ACS patients) - ticagrelor vs. clopidogrel				
Clopidogrel	£17,982	7.58	6.31	-
Ticagrelor	£18,409	7.72	6.42	£3,443
Subgroup analysis 1 (PLATO-INVASIVE) - ticagrelor vs. prasugrel				
Prasugrel	£23,269	11.22	9.27	-
Ticagrelor	£23,510	11.30	9.34	£3,882
Subgroup analysis 2 (PLATO-STEMI) – ticagrelor vs. clopidogrel				
Clopidogrel	£19,336	8.54	7.01	-
Prasugrel	£19,640	8.79	7.28	£1,126
Ticagrelor	£19,825	8.79	7.28	<i>Dominated</i>

Results from the NICE TA236 concluded that, in the PLATO-INVASIVE subgroup, ticagrelor was highly cost-effective vs. prasugrel with a cost per QALY of £3,482 at a 40-year time horizon [AstraZeneca 2010a]. The replicated analysis obtained similar results with a probabilistic mean ICER of £3,882 per QALY. Using the HRs obtained by IQWiG for the PLATO-STEMI subgroup in my model, ticagrelor was dominated and prasugrel to be most cost-effective. The uncertainty in the ICERs for ticagrelor vs. prasugrel is depicted in the cost-effectiveness planes for both subgroups in Figure 3a and 4a. Figure 3b and 4b plot the CEAC for INVASIVE and STEMI (PCI) subgroups analyses, respectively, on a willingness to pay scale of £0 to £40,000. The contrasting shapes of the CEACs illustrate the differences in the probabilities of ticagrelor being cost-effective vs. prasugrel in the two patient populations selected by NICE and IQWiG.

Figure 3 (a) Cost-effectiveness plane and (b) CEAC of ticagrelor vs. prasugrel in INVASIVE population (subgroup analysis 1)

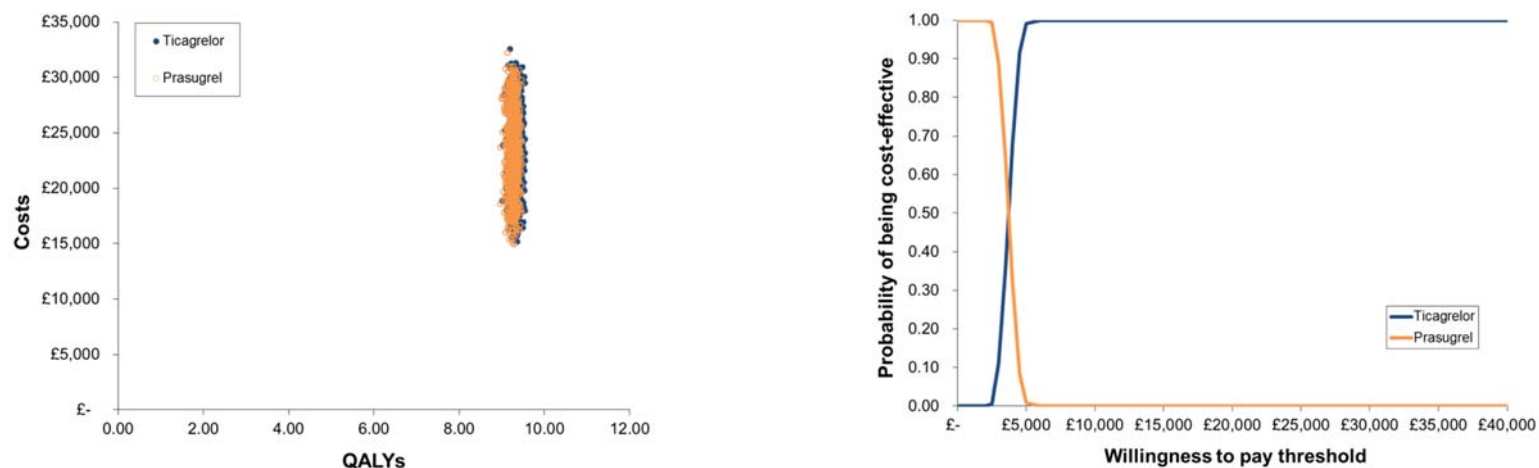
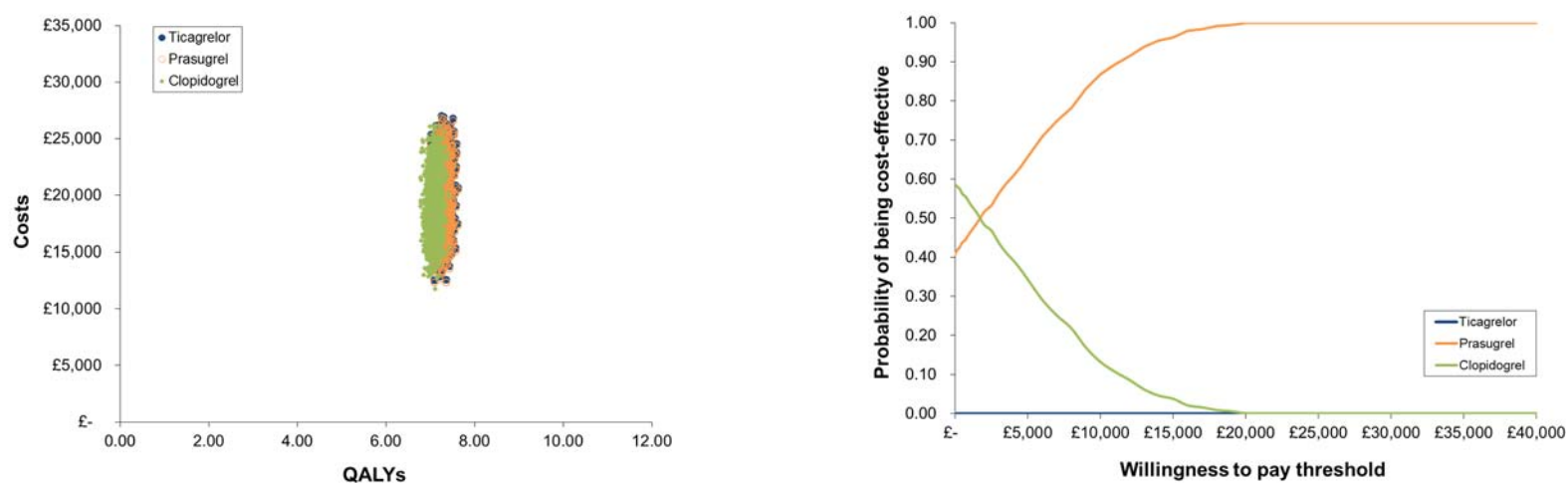


Figure 4 (a) Cost-effectiveness plane and (b) CEAC of ticagrelor vs. prasugrel in STEMI (PCI) population (subgroup analysis 2)



4.5 Discussion

The REA of ticagrelor presented a gap in the evidence for the comparison vs. prasugrel in an appropriate subset of patients with ACS. Although all seven HTA reports reviewed adhered to their local guidelines to select population(s) and comparator(s) of interest within their jurisdiction; different interpretations of the indication and clinical trial evidence for ticagrelor resulted in different treatment comparisons.

Despite using the same evidence base, cross-country comparisons of ticagrelor HTA reports suggested that contextual factors could play a key role in explaining international discrepancies in HTA recommendations. In this chapter, I focused on understanding why IQWiG recommended ticagrelor's use be restricted in Germany to NSTEMI and unstable angina patients, favouring treatment with prasugrel for STEMI patients undergoing PCI. This conclusion differed from the NICE TA236 guidance which found ticagrelor to be cost-effective and within the acceptable willingness to pay threshold for the NHS compared to existing pharmaceutical alternatives (i.e. clopidogrel and prasugrel).

The critical review of NICE and IQWiG's assessments highlighted the different subgroups of interest chosen for comparison against prasugrel. Despite limitations (cf. section 4.5.1), both agencies considered indirect evidence in the economic evaluation of ticagrelor vs. prasugrel. NICE and manufacturers argued that the PLATO-INVASIVE subpopulation for ticagrelor was the most closely matched to the TRITON-TIMI 38 population in which prasugrel was evaluated. On the other hand, IQWiG used the PLATO-STEMI subgroup. No judgement on the appropriateness of which subgroup should be used was made in this analysis; however, both premises were examined in a cost-effectiveness model for the UK setting. The subgroup analyses conducted in the reconstructed economic model to evaluate ticagrelor vs. prasugrel, *ceteris paribus*, led to irreconcilable cost-effectiveness estimates between the INVASIVE and STEMI (PCI) patient populations. The CEACs for ticagrelor in both populations suggested that the choice of subgroup by NICE may have influenced recommendations for ticagrelor in the UK.

The example explored in this chapter emphasised the importance of contextual factors, such as the interpretability and quality of clinical evidence, they not only pose a challenge to the transferability of REAs across countries but can also influence HTA

outcomes. Chapters 5 and 6 will explore how evidence and methods in REA can also impact results.

4.5.1 Caveats and model limitations

The most important caveat to this example is the controversial ITC of ticagrelor vs. prasugrel based on the pooling of the PLATO and TRITON-TIMI 38 studies. As mentioned in section 4.3.2, a majority of HTA agencies had raised concerns regarding the dissimilar patient populations included in both RCTs. CADTH and PBAC had already omitted prasugrel as a relevant comparator based on the inconsistent indications for ticagrelor and prasugrel in ACS. In addition, I did not perform my own ITC and only had access to the pooled estimates from the IQWiG report with no information as to how the evidence synthesis was conducted or the resulting correlation structure between treatment effects.

Findings from my analysis may not be applicable to the scope of ticagrelor's assessment under the NICE perspective. However, by recreating the cost-effectiveness analysis, I was able to evaluate the impact of a key structural assumption for ticagrelor's economic evaluation that of the choice of relevant patient subgroups and comparators.

Chapter 5

Searching for indirect evidence: NMA of interventions for VTE prevention

5.1 Background

In Chapter 3, I highlighted the conceptual and practical assumptions required to perform NMA including the need for trials to be ‘connected’ or ‘anchored’ by at least one common intervention to form a single network of evidence. For example, when no head-to-head trial is available, studies evaluating A vs. B and B vs. C can be used to compare A and C indirectly, i.e. via treatment B. However, additional intermediate connections may be required to link two treatments of interest in a larger network; thereby increasing the degree of ‘removal’ or ‘separation’ between comparisons and decreasing the degree of influence on the analysis [Hawkins 2009a, Jansen 2011, Caldwell 2015]. When extending an evidence base for NMA, a key methodological concern is how best to identify relevant trials and select treatment comparisons to optimise the network shape and size. Indeed, Caldwell remarks: “*the biggest deviation [of NMA] from a pairwise systematic review is in the definition of treatments in the network*” [2014]. In this chapter, I evaluate the impact of study identification and network size on NMA to compare pharmacological interventions for the prevention of venous thromboembolic events (VTE) following total knee replacement (TKR) surgery. This chapter is largely based on a recent publication by Dequen *et al.* (cf. Appendix F) [2014].

5.1.1 Apixaban for VTE prevention

Venous thrombosis is a blood clot that forms within a blood vessel. Deep vein thrombosis (DVT) refers to a clot occurring in the ‘deep veins’ of the body, most commonly in the legs. If a clot breaks off and travels through the circulatory system (i.e. embolization) lodging itself in the lungs and obstructing blood vessels, it is called a pulmonary embolism (PE). Venous thromboembolism or VTE collectively describe DVT, PE, or a combination of both events. Possible clinical symptoms of thrombosis

include swelling, discoloration, tenderness, and/or pain; however, VTE is often asymptomatic and has been described in the literature as a ‘silent killer’ [Futterman 2004, NICE 2010a]. VTE is a major cause of death in the UK, with an estimated 25,000 people in England dying from preventable hospital-acquired VTE every year [NICE 2010a, House of Commons (HoC) 2005]. Hospitalised patients are particularly at high risk of developing VTE due to inactivity and reduced mobility; around 25% of all VTE cases are attributed to hospitalisation following illness or surgery [Francis 2007, Geerts 2008]. For example, the HoC Health Committee reported that 45-51% of patients undergoing orthopaedic surgery would develop DVT without adequate thromboprophylaxis [HoC 2005]. Although VTE represents a considerable morbidity burden and can be fatal, it is preventable and the use of pharmacological, as well as mechanical, prophylaxis is now common practice in the UK.

In 2010, NICE published a clinical guideline on reducing the risk of VTE in patients admitted to hospital. At the time, five drugs were recommended following elective orthopaedic surgery: dabigatran etexilate, fondaparinux sodium, low molecular weight heparins (LMWH), rivaroxaban, and unfractionated heparin for patients with renal failure [NICE 2010a]. These drugs were evaluated in a series of single technology appraisals (STA) and all shown to be highly cost-effective within their given indication [NICE 2008b, 2009b, 2012a]. The oral anticoagulant apixaban (Eliquis™, Bristol-Myers Squibb (BMS) and Pfizer) was later recommended by NICE, in 2012, to prevent blood clots in adult patients scheduled for total hip or knee replacements [NICE 2012b].

Postoperative apixaban 2.5mg twice daily demonstrated superiority compared to enoxaparin—the most widely used LMWH—in reducing VTE and all-cause death in patients undergoing major joint replacement surgery [Lassen 2008]. The relative safety and efficacy of apixaban vs. alternative VTE prophylaxes were assessed by NICE based on an ITC and MTC submitted by manufacturers during the STA process [BMS and Pfizer 2011]. The interventions of interest defined by the NICE scope and included by the manufacturers in the I/MTC—rivaroxaban, dabigatran, fondaparinux, and LMWHs—formed a connected network ‘anchored’ by the common comparator enoxaparin 40mg once daily. RCTs to inform the network of studies were identified by systematic literature review; however, solely head-to-head comparisons of interventions considered in the NICE scope were included for analysis. Whilst the ERG found that the manufacturers’ approach “*satisfactorily*” addressed the decision problem

posited by NICE, a substantial number of trials were disregarded from the statistical analysis and results from the MTC were overlooked as the network was found to be inconsistent [Riemsma 2011, NICE 2013b].

5.2 Objectives

Based on my findings in Chapter 2, no HTA guidance currently advises on the search methodology or study selection criteria specifically required to target the identification of evidence for NMA. However, the latest NICE *Guide to the methods of technology appraisal* states the following:

“Ideally, the network meta-analysis should contain all treatments that have been identified either as an intervention or as appropriate comparators in the scope. Therefore, trials that compare at least 2 of the relevant (intervention or comparator) treatments should be incorporated even if the trial includes comparators that are not relevant to the decision problem.”

[2013a: p41]

Scientific bodies such as the NICE DSU and the Cochrane Collaboration on Multiple Interventions Methods Group (CMIMG) recommend that the good principles of conducting a standard systematic review and meta-analysis be extended to NMA [Dias 2011, CMIMG 2016]. The ISPOR-AMCP-NPC²¹ Good Practice Task Force has recently produced a two-part report on how to conduct ITC, as well as, a questionnaire to assess its relevance and credibility to inform decision-making [Jansen 2011, Hoaglin 2011, Jansen 2014]. All three publications suggest a ‘good effort’ should be made not only to identify but also to include all available and relevant published RCTs for NMA. Moreover, the Task Force report (part-2) advocates the use of a ‘staged’ search strategy as best practice when performing an NMA [Hoaglin 2011]. This ‘staged’ or iterative search methodology was initially proposed and trialled by Hawkins *et al.* to maximise a network of studies by more efficiently identifying indirect evidence [Hawkins 2009a, 2009b]. Additional connections can provide useful information, but authors warn that if more than a few links separate treatments, results may be unreliable. For example, a larger network may connect interventions otherwise unconnected but may also increase between-study heterogeneity and uncertainty around estimates, as well as, introduce

²¹International Society for Pharmaceutical and Outcomes Research—Academy of Managed Care Pharmacy—National Pharmaceutical Council (ISPOR-AMCP-NPC)

inconsistency between direct and indirect comparisons [Hawkins 2009a, 2009b, Hoaglin 2011, Caldwell 2014].

Adopting the Hawkins *et al.* search strategy, I evaluate the impact of different network sizes on the relative effectiveness and cost-effectiveness of apixaban vs. recommended interventions to prevent VTE in adult patients undergoing elective TKR surgery in the UK. The specific chapter objectives are:

- i. to conduct a breadth-first systematic literature search to identify all relevant indirect evidence on pharmacological thromboprophylaxes for TKR;
- ii. to perform a series of NMA for each network size obtained from the ‘staged’ searches; and
- iii. to evaluate the impact of network size on the cost-effectiveness of VTE interventions following TKR from an NHS/PSS perspective.

5.3 Systematic literature review

5.3.1 Methods

Breadth-first searching is based on graph theory; it is an uninformed or ‘naïve’ search process which aims to exhaustively search a sequence or combination of sequences from a ‘root’ node on a graph, to all ‘neighbouring’ nodes without considering a final limit until it is reached [Hawkins 2009a]. A parallel can be drawn between nodes on a graph to interventions on a network map and the need to identify all ‘links’ (i.e. comparisons) without knowing the final size or shape of the network. Caldwell [2014] divides interventions within a network as “*decision*” and “*supplementary*” sets, whereby *decision* interventions describe the subset of treatments of most interest for the systematic literature review and *supplementary* interventions are included to provide additional evidence on the *decision* set [Caldwell 2014]. In HTA practice, a *decision* set of interventions should be defined in the research question, scope and search protocol. These interventions and/or comparators of interest are often used to restrict the selection of studies for NMA. However, a breadth-first search allows investigators to identify all the ‘nodes’ and ‘links’ forming a network of evidence without pre-specifying a *decision* set or even knowing the extent of the *supplementary* set of interventions available in the literature.

Hawkins *et al.* refer to search ‘orders’ and associated search comparators to describe each sequential step in the breadth-first search [2009a]. A generic description of the breadth-first search strategy is summarised in Table 19.

Table 19 Breadth-first search strategy

Search order	Search iteration	Search comparators
1	i	All first order comparators <i>except one</i>
	ii	First order comparator previously omitted
2	iii	All second order comparators <i>except one</i>
	iv	Second order comparator previously omitted
3	v	All third order comparators <i>except one</i>
	vi	Third order comparator previously omitted

Adapted from Table 1 of Hawkins et al. [2009a]

Treatments directly compared to first order comparators following first order searches become second order comparators, and so on. The sequence of searches in Table 19 progressively includes first, second and third order comparators to identify all RCTs contributing to a network of evidence, until no further comparators are identified. From the set of identifiable trials, all relevant indirect comparisons are also identified at any given order.

In accordance with Hawkins *et al.* searches were divided further for each order [2009a]. In Table 19, search orders are numbered 1-3 and searches within each order i-vi. For example, in the first order searches, all but one first order comparator are included in the search terms (cf. search 1(i) in Table 19). The omitted comparator is searched separately in a subsequent search iteration to ensure all trials including one or more first order comparators are captured and all possible second order comparators identified (cf. search 1(ii) in Table 19). Search (1i) will identify all trials comparing more than one of the first order treatments, thus identifying any direct head-to-head evidence, albeit one of the treatments is not included in the search syntax. If the objective is to capture only first-order (i.e. direct) comparisons, the subsequent search (1ii) of the omitted comparator is not required. In this instance, dividing the search in two steps has the potential to reduce the search burden if a particular comparator is associated with a large number of hits. Hawkins *et al.* thus recommend omitting a widely-used comparator such as placebo or best supportive care. If further search orders are conducted and abstracts reviewed, search (1ii) is redundant and all order comparators

could be searched at once. First order comparators can be arbitrarily selected within or outside the original decision problem and include treatments not of interest for appraisal. Moreover, study selection is intentionally broadened to include all RCTs evaluating a first order comparator without a restriction on comparator criteria, allowing for treatments which may not fall within the scope for appraisal or the ‘decision set’, such as unlicensed drugs, non-relevant treatments for decision-making or non-pharmacological interventions, to contribute to the network of evidence.

The clinical evidence review for apixaban and original NMA—as submitted by manufacturers for NICE TA245—was based on a systematic literature search of RCTs evaluating VTE prophylactic interventions following TKR [BMS and Pfizer 2011, NICE 2012b]. I adapted the final reported clinical effectiveness search strategy reworked by the ERG to identify relevant RCTs for NMA using Hawkins *et al.*’s breadth-first search methodology [Hawkins 2009a, Riemsma 2011]. In October 2012, I performed a stepwise search including three search orders and six search iterations in Medline®, Medline-in-Process®, OLDMedline®, EMBASE, and the Cochrane Library. Table 20 lists the search comparators in VTE prevention for each search order and search iteration; first order search comparators were extracted from the original industry submission search strategy.

Table 20 Breadth-first search strategy for VTE prevention

Search order	Search iteration	Search comparators
1	i	antixarin, apixaban, ardeparin, bemiparin, calciparine, certoparin, dabigatran, dalteparin, deligoparin, enoxaparin, fondaparinux, fraxiparine, heparin/LWMH, idraparinux, livaraparin-calcium, lomorin, minidaltin, monoparin, nadroparin, parnaparin, parvoparin, reviparin, sandoparin, seleparin, semuloparin, tafoxiparin, tedegliparin, tedelparin, tinzaparin
	ii	rivaroxaban
2	iii	acenocoumarol, ancrod, aspirin, acetylsalicylic acid, ave5026, betrixaban, continuous enhanced circulation therapy, desirudin, dextran, dihydroergotamine, edoxaban, foot pump, garment, graduated compression stocking, hirudin, indomethacin, intermittent/pneumatic compression, lomoparin, LY517717, melagatran, rosuvastatin, synchronisation technology, TAK442, TB402, warfarin, ximelagatran
	iv	placebo

Search order	Search iteration	Search comparators
3	v	antistenocardin, danaparoid, dipyridamole, foot system, inflation/equential/pneumatic/ plantar/intermittent compression, lidocaine, methylprednisolone, orgaran, persantin, tocainide, triflusal
	vi	steroid

Table 21 and Table 22 present both iterations of the first order search strategy run in Ovid for all Medline® and EMBASE online resources²². In order to replicate the original search conditions and provide comparable results to those reported by the manufacturers and ERG, searches were further restricted by date to studies published prior to September 2011. The second and third order search strategies are included in the Appendix C in Tables C1-4.

Table 21 First order search strategy without rivaroxaban (cf. Table 19 search 1i)

#	Search terms	Hits
1	exp Thromboembolism/	345649
2	exp Pulmonary Embolism/	83761
3	exp Venous Thrombosis/	123044
4	((venous or vein) adj (thrombosis or thromboses or thrombus or thromboembolism)).mp.	132049
5	(dvt or vte).mp.	22959
6	((pulmonary or lung) adj6 (embolism or emboli)).mp.	100038
7	thrombophlebitis.mp.	42082
8	or/1-7	435855
9	(fondaparin* or arixtra or ic851589 or org31540 or quixidar or sr90107*).mp.	5592
	(rivaroxaban or bay597939).mp.	
10	(dabigatran or rendix or pradaxa or bibr1048).mp.	3369
11	(apixaban or eliquis or bms562247).mp.	1536
12	exp Heparin/ or exp Heparin, Low-Molecular-Weight/ or (LMWH or low molecular weight heparin).mp.	187499
13	(dalteparin or fragmin* or k2165).mp.	7007
14	(enoxaparin or clexane or klexane or lovenox or pk10169).mp.	17548
15	(nadroparin or fraxiparin* or fraxodi or seleparine or tedegliparin or cv216).mp.	4360
16	(ardeparin or normiflo or wy90493).mp.	366

²² Ovid Medline® In-Process & Other Non-Indexed Citations and Ovid Medline® 1946 to Present; Ovid OLDMedline® 1946 to 1965; EMBASE 1974 to 2012 August 27.

#	Search terms	Hits
17	(tinzaparin or innohep or logiparin of lhn1).mp.	2612
18	(certoparin or sandoparin or emborex or monoemborex).mp.	852
19	(parnaparin or fluxum or lohepa or minidaltan or parvoparin or op2123).mp.	321
20	(reviparin or cilvarin* or lomorin or lu47311).mp.	979
21	tedelparin.mp.	53
22	(calciparine or monoparin or bemiparin or hibor or phivor).mp.	686
23	(livaraparin-calcium or tafoxiparin or idrabiotaparinux or rd-11885 or idraparinux or semuloparin or cy-222 or deligoparin or antixarin).mp.	1088
24	or/9-23	191079
25	Randomized controlled trials as Topic/	102101
26	Randomized controlled trial/	665396
27	Random allocation/	134748
28	Double blind method/	229566
29	Single blind method/	32898
30	Clinical trial/	1347882
31	exp Clinical Trials as Topic/	299876
32	or/25-31	1842964
33	(clinic\$ adj trial\$1).tw.	424839
34	((singl\$ or doubl\$ or treb\$ or tripl\$) adj (blind\$3 or mask\$3)).tw.	269088
35	Placebos/	247696
36	Placebo\$.tw.	327596
37	Randomly allocated.tw.	32529
38	(allocated adj2 random).tw.	1482
39	or/33-38	932608
40	32 or 39	2202449
41	Case report.tw.	423345
42	Letter/	1545999
43	Historical article/	285798
44	Review of reported cases.pt.	0
45	Review, multicase.pt.	0
46	or/41-45	2240515
47	40 not 46	2144517
48	Meta-Analysis as Topic/	16845
49	meta analy\$.tw.	103143
50	metaanaly\$.tw.	4088
51	Meta-Analysis/	101122
52	(systematic adj (review\$1 or overview\$1)).tw.	83565
53	exp Review Literature as Topic/	50800
54	or/48-53	257434
55	cochrane.ab.	49961

#	Search terms	Hits
56	embase.ab.	44232
57	(psychlit or psyclit).ab.	1816
58	(psychinfo or psycinfo).ab.	13555
59	(cinahl or cinhal).ab.	15798
60	science citation index.ab.	3475
61	bids.ab.	745
62	cancerlit.ab.	1217
63	or/55-62	78177
64	reference list\$.ab.	16731
65	bibliograph\$.ab.	23853
66	hand-search\$.ab.	7127
67	relevant journals.ab.	1291
68	manual search\$.ab.	4114
69	or/64-68	47727
70	selection criteria.ab.	35862
71	data extraction.ab.	18440
72	70 or 71	51625
73	Review/	3653382
74	72 and 73	32593
75	Comment/	514915
76	Letter/	1545999
77	Editorial/	765479
78	animal/	6825372
79	human/	26193069
80	78 not (78 and 79)	5024451
81	or/75-77,80	7386916
82	54 or 63 or 69 or 74	313357
83	82 not 81	295416
84	47 or 83	2326847
85	Orthopedics/	32293
86	arthroplasty, replacement, hip/ or arthroplasty, replacement, knee/	42036
87	((hip or knee or femoral head) and (replac\$ or arthroplast\$ or prosthesis\$ or surgery or surgical or implant\$)).mp.	191424
88	or/85-87	221271
89	8 and 24 and 84 and 88	2068
90	limit 89 to yr="2012 -Current"	104
91	89 not 90	1964
92	Remove duplicates from 91	1422

Adapted from Appendix 1A Clinical effectiveness search reworked by ERG to maximise results Medline® (OvidSP): 1948 to August Week 5 2011 [Riemsma 2011]

Table 22 First order search strategy rivaroxaban only (cf. Table 19 search 1ii)

#	Search terms	Hits
93	(rivaroxaban or bay597939).mp.	2,923
94	8 and 93 and 84 and 88	459
95	limit 94 to yr="2012 -Current"	37
96	94 not 95	422
97	remove duplicates from 96	340
98	97 not 92	3

Studies were selected in two stages according to a pre-defined set of inclusion and exclusion criteria taken from the TA245 manufacturer submission and limited to the English language [Riemsma 2011]. First, abstracts were screened and in a second instance, full-text publications were retrieved and reviewed to meet the population, outcomes, and study design criteria listed in Table C5 in the Appendix. As aforementioned, a comparator criterion was not defined so as to not limit the identification of search comparators in the breadth-first search. The study selection process was repeated for each search iteration until no additional comparators were identified.

5.3.2 Results

In total, 53 RCTs met the inclusion criteria across three network orders. Figure 5 shows the study selection flow diagram broken down by search and network order. The numbers of studies included and excluded for each search iteration are also presented in Figure 5 and totalled by network order. Trials enrolling patients who had undergone either total hip or knee replacement surgery were included if results were reported for the TKR population separately. Figure C1 in the Appendix illustrates in different colours the network map representing all treatment comparisons identified by the successive search orders.

The number of RCTs included in the NMA was limited to focus solely on treatment comparisons that would inform the relative effectiveness estimates for apixaban vs. relevant comparators for decision-making (i.e. dabigatran etexilate 220mg/qd, enoxaparin 40mg/qd, rivaroxaban 10mg/qd). Graphically, these comparisons are referred to as ‘closed loops’ within the network of studies. Focusing on these loops reduced the size of the evidence base and made datasets more manageable without

biasing results, since excluded studies did not contribute to indirect comparisons relevant to the decision space.

Figure 6 illustrates the network diagrams for each search order including only the ‘closed loops’²³. The base case network for the ITC used by the manufacturers to inform the economic model is provided in Figure 6a, for reference. Not all studies reported outcomes of interest and were *de facto* excluded from the NMA. The final numbers of studies in each NMA order for TKR are included in Figure 5 and presented in tabular format in the Appendix (cf. Table C6). Lastly, interventions from 3-arm trials were included even if only one treatment comparison from the trial was of interest, such as in Wang *et al.* comparing placebo, fraxiparine (nadroparin calcium) 0.2-0.4mL/qd and indomethacin 25mg/bd [2004].

²³Asterisks in Figure 5 indicate one node representing multiple drug dosages; although different dosages were considered as individual treatments in the analyses they were not illustrated in the networks for readability.

Figure 5 Study selection flow diagram

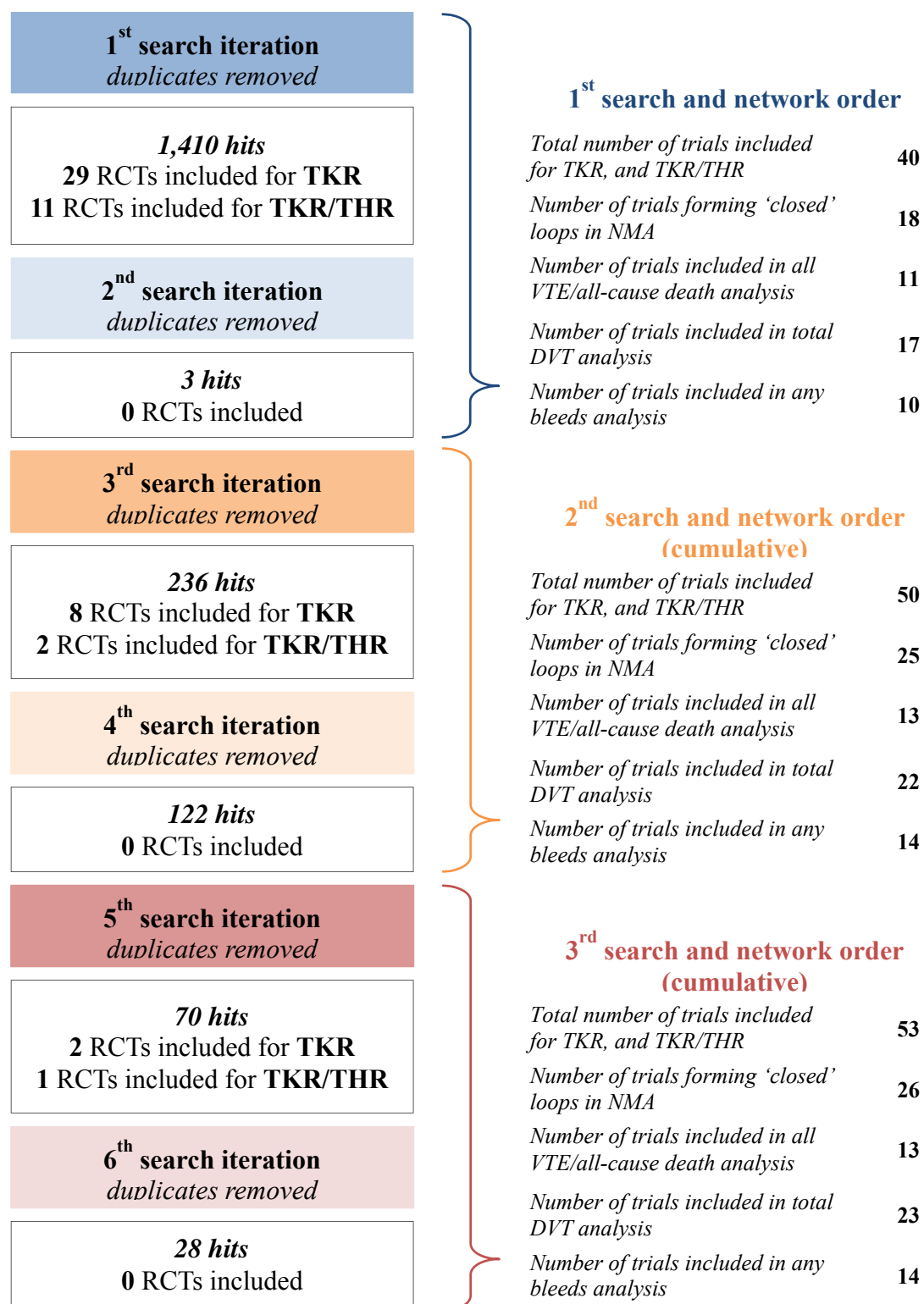
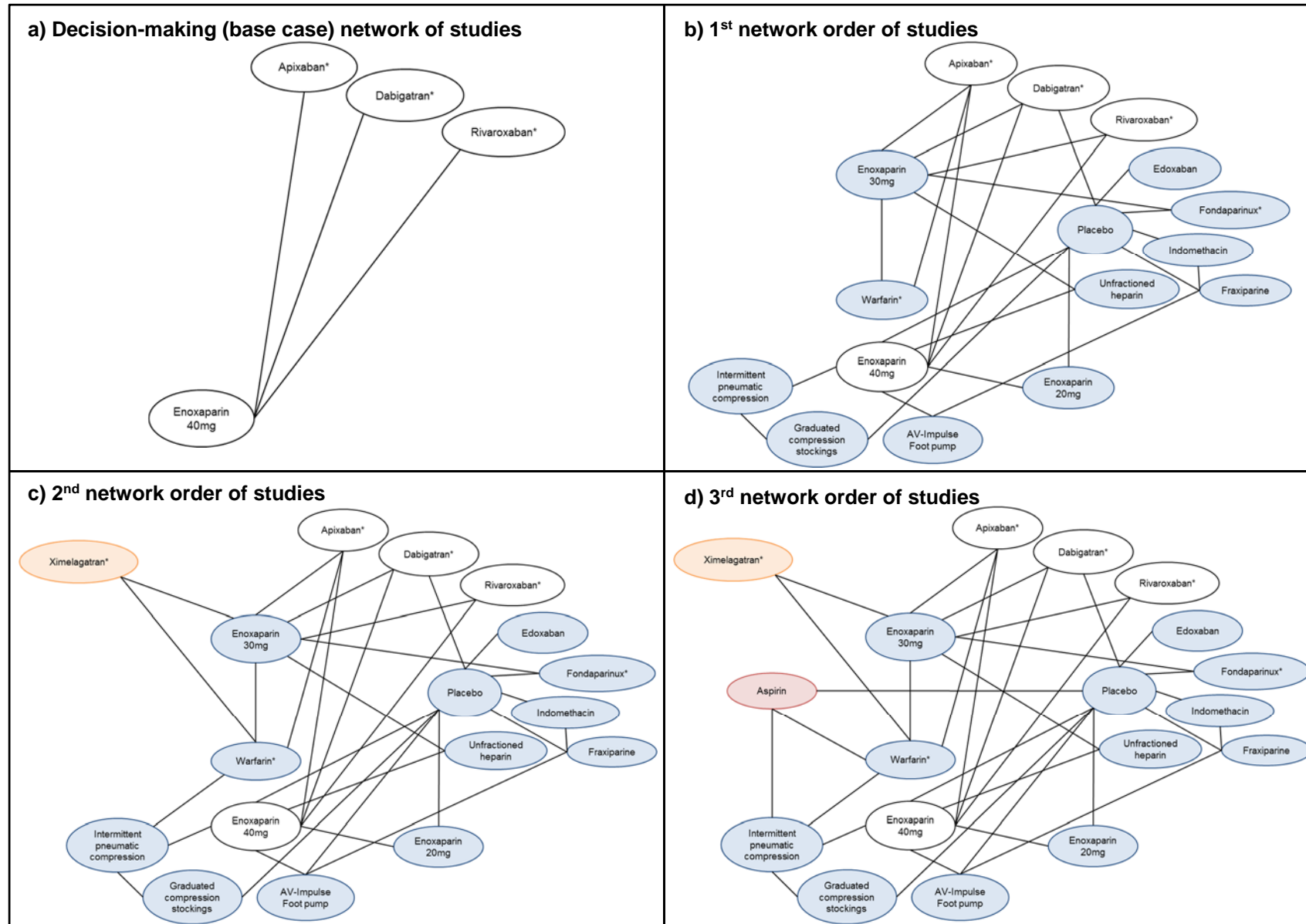


Figure 6 Network of studies including only ‘closed loops’ based on search/network orders (* indicates multiple dosages)



5.4 Network meta-analysis

5.4.1 Methods

A Bayesian NMA was conducted for each network order for the composite outcome of total VTE and all-cause death, as well as for total DVT, and any bleeds. Network sizes were based on the studies selected following each search order (i.e. first, second, and third network orders). Multiple outcomes were analysed for economic modelling purposes and in order to curb potential outcome reporting bias for the composite measure of all VTE/all-cause death used in more recent trials as primary outcome measure but not frequently calculated in older studies [Eriksson 2007, Lassen 2012, Lassen 2008]. Fixed and random effects NMA models adjusted for multi-arm trials—taken from Ades *et al.*—were used in WinBUGS version 1.4.3 to estimate ORs for each outcome of interest [Lunn 2000, Dias 2011]. The WinBUGS code for the third network order random effects NMA model, adjusted for multi-arm trials, is provided in Figure A6.2 in the Appendix including the extracted data from included study publications for each outcome.

Model fit was evaluated using the total residual deviance and the deviance information criterion (DIC) for each network order [Spiegelhalter 2002]. Between-study heterogeneity was compared using the standard deviation across random effects models [Spiegelhalter 2003]. Inconsistency was assessed by plotting the residual deviances against the number of intervention arms in each included study, and looking at the proportion of mixed p-values under 5% and 10% significance [Marshall 2003, Caldwell 2010]. If there was no inconsistency, the residual deviance would equal the number of arms in each trial because it should be equal to one for each data point. Mixed p-values provide an approximation to cross-validation p-values, which can be calculated in a single model run. According to Welton *et al.*, mixed p-values calculated from the same dataset should follow a Uniform distribution on the interval (0,1) [Welton 2012]. I plotted the ordered p-values for each study and each network order against Uniform order statistics to evaluate inconsistency looking at unusually small or large p-values [Welton 2012].

5.4.2 Results

Forest plots in Figure 7 summarise the mean ORs and 95% credible intervals (CrI) for all VTE/all-cause death, total DVT, and any bleeds obtained for the base case ITC and three different network order NMAs.

Given the numbers of studies included (cf. Figure 5), second and third order NMAs for all VTE/all-cause death and any bleeds were very similar and results in Figure 7 are only presented for completeness. The growing evidence from base case to first network order marginally increased precision (i.e. decreased width of credible intervals) around the mean ORs for all outcomes. For example, the all VTE/all-cause death mean OR for dabigatran vs. enoxaparin decreased from 0.95 (95% CrI 0.74;1.22) to 0.90 (0.73;1.10) between base case to first order analysis; similarly the uncertainty in the any bleeds mean OR for apixaban vs. enoxaparin was reduced from 0.78 (0.51;1.26) to 0.72 (0.55;0.97). Apixaban and rivaroxaban were superior to enoxaparin 40mg for both efficacy outcomes; however, ORs for dabigatran vs. enoxaparin were inconclusive. Results favoured apixaban over dabigatran for all VTE/all-cause death for all network orders with the same mean OR of 0.65 (0.51;0.85) for first and second order analyses. The NMA also estimated that patients are less likely to experience a VTE event/death with rivaroxaban compared to apixaban at higher network orders, although the base case ITC did not support the statistical superiority of rivaroxaban and this was not demonstrated for total DVT. Apixaban showed the most favourable safety profile vs. enoxaparin and vs. rivaroxaban for first and second order NMA.

For each model, the first 20,000 iterations were discarded as a burn-in and achieved reasonable convergence according to visual inspection of trace and history plots. The main analyses were based on a further 50,000 iterations to ensure the robustness of results. Fixed effects models for all network orders and all outcomes were used as they provided the best fit to the data according to the DIC. Goodness-of-fit statistics for the fixed and random effects NMA models are presented in Table 23; both models were thought to provide a similar fit to the data.

Figure 7 Odds ratios for all VTE/all-cause death, total DVT, and any bleeds from fixed effects NMA models

— base case — 1st order — 2nd order — 3rd order

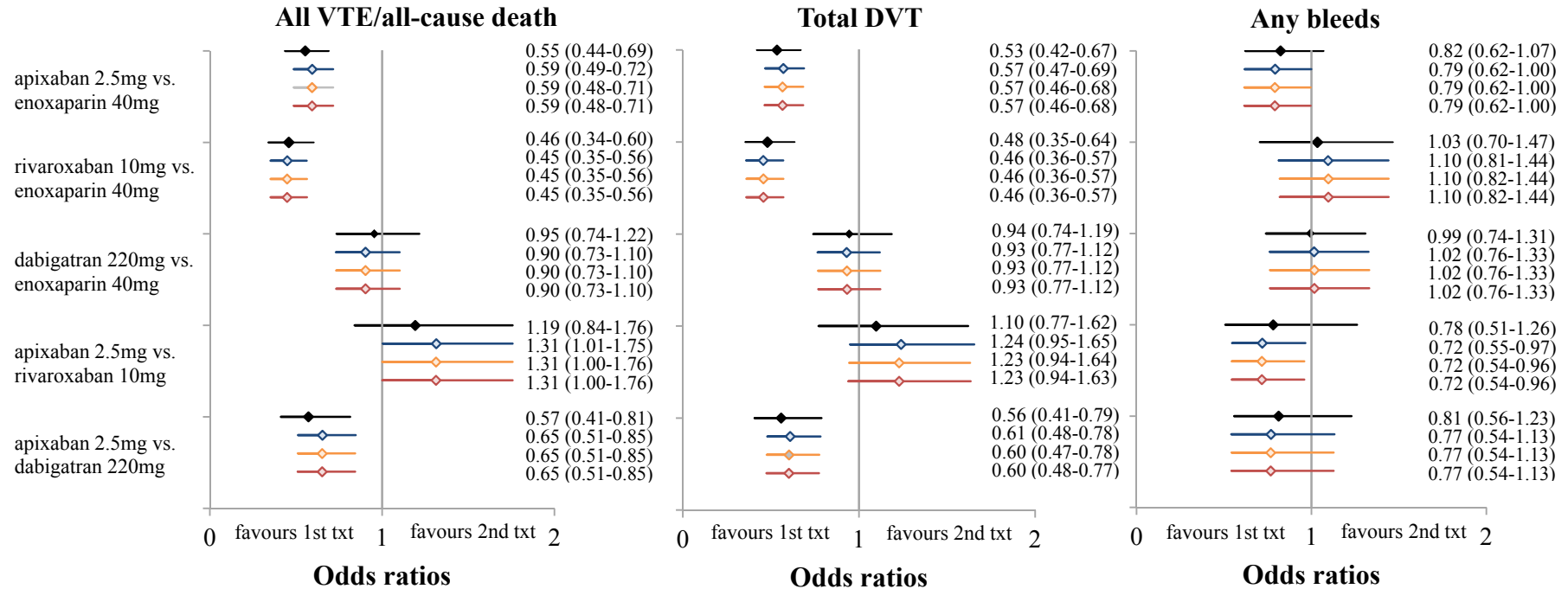


Table 23 Goodness-of-fit statistics for fixed and random effects NMA models by outcome and by network order

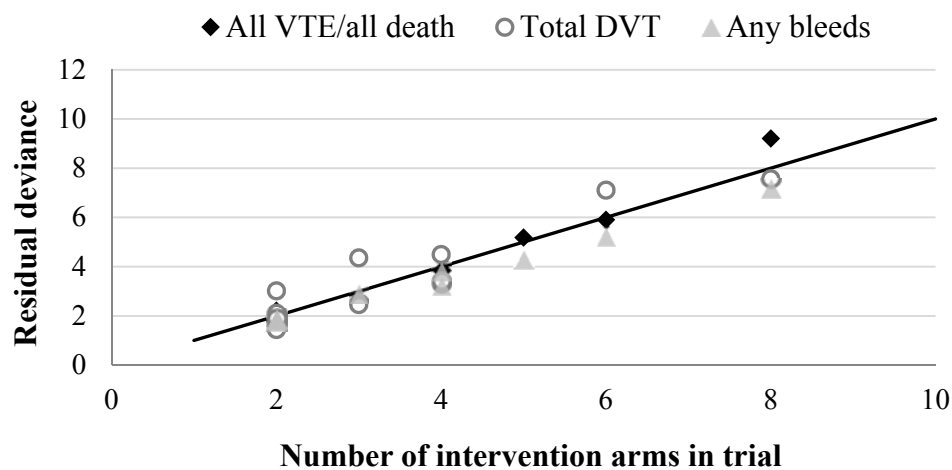
	Fixed effects		Random effects		
	DIC	Total residual deviance	DIC	Total residual deviance	Standard deviations (95% CrI)
All VTE/all-cause death					
1 st order	260.97	39.92	262.27	39.33	0.156 (0.005 - 0.588)
2 nd order	303.14	44.23	304.45	43.95	0.108 (0.004 - 0.379)
3 rd order	n/a	n/a	n/a	n/a	n/a
Total DVT					
1 st order	366.15	52.48	369.14	52.96	0.092 (0.002 - 0.307)
2 nd order	468.65	70.59	471.05	69.45	0.112 (0.006 - 0.341)
3 rd order	490.00	80.1	492.11	77.98	0.138 (0.015 - 0.391)
Any bleeds					
1 st order	237.87	33.46	239.46	34.12	0.115 (0.003-0.569)
2 nd order	303.89	42.29	305.36	42.86	0.108 (0.004 - 0.350)
3 rd order	n/a	n/a	n/a	n/a	n/a

Although the fixed effects provided the best model fit for all outcomes and all network orders; the random effects models were used to assess between-study heterogeneity and consistency of the evidence. Results for the random effects models are included in the Appendix (cf. Figure C3). Overall, results were consistent across all network orders for both fixed and random effects models with little variation between respective point estimates and credible intervals. The between-study heterogeneity estimates and credible intervals were reduced for all VTE/all-cause death from 0.156 (0.005 - 0.588) to 0.108 (0.004 - 0.379) and from 0.115 (0.003-0.569) to 0.108 (0.004 - 0.350) for any bleeds, from first to second order NMA. The standard deviations increased, but not considerably, from 0.092 (0.002 - 0.307), to 0.112 (0.006 - 0.341), and 0.138 (0.015 - 0.391) as the network of studies grew across all three total DVT networks. Spiegelhalter *et al.* provide a possible interpretation of the random-effects standard deviation by describing a ‘range’ of ORs [2003]. This range is in fact the ratio of the 97.5% to the 2.5% point of the distribution of ORs for any given relative treatment effect. They state that standard deviations on the OR scale of 0.1 or 0.2 will only ever correspond to a range of ORs of 1.48 or 2.19, respectively [Spiegelhalter 2003]. Therefore, the standard deviations reported in Table 23, all smaller than 0.2, showed little evidence of between-study heterogeneity further justifying the use of a fixed effect model.

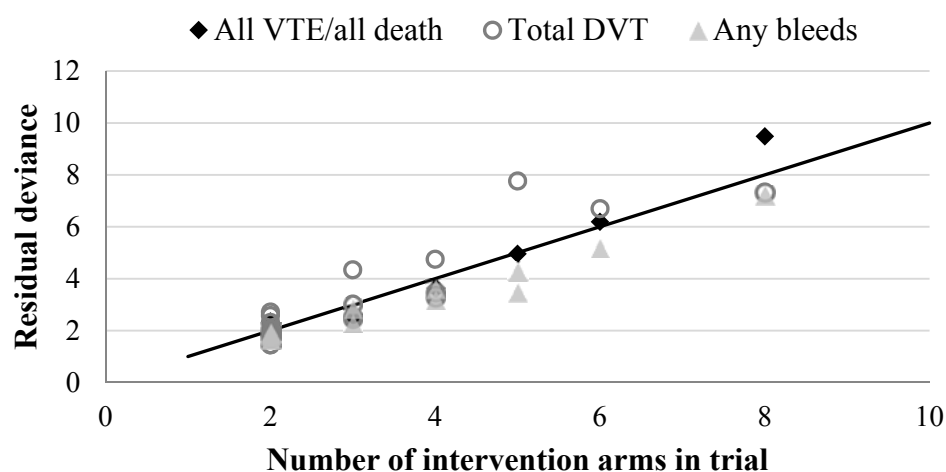
Investigatory plots of residual deviances against number of intervention arms for each trial, outcome, and network order, as shown in Figure 8, do not suggest any inconsistencies between direct and indirect evidence across all models. Additional plots of the ordered mixed predicted p-values against Uniform order statistics suggest the evidence is consistent across the three outcomes and network orders (cf. Figure C4 in the Appendix C). Although the plotted mixed p-values appear to deviate slightly from a Uniform distribution, no individual p-value was significant as 5% or, more appropriately, at 10% due to the estimates being conservative by nature [Welton 2012]

Figure 8 First, second and third order NMA inconsistency plots

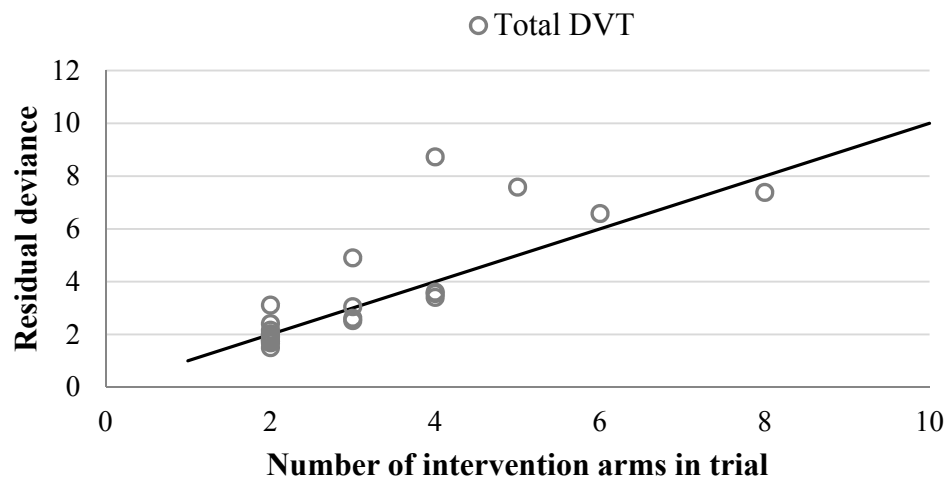
a) 1st network order



b) 2nd network order



c) 3rd network order



I did not find an in-depth exploration of heterogeneity and inconsistency (e.g. node-splitting) across order NMAs was warranted given results and therefore it was not performed. Lastly, the analysis of both efficacy outcomes—i.e. all VTE/all-cause death and total DVT—showed little variation largely due to the relatively low risks of PE (fatal and non-fatal) and death amongst surgical patients suggesting no outcome reporting bias for composite measures in the VTE literature.

The results from all the network orders were used as clinical input parameters to populate the cost-effectiveness analysis.

5.5 Cost-effectiveness analysis

5.5.1 Methods

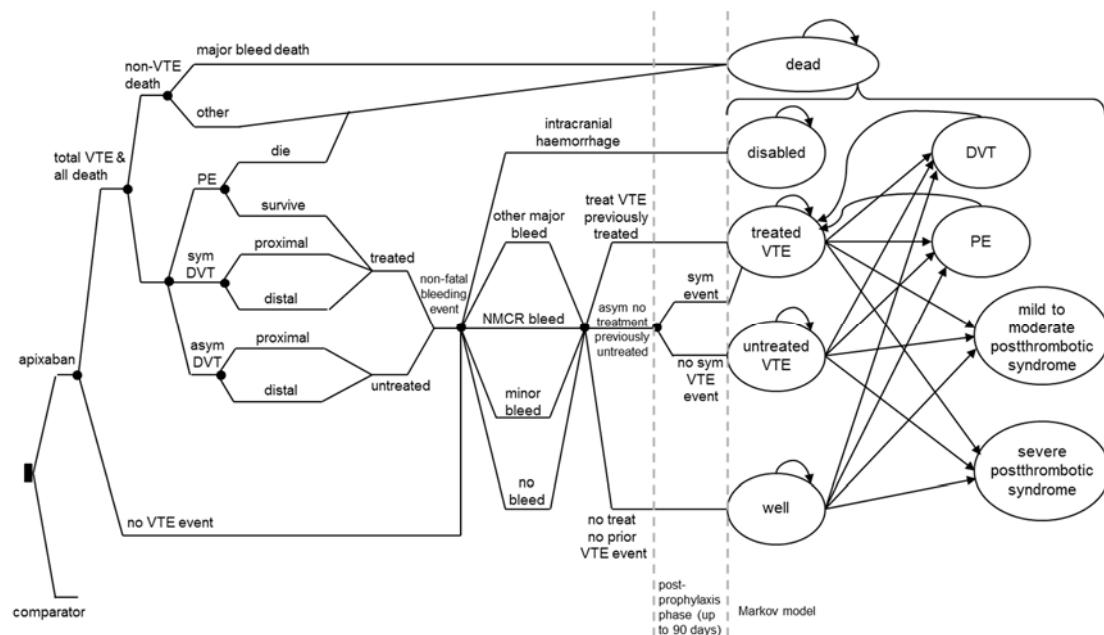
The base case was defined a priori in the apixaban appraisal from three Phase III clinical trials comparing apixaban 2.5mg/bd, dabigatran etexilate 220mg/qd, and rivaroxaban 10mg/qd to enoxaparin 40mg/qd, respectively [Eriksson 2007, Lassen 2012, Lassen 2008]. In accordance with the manufacturers' submitted economic model for apixaban, these interventions form the decision space for VTE prevention following TKR and are routinely used in clinical practice in the UK [BMS and Pfizer 2011]. A comparison with fondaparinux was not considered relevant by manufacturers or the ERG due to its low market share in the UK and was therefore excluded from the cost-effectiveness analysis.

5.5.1.1 Model structure

A combined decision-tree and Markov model was built in Excel 2010 to model an initial prophylaxis phase following TKR and a lifetime horizon thereafter. The decision-tree modelled post-surgery clinical outcomes and costs over 90-days; whilst 35 yearly Markov cycles were considered to span the remainder of patients' lifetime. The economic model was rebuilt using the *de novo* model structure, assumptions, and input data published in the apixaban manufacturer submission and ERG report [BMS and Pfizer 2011, Reimsma 2011]. Figure 9 illustrates the two-phase model diagram.

The model structure reflects the clinical pathways of care for patients undergoing TKR in the UK and was in part informed from the outcomes collected from the pivotal ADVANCE -2 trial for apixaban 2.5mg [Lassen 2010]. The clinical pathways in the decision-tree and the health states in the Markov model were designed to capture all the risks of VTE (i.e. PE, symptomatic DTV, asymptomatic DVT), benefits of prophylaxis, adverse events (i.e. bleeds, intracranial haemorrhage), and associated costs to the NHS.

Figure 9 Economic model structure diagram



asym: asymptomatic; NMCR: non-major clinically relevant; sym: symptomatic
Adapted from Figure 8 and 9 from apixaban manufacturer submission [BMS and Pfizer 2011]

5.5.1.2 Clinical model parameters

Treatment effects were only demonstrated in the first 90 days of the clinical pathway for all VTE and all death, and non-fatal bleeding event. The ORs for all VTE/all-cause death and any bleeds from the base case, first order, and second order NMAs were used to adjust a baseline risk and inform transition probabilities in the decision-tree. Pooled results for total DVT were not directly modelled in the decision-tree. Since additional studies were only identified for this outcome in the third order searches; an economic analysis based on the ORs from the third order NMA was redundant. Baseline risks were extracted from the control arm of the ADVANCE-2 trial for enoxaparin 40mg [Lassen 2010].

In accordance with the manufacturers' economic model, all the remaining transition probabilities in the decision tree were assumed to be treatment independent and were the same for the baseline treatment (i.e. enoxaparin), apixaban, rivaroxaban, and dabigatran [BMS and Pfizer 2011]. Where possible, these conditional post-event probabilities for all VTE and bleeding events were synthesised from published new oral anticoagulant trials. When data was not available, probabilities were taken from both arms of the ADVANCE-2 trial [Lassen 2010]. Details about how these transition probabilities were derived are provided in the apixaban manufacturer submission [BMS and Pfizer 2011].

The decision tree gave rise to 33 different clinical pathways—informed by the conditional post-event probabilities aforementioned—feeding into five starting Markov health states: well, treated VTE, untreated VTE, disabled, and dead. Transition probabilities for the Markov model were also assumed to be treatment independent; however, the probabilities for recurrent DVT, PE, mild/moderate post-thrombotic syndrome (PTS) and severe PTS were time-dependent. All clinical input parameters used in the cost-effectiveness analysis were taken from the TA245 manufacturer submission [BMS and Pfizer 2011 (Tables 61-69)].

Uncertainty around parameters was expressed in distributions; a PSA was performed using 1000 model runs sampling from these distributions. ORs for all VTE/all-cause death and any bleeds were sampled from 10,000 MCMC simulations extracted from WinBUGS output for each NMA order which maintained the correlation structure [Welton 2012]. QALYs were used to estimate ICERs compared to enoxaparin

40mg/qd; the 2.5th and 97.5th percentiles were also extracted to demonstrate the variation in uncertainty around mean ICER estimates at each given order.

5.5.1.3 Utility valuation

A VTE or bleed can significantly impact a patient's quality of life both in the short-term and long-term, for example if an event leads to a PE or disability. HRQoL data was not collected alongside the ADVANCE-2 trial, therefore utility measures from the literature were used to populate the economic model. EQ-5D UK population norms were used to represent the HRQoL of a fully recovered patient following surgery, i.e. Markov state 'well' [Kind 1998]. Each event was associated with a utility decrement; the subsequent utilities representing a 'worsened' HRQoL were then assumed to be constant for the duration of the event. A detailed account of the quality of life literature searches and the mapping exercise undertaken to obtain the utilities and utility decrements for the prophylaxis, post-prophylaxis, and long-term Markov phase included in the economic model were provided in the apixaban manufacturer submission [BMS and Pfizer 2011 (Tables 71-74)]. A discount rate of 3.5% was applied yearly to QALYs.

5.5.1.4 Resource use and costs

Drug acquisition costs—summarised in Table 24—were applied based on assumed treatment durations for each intervention. Administration costs were considered both during inpatient stay and following discharge; however, these were assumed to be the same for all drugs and thus excluded from the model. Additional monitoring costs post-discharge were required for enoxaparin, as this treatment course required subcutaneous administrations at home by a community nurse or a one-off training to self-inject, and regular blood tests not common to other interventions [BMS and Pfizer 2011].

Table 24 Drug acquisition costs

Drug	Dose	Cost per day	Days of TKR treatment	Total drug cost per TKR treatment course	Additional administration costs per TKR treatment course
Enoxaparin	40mg/od	£4.04	12	£48.48	£46.32
Apixaban	2.5mg/bid	£3.43	12	£52.97	-
Rivaroxaban	10mg/od	£4.41	8	£33.60	-
Dabigatran	220mg/od	£4.20	12	£41.16	-

Taken from Table 77 of the TA245 manufacturer submission [BMS and Pfizer 2011]

BMS and Pfizer used 2008/09 NHS reference costs in their submission for all clinical events in the decision-tree, Markov health states, as well as included adverse event with associated costs [2011]. The same costs were included in my replicated model to allow for comparison with the published base case using the ITC network with the four interventions of interest. An itemised list of all included costs is presented in Tables 75 to 79 of the original manufacturer submission [BMS and Pfizer 2011]. A half-cycle correction and discount rate of 3.5% were applied yearly to costs.

5.5.2 Results

Apixaban, dabigatran etexilate, and rivaroxaban were found to be cost-effective vs. enoxaparin 40mg for all network orders. These results were in line with findings from the NICE appraisals that recommended these treatments based on their dominance over enoxaparin 40mg [NICE 2008b, 2009, and 2012]. Table 25 presents the PSA means for total costs, total QALYs and ICERs for the base case, first and second network orders. As previously stated, second and third order NMA results for all VTE/all-cause death and any bleeds were the same, CEA results for the third network order were redundant and not included in Table 25. The Markov traces for the model are listed in the Appendix Tables C7 to C10 for reference.

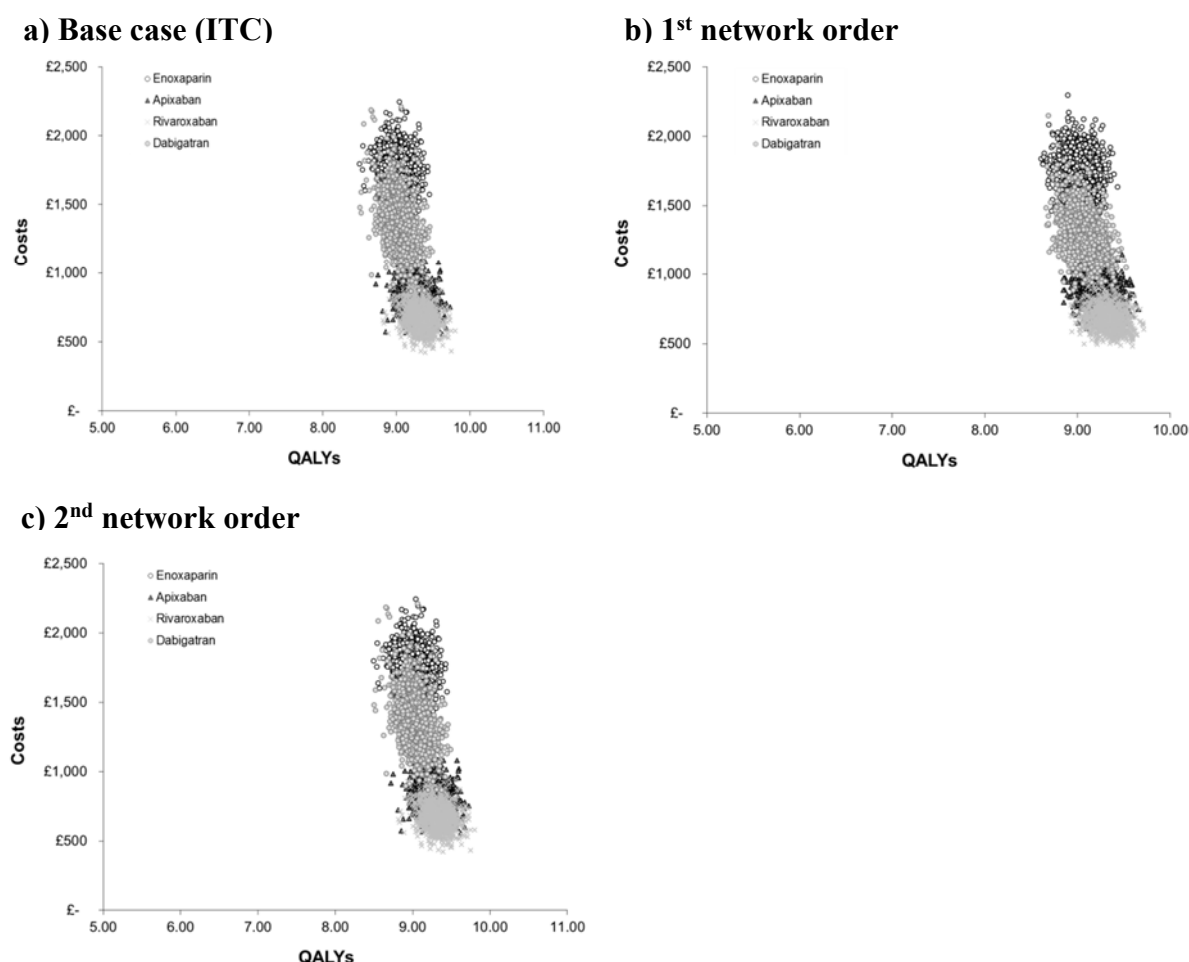
The mean ICERs for rivaroxaban, apixaban, and dabigatran etexilate were dominating vs. enoxaparin 40mg, respectively, across all models suggesting treatments were on average both more effective and less costly than the current standard of care. The cost-utility analysis results showed little variation in outcomes despite the growing evidence base for the NMA parameterising the economic model. Figure 10 shows the cost-effectiveness planes based on the PSA results. At face value, these plots appear uninformative with regards to the impact of network size on the economic evaluation of compared pharmacological treatments for VTE. However, the percentages in Table 25 indicate a reduction in the uncertainty for which treatment is most cost-effective at a willingness to pay threshold of £20,000 from base case to first network order, with rivaroxaban's predicted percentages increasing from 83.2% to 97.1% cost-effective.

Table 25 Cost-effectiveness results for apixaban, dabigatran etexilate, and rivaroxaban vs. enoxaparin 40mg for the base case (ITC), first and second network orders

Base case (ITC)	Total costs	Total QALYs	ICERs	% cost-effective at £20,000	% cost-effective at £30,000
Enoxaparin 40mg	£ 1,746	9.02		0%	0%
Dabigatran etexilate	£ 1,377	9.04	<i>Dominated</i>	0%	0%
Apixaban	£ 810	9.27	<i>Dominated</i>	16.8%	16.7%
Rivaroxaban	£ 703	9.32	<i>Dominant</i>	83.2%	83.3%
First order	Total costs	Total QALYs	ICERs	% cost-effective at £20,000	% cost-effective at £30,000
Enoxaparin 40mg	£ 1,748	9.03		0%	0%
Dabigatran etexilate	£ 1,275	9.09	<i>Dominated</i>	0%	0%
Apixaban	£ 860	9.26	<i>Dominated</i>	2.9%	2.9%
Rivaroxaban	£ 688	9.34	<i>Dominant</i>	97.1%	97.1%
Second order	Total costs	Total QALYs	ICERs	% cost-effective at £20,000	% cost-effective at £30,000
Enoxaparin 40mg	£ 1,754	9.02		0%	0%
Dabigatran etexilate	£ 1,293	9.08	<i>Dominated</i>	0%	0%
Apixaban	£ 868	9.25	<i>Dominated</i>	3.7%	3.7%
Rivaroxaban	£ 695	9.33	<i>Dominant</i>	96.3%	96.3%

Third network orders for included model inputs are the same as second network orders so model results are not presented above.

Figure 10 Cost–effectiveness planes by network order



5.6 Discussion

A network of evidence can take many shapes and vary considerably in size and complexity based on the number of included interventions, the number of included studies, and the diversity of the treatment comparisons represented. As discussed in Chapter 2 and 4, the choice of relevant interventions and comparators when framing a healthcare decision problem is often arbitrary and subject to contextual factors. Jansen and Salanti respectively describe some common patterns of comparisons, such as a closed ‘loop’ or a star shape, but the geometry of a network of evidence depends on scoping decisions and search protocols [Jansen 2011, Salanti 2008]. Recent methodological research suggests that the assessment of clinical evidence should be systematic and transferable across countries [Laws 2014, Kleijnen 2012]; however how to define a network size for NMA remains an “*unsolved issue*” [Sturtz 2012]. Using a breadth-first search strategy specifically designed to optimise the identification of indirect evidence

and extend the network of relevant studies for analysis could help standardise the REA of new interventions.

VTE is a common complication of surgical procedures and is associated with substantial morbidity and mortality in the UK; therefore, ensuring that the maximum and most appropriate evidence base is considered when evaluating new thromboprophylactic drugs is essential. In this example, extensions of the network maximised the number of indirect comparisons between existing pharmacological interventions and precision was increased from base case to first network order as additional studies reduced the uncertainty around mean ORs for all VTE/all-cause death, total DVT, and any bleeds. However, estimates became more stable as fewer studies were included in the evidence networks with each subsequent search order. I found that additional information provided by trials comparing existing treatments to a lower dose of enoxaparin (30mg/bd) identified in first order searches contributed in large part to the increased precision across all outcome estimates. Overall, results from the NMA were consistent across network orders and extending the networks did not increase heterogeneity or inconsistency between studies. The cost-utility analysis was insensitive to NMA results, variation in the clinical input data according to network order did not impact mean ICERs but reduced the uncertainty in outcomes without influencing the acceptability of interventions.

5.6.1 Caveats and model limitations

The selection of first order comparators was arbitrary as no clear definition of how to optimally choose these search terms currently exists. Hawkins *et al.* start the iterative searches in their practical example looking at all currently licensed treatments for non-small cell lung cancer (NSCLC) across regulatory jurisdictions [2009b]; in this example, first order search considered indicated pharmaceutical interventions for VTE prophylaxis in the UK. Although the NICE scoping process can provide some grounds for defining first order comparators, depending on the therapeutic area these can include four interventions, i.e. for second-line stage III/IV NSCLC, or 30 in this case study. The choice of first order comparators should not influence final search results but could impact how many search iterations are needed in the breadth first strategy.

I found no particular gains were achieved from further dividing search orders into two distinct search iterations, as the additional burden of including all comparators, even placebo, rather than all but one comparator was marginal. Ultimately, all relevant

comparators will be identified in the sequence of searches; however, the incremental value of higher search and network orders for NMA should be weighed against the associated additional search and computational burden. Caldwell writes: “*The larger the network the more intensive the assessment of transitivity, data extraction, risk of bias assessment and tabulation of results is likely to be*” [2014]. For example, initially splitting each search order as recommended by Hawkins *et al.* to minimise the search burden, i.e. searching for ‘all except one’ comparators and subsequently searching the omitted comparator separately, proved inefficient. I agree with Hawkins *et al.* that such omission is redundant if the next search order is conducted and abstracts reviewed, as was the case in this example [2009a]. In HTA, searches conducted as part of a clinical evidence review could inform first network order searches, even if NMA specific study selection criteria may be required, and could help alleviate the search burden.

A number of additional limitations may hinder the selection of studies for NMA; Caldwell argues that a pre-specified search strategy to extend the network could “*mitigate but not eliminate the risk of post hoc inclusion/exclusion of treatments*” [2014]. For example, efforts to widen an evidence base for analysis are highly dependent on the available literature. Salanti suggests a number of inherent biases in the medical literature impede the optimal search for evidence:

“what is studied is not necessarily what is eventually published: selective reporting biases, publication bias, time lag bias, and other selection forces further affect the amount of publicly visible evidence on specific treatments”
[2008: p545]

Such biases can also influence the extent to which the transitivity assumption holds for studies included in NMA. In addition, a greater number of treatment comparisons can either contribute to increase between-study heterogeneity or provide a more precise estimate of it [Caldwell 2014]. As the breadth of searches is extended, small-sample size and older studies with different populations are more likely to be identified and increase the potential for time lag bias. Thus heterogeneity and inconsistencies may also increase, although this didn’t appear to be the case here, adjusting for baseline risk could overcome this and still allow for the potential benefits of a larger network [Achana 2013]. I could have also considered node-splitting as another way of assessing inconsistency but as none was identified using other methods, I expect this would not have added anything [Dias 2010]. Another limitation is which and how outcomes are reported in trial publications.

Across all networks, between three and 13 studies were excluded from the analyses because they did not report outcomes of interest.

Recent work in multiple outcomes analysis could help maximise the evidence base and improve NMA methods [Lu 2007, Welton 2008, König 2013]. Moreover, König *et al.* propose a new method to characterise the flow of evidence in an NMA using linear coefficients to interpret the “*parallelism*” and “*indirectness*” of networks to gauge the risk of bias, heterogeneity, and inconsistency within an NMA [2013]. Such methodological extensions to understand an evidence base, including how searching and identifying indirect evidence could be examined quantitatively to optimise network shape and size, are desirable.

The application of Hawkins *et al.* search methods in this Chapter to evaluate the cost-effectiveness of VTE prophylaxis suggests that more exhaustive searches to identify indirect evidence can provide valuable additional information for NMA [2009a]. Given the contradictory results found by Hawkins *et al.* in their similar study evaluating relative effectiveness estimates for NSCLC treatments across multiple network sizes, the impact of extending the network size on uncertainty remains case-specific [2009b]. However, taken together with my findings, this highlights the case for examining a wider network of evidence for clinical review and quantitative data synthesis in HTA. In the absence of current guidelines on searching for indirect evidence, I recommend Hawkins *et al.*’s search strategy to future researchers and reviewers. In addition and in accordance with NICE guidelines for technology appraisal, sensitivity analyses should be performed on studies which could extend the network of evidence and the use of narrower networks for economic modelling and decision-making should be adequately justified [NICE 2013a, Caldwell 2014]. This awareness should prevent, or at least discourage, ‘gaming’ when undertaking and reporting NMAs. Future work, such as a simulation study, to evaluate the impact of network size and shape for NMA would provide generalisable findings and help formalise guidance on the added value of indirect searching and network extensions.

Chapter 6

Mitigating bias in observational data: meta-analysis of vertebral augmentation procedures

6.1 Background

Chapters 4 and 5 have explored evidential issues associated with the identification, selection and synthesis of RCT data. However, in order for HTA to inform relevant health care decisions, it should reflect and be applicable to ‘real-world’ clinical settings. For this reason, international HTA agencies often use non-randomised evidence to supplement RCT data and to help determine whether efficacy claims in trial conditions can be extended to routine practice [Goodman 2014], as highlighted in the review of HTA guidelines presented in Chapter 2.

In this chapter, I explore how observational studies can provide additional evidence on the effectiveness of a new technology by combining RCT and registry data in an adjusted meta-analysis to compare percutaneous vertebroplasty (PVP) vs. optimal pain management (OPM) to treat symptomatic osteoporotic vertebral compression fractures (VCFs).

6.1.1 Surgical augmentation procedures for VCFs

A compression fracture occurs as a result of a weakened or injured vertebra causing a break or collapse in the vertebral body most often characterised by sudden and severe back pain. Osteoporosis is a common condition in which bones lose their strength, making them less dense and more fragile; it is the leading cause of VCFs with an estimated 25,000 to 40,000 osteoporotic spinal/vertebral fractures treated each year in the UK [Burge 2001]. Back pain is the main symptom of VCFs in the short-run but may subside once the fracture is fully healed; however, VCFs are also associated with a number of complications including spinal curvature or deformities (e.g. kyphosis), impaired functional status, difficulties breathing, and loss of mobility [Garfin 2001, Longo 2012]. Longo *et al.* point out that these comorbidities can lead to “[a] diminished ability to perform activities of daily living and a reduction in the quality of life” [2012]. In addition,

patients with osteoporosis who have experienced at least one ‘fragility’ fracture such as a VCF are at higher risk for subsequent fractures, chronic pain, and increased mortality [Klotzbuecher 2000].

In April 2013, two vertebral augmentation procedures—PVP and balloon kyphoplasty (BKP)—were recommended in the UK by NICE to treat painful VCFs due to osteoporosis [NICE 2013c]. Both surgical procedures are minimally invasive and consist of injecting bone cement in the fractured vertebra to stabilise the spine and relieve pain. During vertebroplasty, the cement is directly injected through a small puncture in the skin—percutaneously—replacing the need for open back surgery. For BKP, a balloon is initially inserted in the fragmented vertebra and inflated to create a cavity within the vertebral body, compacting the bone around it and elevating the fracture [Garfin 2001, Denaro 2009]. The balloon is then deflated and removed before cement is injected to fill in the space created by the balloon, a metal stent may also be used for additional support [NICE 2013c]. Unlike PVP, BKP aims not only to reduce pain from symptomatic VCFs but also to correct spinal deformities by restoring some or all of the vertebral body height.

The multiple technology appraisal of PVP and BKP demonstrated the cost-effectiveness of both procedures compared to the conservative/non-invasive management of VCFs in the UK [NICE 2013c]. At the time of the effectiveness assessment, a total of 9 RCTs investigated the added clinical value of PVP²⁴ and/or BKP²⁵ in patients with painful osteoporotic VCFs. Only one trial compared BKP directly to PVP [Liu 2010] whilst the majority of the remaining RCTs evaluated PVP vs. conservative treatment, OPM, or operative placebo with local anaesthesia (OPLA). Although there is no gold standard for the non-invasive treatment of VCFs it usually refers to non-operative measures focused on alleviating pain and supporting the spine such as analgesic use (i.e. OPM), bed rest, and/or back braces [NICE 2013c].

Pain scores and back-specific functional status, as assessed by pain visual analogue scales (VAS) and disability questionnaires, were the most widely reported primary outcomes in randomised trials. All-cause mortality was assessed as a secondary outcome; however, included RCTs were not powered to detect a mortality difference and trial results did not

²⁴Voormolen *et al.* 2007 (VERTOS), Buchbinder *et al.* 2009, Kallmes *et al.* 2009 (INVEST), Rousing *et al.* 2009, Klazen *et al.* 2010 (VERTOS II), Liu *et al.* 2010, Rousing *et al.* 2010, Farrokhi *et al.* 2011, Blasco *et al.* 2012.

²⁵Wardlaw *et al.* 2009 (FREE), Liu *et al.* 2010, Boonen *et al.* 2011 (FREE).

achieve statistical significance, even when pooled, comparing operated patients to patients receiving only OPM. Although both augmentation procedures were recommended, the ERG described the link between VCFs and mortality as: “*an important, yet inadequately understood issue*” [Stevenson 2012]. The impact of augmentation on overall survival was explored in the assessment group’s economic model through various assumptions informed by the available randomised evidence and registry data; however, no explicit inclusion of the observed treatment effects was considered.

6.2 Objectives

Based on NICE’s appraisal (TA279), improvements in pain scores and functional status from included RCTs were sufficient to demonstrate the added value for money of PVP and BKP (without stenting) compared to non-invasive management, respectively, in reducing pain and disability in patients with osteoporotic VCFs [2013c]. There was little trial evidence available on the long-term impact of VCF treatments on morbidity and mortality, including a potential sustained survival benefit for patients having undergone surgical augmentation. However, recently published large-scale registry studies from Germany and the USA found a significant mortality difference between patients having undergone PVP or BKP and non-invasively managed patients.

For economic evaluation, a lifetime horizon is often used to capture all the relevant costs and health consequences of a new treatment over a patient’s lifetime. When RCT data is not available or insufficient, observational data can inform missing model inputs and in this example, long-term claims data was used to account for a mortality benefit in the PVP and BKP treatment arms. The assessment group tested different scenarios to evaluate the impact of a differential mortality rate on the cost-effectiveness of both augmentation procedures vs. OPM [Stevenson 2012].

The analysis sought to estimate the mortality differences between treatments for osteoporotic VCFs by pooling randomised and observational data using a power adjustment to transform prior distributions and a bias allowance within a Bayesian pairwise meta-analysis model. Both approaches sought to demonstrate the impact of including ‘real-world’ evidence on the relative effectiveness and cost-effectiveness assessments of PVP vs. OPM. The specific chapter objectives are:

- i. to use bias adjustment methods to pool both randomised and observational data available from Germany and the USA to estimate mortality differences between treatments for osteoporotic VCFs;
- ii. to evaluate the impact of mortality differences on the cost-effectiveness of PVP vs. OPM from an NHS/PSS perspective.

6.3 Adjusted meta-analysis

6.3.1 Mortality data

A number of studies have explored the link between the occurrence and recurrence of vertebral fractures and an increased risk of death [Kado 1999, Cauley 2000, Jalava 2003]. Although this link is widely reported in the literature, a number of explanations have been put forward to justify a potential causality between VCFs and death. Namely, it has been suggested that the added pressure on the pulmonary and gastrointestinal systems associated with VCFs and kyphosis could be the primary reason for the excess mortality associated with vertebral fractures [Gangi 2006, Stevenson 2012]. Stevenson *et al.* also acknowledge that “*the increased risk of death may also be due, at least in part, to the co-existence of serious underlying diseases in many individuals with VCF.*” [2012].

The mortality benefit from vertebral augmentation procedures is not as well-documented and considerable uncertainty remains as to whether operated patients with VCFs live longer. At present, RCTs have been too small or too short to detect a mortality difference between BKP, PVP, and the non-invasive management of osteoporotic VCFs. However, recently published large-scale registry data from Germany and the USA shows a significant improvement in survival rates for patients having received vertebral augmentation. This section provides a summary of the available randomised and non-randomised evidence on mortality for BKP and PVP vs. OPM, as well as the statistical methods used to synthesise both types of evidence, prior to presenting the results of the adjusted meta-analysis.

6.3.1.1 Randomised evidence

Table 26 summarises the overall mortality results available extracted from 7 out of the total 9 RCTs included in the clinical effectiveness assessment of BKP and PVP. Randomised patients all suffered from at least one painful osteoporotic vertebral fracture, most often refractive to medical therapy and with a clinical onset of no longer than one year; the number of participants ranged from 34 to 300 and the length of follow-up varied

from 2 weeks to 36 months. Only the FREE trial reported on serious adverse events resulting in death for BKP vs. nonsurgical management [Boonen 2011]; the remaining six trials compared PVP to a control group. Unsurprisingly given the RCTs small sample sizes, none of the calculable HRs at the different time points provided were statistically significant. Stevenson *et al.* [2012] performed a meta-analysis on the mortality rates at 12 months for PVP vs. OPM reported by Rousing *et al.* [2009], Klazen *et al.* (VERTOS II) [2010], and Blasco *et al.* [2012]. Pooled results—HR of 0.68 (95%CI 0.30;1.57)—suggested that PVP might be associated with a reduction in mortality but did not reach statistical significance.

Table 26 Overall mortality at different time points from RCTs

	Study <i>Comparison</i>	Experimental arm		Control arm		Hazard ratio (95%CI)
		Events	Total	Events	Total	
2 wk.	VERTOS [2007] <i>PVP vs. control</i>	0	18	0	16	Not calculable
3 mo.	INVEST [2007] <i>PVP vs. control</i>	0	68	0	63	Not calculable
6 mo.	Buchbinder <i>et al.</i> [2009] <i>PVP vs. control</i>	2	38	1	40	2.11 (0.20-22.28)
12 mo.	Rousing <i>et al.</i> [2009] <i>PVP vs. control</i>	1	25	1	24	0.96 (0.06-14.50)
	VERTOS II [Klazen 2010] <i>PVP vs. control</i>	5	101	6	101	0.83 (0.26-2.64)
	Blasco <i>et al.</i> [2012] <i>PVP vs. control</i>	3	64	6	61	0.48 (0.12-1.82)
24 mo.	FREE [2011] <i>BKP vs. control</i>	12	149	11	151	1.11 (0.50-2.43)
36 mo.	Farrokhi <i>et al.</i> [2011] <i>PVP vs. control</i>	2	40	1	42	2.10 (0.20-22.26)

mo.: months, wk.: weeks

6.3.1.2 Non-randomised evidence

Although the randomised evidence was inconclusive, observational data from large medical claims databases from the USA and Germany suggest that both PVP and BKP can improve operated patients' survival over a number of years. In 2011, Edidin *et al.*

explored the association between operative treatments for VCFs and mortality in elderly patients in the entire USA Medicare population. Their study identified 858,978 patients with a newly diagnosed VCF from the 100% Medicare inpatient and outpatient claims data²⁶ between January 2005 and December 2008 [2011]. Patients were stratified according to the procedure they received: BKP or PVP—‘operated’ or ‘augmented’ group—and compared to those that did not undergo surgery—‘nonoperated’ group. Survival times were estimated using the index diagnosis date until the date of death recorded in Medicare enrolment files, or the end of follow-up, right-censoring patients who remained alive on December 31, 2008. Edidin *et al.* included 119,253 patients who underwent BKP (13.9%), 63,693 patients who underwent PVP (7.4%), and the remaining 676,032 patients were categorised as the ‘nonoperated’ cohort (78.7%) [2011].

The Medicare claims database was also used by McCullough *et al.* [2013] and Chen *et al.* [2013] to compare mortality rates in operated and nonoperated patients with VCFs restricted to fractures associated with osteoporosis, from 2002 through to 2006, and 2006 only, respectively. The Medtronic manufacturer’s submission also presented an updated analysis—Exponent—with up to five years follow-up claims data from the USA Medicare; however, results were yet unpublished and redacted as commercial-in-confidence [2012].

In Germany, the AOK Niedersachsen is one of the largest providers of statutory health insurance and covers just under a third of the German population [2016]. Lange *et al.* reviewed claims from the AOK Niedersachsen database for patients ≥ 60 years that had at least one osteoporotic VCF diagnosis in the inpatient sector or two secured diagnoses in the outpatient sector between January 2006 and December 2010 [2014]. Using a similar approach to Edidin *et al.* [2011], survival was calculated from patients’ first inpatient or outpatient osteoporotic VCF diagnosis until death or end of follow-up, right-censoring patients who remained alive on December 31, 2010 [Lange 2014]. A total of 3,607 elderly patients with a newly diagnosed osteoporotic VCF were included in the Lange *et al.* study: 441 patients who underwent BKP (12.2%), 157 patients who underwent PVP (4.4%), and the remaining 3,009 patients were categorised as the ‘nonoperated’ cohort (83.4%).

²⁶ Patients identified using the ICD-9-CM diagnosis codes of 733.13 (pathologic fracture of vertebrae) or 805.0, 805.2, 805.4, 805.6, or 805.8 (cervical, thoracic, lumbar, sacrum/coccyx, and other unspecified vertebral fractures) [Edidin 2011].

Table 27 presents the unadjusted and adjusted HRs at different time-points extracted from the observational studies, manufacturers' submissions and technology assessment report for TA279 comparing PVP, BKP, and non-operative management (i.e. control).

Table 27 Overall mortality at different time points from long-term claims databases

	Study Comparison	Unadjusted hazard ratio (95%CI)	Adjusted hazard ratio (95%CI)
30 days	McCullough <i>et al.</i> [2013] <i>Operated vs. control</i>	0.29 (0.20;0.41)	0.61* (0.39;1.04)
12 months	McCullough <i>et al.</i> [2013] <i>Operated vs. control</i>	0.83 (0.75;0.92)	0.92* (0.81;1.04)
	Lange <i>et al.</i> [2014] <i>PVP vs. control</i>	0.61 (0.41;0.90)	0.70* (0.46;1.07)
	Lange <i>et al.</i> [2014] <i>BKP vs. control</i>	0.74 (0.62;0.89)	0.86* (0.58;1.27)
	Lange <i>et al.</i> [2014] <i>BKP vs. PVP</i>	1.22 (0.98;1.52)	1.22* (1.18;1.26)
	Chen <i>et al.</i> [2013] <i>PVP vs. control</i>	0.81 (0.78;0.84)	0.85 [‡] (0.81;0.88)
	Chen <i>et al.</i> [2013] <i>BKP vs. control</i>	0.58 (0.57;0.60)	0.68 [‡] (0.66;0.70)
	Chen <i>et al.</i> [2013] <i>BKP vs. PVP</i>	0.72 (0.69;0.75)	0.80 [‡] (0.77;0.84)
48 months	Edidin <i>et al.</i> [2011] <i>Operated vs. control</i>	n/a	0.63 [‡] (0.62;0.64)
	Edidin <i>et al.</i> [2011] <i>PVP vs. control</i>	n/a	0.76 [‡] (0.75;0.77)
	Edidin <i>et al.</i> [2011] <i>BKP vs. control</i>	n/a	0.56 [‡] (0.55;0.57)
	Edidin <i>et al.</i> [2011] <i>BKP vs. PVP</i>	n/a	0.77 [‡] (0.75;0.78)
60 months	Lange <i>et al.</i> [2014] <i>Operated vs. control</i>	n/a	0.57* (0.48;0.70)

*Adjusted using propensity score matching; [‡]adjusted using Cox proportional hazards regression model.

Unlike RCTs, observational data is not generated in a controlled environment therefore as Faria *et al.* describe: “no individual is observed in both the treated and non-treated state and therefore the counterfactual is not observed” [2015] which complicates the estimation of a treatment effect. The main concern in interpreting and synthesising observational data is the high risk of bias and confounding due to the non-random assignment of patients in a study. In this example, selection bias could be present if patients in the treated and control groups have inherently different probabilities of death; irrespective of the treatment they underwent [Eddy 1992]. Confounding refers to the

presence of unobserved or unknown variables—confounders—that effect both the selection for treatment and the outcomes of interest.

Authors used a number of regression and matching methods to adjust the differences in the mortality risks obtained from the retrospective analyses of national claims databases. McCullough *et al.* and Lange *et al.* used propensity score matching methods to better account for the selection bias and uneven covariates distribution between the BKP, PVP, and control groups inherent in observational studies [2013, 2014]. Chen *et al.* performed a multivariate analysis with a Cox proportional hazard regression model to control for possible confounding by comorbidities, age, sex, and race [2013]. Edidin *et al.* also conducted a Cox regression to account for variation in gender, age, race/ethnicity, patient health status, type of diagnosed fracture, year of diagnosis, etc [2011]. As shown in Table 27, the adjusted HRs provide a more conservative estimate of treatment effect but support a survival benefit for operated vs. nonoperated patients and BKP vs. PVP.

6.3.2 Methods

A number of statistical techniques are available to adjust observed treatment effects, such as propensity scoring or regression methods, used by Lange *et al.* [2014] and Edidin *et al.* [2011], respectively, to model study-specific biases based on individual studies characteristics. These methods are employed prior to any evidence synthesis to adjust for differences in case-mix between treatment groups, they often but not always provide a more conservative estimate of treatment effect, see Table 27 [Sterne 2002, Deeks 2003]. Another approach is to down-weight high risk of bias studies in evidence synthesis, i.e. to assign a quality weight to each study reflecting discrepancies in both internal and external validity in accordance with a study quality assessment tool.

Bayesian methods provide a flexible framework to include all the evidence available to estimate the mortality difference between vertebral augmentation procedures and the non-invasive management of osteoporotic VCFs. This section describes two Bayesian models: the power transform prior model and the bias allowance model. I focused on the comparison between PVP and OPM at 12 months building on the meta-analysis conducted by Stevenson *et al.*, cf. section 6.3.1.1 [2012]. First, the observed log HRs for PVP vs. control for overall mortality at 12 months were pooled in a fixed effect model. The adjusted treatment effects reported in Table 27 were used in the main analysis; and in order to maximise the non-randomised evidence available at one year follow-up,

adjusted survivorship data were extracted from Kaplan-Meier curves in Edidin *et al.* [2011]. The Kaplan-Meier data extraction method is described in section 7.3.1.1 in Chapter 7. Next, a Bayesian pairwise meta-analysis was conducted using the power transform prior model and the bias allowance model to explicitly combine the RCT and observational data using the long-term claims data as prior information. All analyses were performed in WinBUGS version 1.4.3 [Lunn 2000].

6.3.2.1 Model 1 – Power transform prior

Ibrahim *et al.* write: “the power prior provides a useful class of informative priors for Bayesian inference” [2003]. The rationale underpinning the power transform prior model is that a non-informative prior, often used in designing traditional analyses, does not take into account ‘real’ information that could influence posterior estimates. Adapting the notation of Chen and Ibrahim, I defined a joint probability distribution for a parameter of interest θ given by a general likelihood function of θ based on the RCT data, such as a fixed effect model, and the likelihood function of θ based on the observational data [2000]:

$$P(\underline{\theta}|Data) = L(\underline{\theta}|RCTs) * [L(\underline{\theta}|Obs)]^{\alpha} \times P(\underline{\theta}) \quad (2)$$

With α a scalar prior parameter that weighs the observational data relative to the likelihood of the randomised data. α determines the influence of the observational data on the full data meta-analysis, such that if $\alpha = 0$, the observational data is totally discounted and if $\alpha = 1$, the observational data is considered at ‘face value’ and given the same weight in the analysis as RCT data. This model is best used in a sensitivity analysis to test a range of power priors as it may prove difficult to quantify the amount of bias in the observed data; the model 1 code is provided in Appendix D2 for $0.0001 \leq \alpha \leq 1$.

6.3.2.2 Model 2 – Bias allowance

A number of meta-epidemiological studies have investigated the potential extent of bias in observational evidence of effectiveness by systematically comparing the results from randomised and observational studies, assuming RCTs were unbiased [Sterne 2002, Welton 2012]. Sacks *et al.* [1982] and Schultz *et al.* [1995] reported a bias of $\pm 30\%$, whilst Ioannidis *et al.* [2001] suggested $\pm 50\%$, and MacLehose *et al.* [2000] concluded the bias could extend to $\pm 100\%$. Although all the studies indicated that observational studies are potentially biased, the values varied widely and the direction of bias remained

largely unpredictable [Welton 2012]. Nevertheless, these values can be modelled as an extra variance parameter to represent the bias in observational studies and compute a ‘bias-adjusted’ estimate. Similarly to model 1, the adjusted observed estimates can be used to specify a prior distribution for the meta-analysis of randomised data [Welton 2012]. The bias allowance model is defined by Spiegelhalter *et al.* [2004] as below:

$$\text{RCTs: } y_i \sim N(\theta_i, s_i^2) \text{ \& } \theta_i \sim N(\mu, \tau^2) \quad (3)$$

$$\text{Obs: } z_j \sim N(n_j, v_j^2) \text{ \& } n_j \sim N(\phi, \omega^2) \quad (4)$$

Where y_i is the effect size for the i^{th} trial with variance s_i^2 ; θ_i is the estimated effect size with μ the pooled effect size for RCT. The same model is fitted to the observational data, with j studies, to obtain a prior for μ given all the observational data \underline{z} .

$$\mu \sim \phi \mid \underline{z} \quad (5)$$

$$\phi = \phi^* + \delta \quad \text{where } \delta \sim N(0, \sigma^2) \quad (6)$$

In order to allow for some bias, δ is introduced as the bias associated with observational evidence to ϕ^* the unbiased true effect in observational studies. δ is given a mean of 0 and σ^2 represents *a priori* beliefs regarding the extent of the bias. Based on the findings in the literature [Sterne 2002], σ^2 was given the following values corresponding to 0%, 30%, 50%, and 100% bias: 0, 0.02, 0.08, and 0.24, respectively. For reference only, infinite bias was assumed for $\sigma^2=1$. The model 2 code is provided in Appendix D3.

6.3.3 Results

For all the results presented, the first 20,000 MCMC simulations were discarded as a ‘burn-in’ and each model achieved reasonable convergence according to visual inspection of density and history plots. The main analyses were based on a further 30,000 iterations to ensure the robustness of the results. Table 28 presents the independently pooled overall mortality HRs at 12 months for the observational and randomised studies. Given the limited number of studies included in the meta-analysis, fixed-effects models were found to be most appropriate.

Table 28 Fixed effect model results for PVP vs. OPM for the observational and randomised data extracted at 12 months

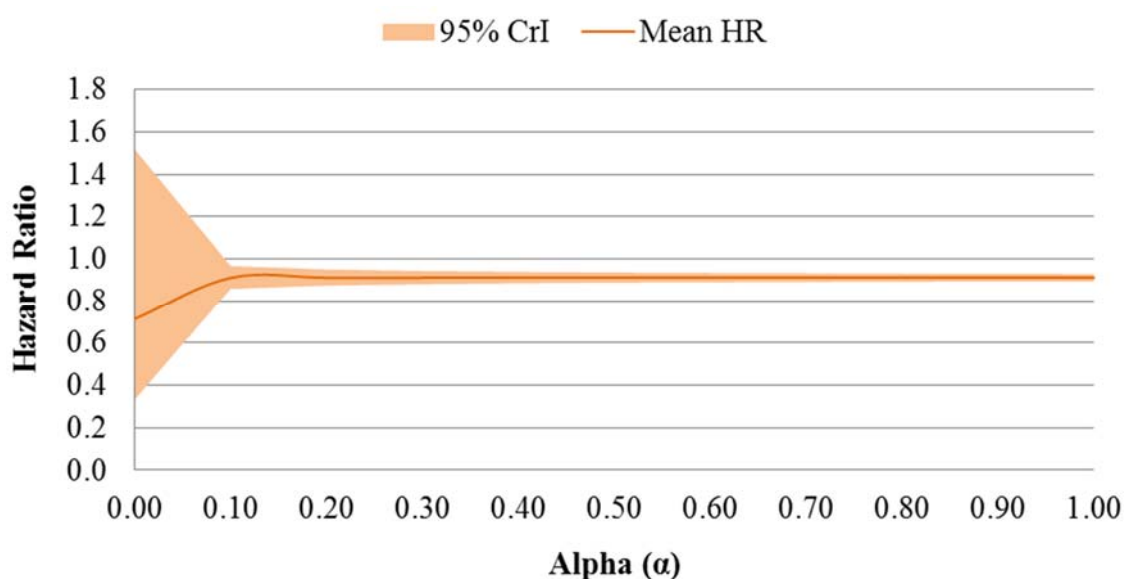
	log(HR)	log(SD)	HR (95% CrI)
Observational studies			
Lange <i>et al.</i> [2014]	-0.354*	0.217	0.70 (0.46;1.70)
Chen <i>et al.</i> [2013]	-0.163*	0.021	0.85 (0.81;0.88)
Edidin <i>et al.</i> [2011]	-0.073‡	0.011	0.93 (0.91;0.95)
Pooled effect	-0.092	0.010	0.91 (0.89;0.93)
Randomised studies			
Blasco <i>et al.</i> [2012]	-0.734	0.694	0.48 (0.12;1.82)
VERTOS II [Klazen 2010]	-0.041	1.40	0.96 (0.06;14.5)
Rousing <i>et al.</i> [2009]	-0.186	0.591	0.83 (0.26;2.64)
Pooled effect	-0.385	0.427	0.68 (0.29;1.58)

SD: standard deviation

*Adjusted using propensity score matching; ‡adjusted using Cox proportional hazards regression model.

Figure 11 illustrates the impact of α on the estimated HR of PVP vs. OPM from RCT data using the pooled effect from the observational studies as an informative prior. For $\alpha = 0$, that is fully discounting the observational data, the power transform prior model results equate those of the pooled randomised data with a HR of 0.71 and similar to wide credible interval of 0.33 to 1.52. On the other hand, when considering the observational evidence on the same level as RCTs, i.e. $\alpha = 1$, the mean HR for PVP vs. OPM is 0.91 (95%CrI 0.89;0.93).

Figure 11 Power transform prior model results for PVP vs. OPM according to α



The results from the bias allowance model are summarised in Table 29 for a range of *a priori* beliefs by varying σ^2 from 0 to 1. Conversely to α , if $\sigma^2 = 0$, the observational evidence is taken at ‘face value’ and if $\sigma^2 = 1$ the observed data is fully discounted. Model 2 results showed that a possible 30% bias would broaden the 95% credible interval to include 1.

Table 29 Bias allowance model results for PVP vs. OPM according to σ^2

Prior belief/Source	Bias	σ^2	HR (95% CrI)
‘Face Value’	0%	0	0.91 (0.89;0.93)
Sacks <i>et al.</i> [1982] and Schulz <i>et al.</i> [1995]	30%	0.02	0.89 (0.68;1.15)
Ioannidis <i>et al.</i> [2001]	50%	0.08	0.83 (0.52;1.32)
MacLehose <i>et al.</i> [2000]	100%	0.24	0.77 (0.41;1.45)
Total discounting	$\infty\%$	-	0.71 (0.33;1.54)

Using the relationship between σ^2 and α —see equation 7—I estimated that $\alpha = 0.1$ corresponded to $\sigma^2 = 0.004$ and was the point at which the 95%CrI no longer contained 1.

$$\sigma^2 = \omega^2 * [1 / (\alpha - 1)] \quad (7)$$

where ω^2 is variance of pooled effect size for observational studies.

This in turn corresponded to requiring that the relative bias associated with observational studies was less than 1% on the HR scale in order for the mortality effect to no longer be considered significant.

The results obtained from the above analyses were used to inform the clinical input parameters in the cost-effectiveness analysis.

6.4 Cost-effectiveness analysis

6.4.1 Methods

The economic model was reconstructed in Excel 2010 based on the assessment group's *de novo* cost-effectiveness analysis for the NICE TA279. The scope of the model was limited to include three comparators: BKP, PVP and the control OPM. The original model also included an arm for OPLA but since no mortality data was available on sham procedures and arbitrary values were only used in sensitivity analysis, OPLA was excluded from the reconstructed model.

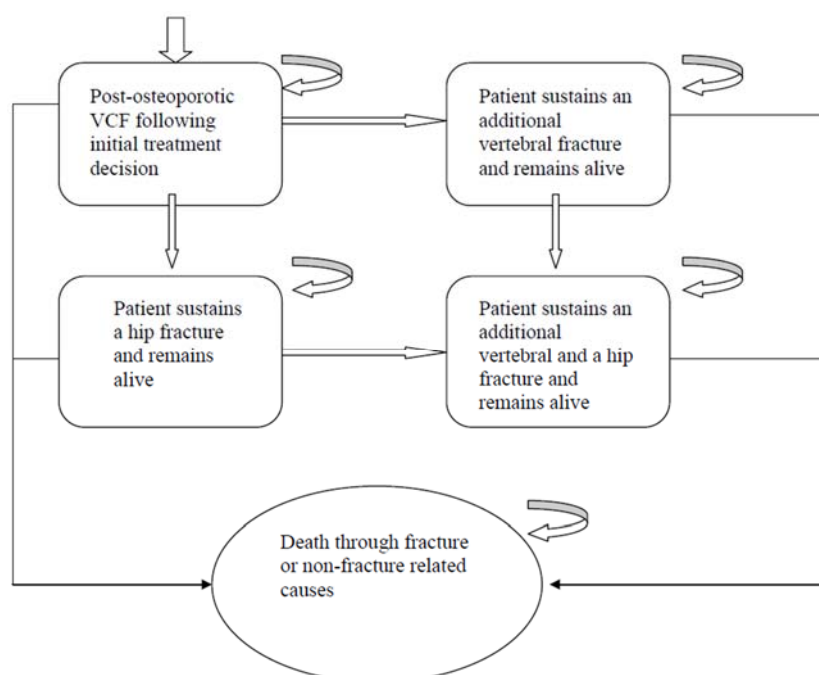
6.4.1.1 Model structure

A five-state Markov model was designed to capture both the short-term quality of life improvements following initial treatment and the long-term differences in mortality rates for each intervention. The increased risk of subsequent fractures following a first VCF was also built into the model. Figure 12 provides a diagrammatical representation of the economic model and illustrates the following five health states:

- 'post-osteoporotic VCF' which was the starting state for patients having been treated with PVP, BKP or OPM;
- a subsequent additional vertebral fracture;
- a subsequent hip fracture;
- both a subsequent additional vertebral and hip fracture; and
- death.

Death is the absorbing state that included both deaths due to a fracture and non-fracture related deaths.

Figure 12 Economic model structure diagram



Taken from technology assessment report for TA279, Figure 21 [Stevenson 2012]

The model structure resembled that of the Medtronic state transition model presented in the manufacturer submission and a previously published economic model by Ström *et al.* comparing BKP to OPM [2010]. The cost-effectiveness analysis employed a 50 year time horizon to model outcomes over a patients' lifetime. A monthly cycle length was used for the first 36 months following initial treatment to take into account utility differences between interventions; followed by 47 yearly cycles.

6.4.1.2 Clinical model parameters

6.4.1.2.1 Transition probabilities

The analysis assumed a cohort of 1,000 patients, 70 year old women with a baseline T-score of -3.0SD, which reflected the patient population of the FREE and VERTOS II trials [Wardlaw 2009, Klazen 2010]. A T-score is a measure of bone density compared with what is normally expected in a healthy adult, it corresponds to the number of standard deviations above or below the average bone mineral density in a healthy subject. The transition probabilities in the model were dependent on age, gender, T-scores, as well as, whether patients were prescribed bisphosphonates to prevent future fractures.

The transition rates for a subsequent additional vertebral fracture or a subsequent hip fracture were taken from Stevenson *et al.* [2009] and calibrated to take into consideration the increased risk based on worsening T-scores, as well as, the additional risks following

an initial fracture for patients aged 70 years and over [Stevenson 2012]. Table 30 and Table 31 summarise the annual risks of vertebral and hip fracture following an initial VCF assumed for patients of a given age and T-score upon entry into the model. These annual risks were also used for patients transitioning to the health state ‘patient sustains an additional vertebral and a hip fracture and remains alive’.

Table 30 Annual risks of vertebral fracture following an initial vertebral fracture based on age and T-Score on entry to the model

	T-Score (SD)			
Age Groups (years)	-2.0	-2.5	-3.0	-3.5
65-69	0.41%	0.56%	0.74%	1.00%
70-74	0.46%	0.62%	0.83%	1.11%
75-79	0.55%	0.74%	0.99%	1.32%
80-84	0.65%	0.87%	1.17%	1.57%
85-89	0.78%	1.05%	1.41%	1.89%

Extracted from technology assessment report for TA279, Table 50 [Stevenson 2012]

Table 31 Annual risks of hip fracture following an initial vertebral fracture based on age and T-Score on entry to the model

	T-Score (SD)			
Age Groups (years)	-2.0	-2.5	-3.0	-3.5
65-69	0.37%	0.60%	0.97%	1.56%
70-74	0.45%	0.72%	1.16%	1.87%
75-79	0.66%	1.07%	1.72%	2.77%
80-84	0.97%	1.56%	2.52%	4.06%
85-89	1.55%	2.51%	4.04%	6.51%

Obtained from Prof. Stevenson via personal communication as an erratum to Table 51 from technology assessment report for TA279 [2012, 2014]

In accordance with the findings from Stevenson *et al.* [2005] and data from Holt and Khaw [2002], a decrease of 0.255SD per 5 year age group was included in the model to account for a patient’s bone density deteriorating over time. The model also incorporated an assumed effect on vertebral and hip fractures for patients taking bisphosphonates, a RR of 0.58 (95%CI 0.50;0.67) and 0.72 (95%CI 0.58;0.88), respectively [Stevenson 2009, 2012]. Using the same assumption as in the original model, the effect of bisphosphonates on the risk of subsequent fractures was assumed to last for five years and to be equal for all ages, a linear wane effect was applied over an additional five year period, so that the RRs for both vertebral and hip fractures were 1 after 10 years [Stevenson 2012]. Table E1 in the Appendix summarises these effects in tabular format.

The underlying all-cause death rates per age and gender were obtained from the National Life Tables for England and Wales from 2010-12 [ONS 2013]. It was assumed that all

patients would die in their 101st year, superseding the 50 year time horizon. The annual all-cause mortality rates for men and women from 60 years old are included in Appendix D4 for completeness.

The mortality rates associated with vertebral fractures were extracted from a UK study by Jalava *et al.* comparing the risk and causes of mortality in patients with osteoporosis and no fracture, and those with one or more prevalent vertebral fractures [2003, Stevenson 2012]. As reported by Stevenson *et al.*, the unadjusted HR of 4.4 (95%CI 1.85;10.6) was employed to inflate the underlying all-cause death rate in the model for the first five years following an initial vertebral fracture; a linear wane effect was also applied over the next five years [2012]. The increased mortality rate was also assigned to the first year following a subsequent vertebral fracture with no dissipated effect over time.

The mortality rates associated with hip fractures were estimated at 6% of patients aged 70-79 years, 11% of patients aged 80-89 years, and 16% of patients aged 90 years and over, based on data from Stevenson *et al.* [2009]²⁷. Further details about the transition probabilities in the economic model and how these were derived can be found in the technology assessment report for TA279 [Stevenson 2012].

6.4.1.2.2 Mortality effect

The mortality benefit of active interventions was a key input in the economic model. The assessment group considered three scenarios based on the following assumptions:

1. BKP had the greatest effect, followed by PVP, compared to OPM;
2. BKP and PVP had the same positive effect compared to OPM;
3. BKP, PVP, OPM had the same effect.

Similarly, Medtronic provided a sensitivity analysis down weighting the relative risk of BKP and PVP vs. control, respectively, to 0%, 50%, and 75% of the reported mortality benefits in their submission [2012]. Unfortunately, the HRs used by Stevenson *et al.* and Medtronic to explore the differential effects of mortality associated with BKP, PVP and

²⁷Note that the following assumption made by the technology assessment group was respected in the economic model:

“It was assumed that the mortality rate following hip fracture could not be lower than either the mortality rate associated with a vertebral fracture, or lower than that of general mortality in the underlying age and gender matched population. In such circumstances the rate of mortality following hip fracture was increased to equal the higher value.” [Stevenson 2012: p183]

OPM were academic in confidence and blacked out from the NICE documentation [2012].

Given that the evidence on BKP compared to either an active or control treatment was not considered in the adjusted meta-analysis (cf. section 6.3.2.), I thought it prudent to solely examine three scenarios in my cost-effectiveness analysis:

1. No mortality benefit for BKP, PVP, or OPM
2. Identical effect of BKP and PVP both being better than OPM (over 5 years)
3. Identical effect of BKP and PVP both being better than OPM (over lifetime)

In an iterative process, the fixed effect model 1 and model 2 results reported in Table D5 in Appendix and Table 29 in section 6.3.3 were used in the economic model to evaluate the impact of a greater mortality benefit for operated patients vs. OPM. In accordance with Stevenson *et al.* [2012], I assumed that the positive mortality effect for both PVP and BKP vs. OPM would only last up to five years and would cease immediately after that time with no waning period. An additional analysis was performed to test this assumption, allowing for the mortality benefit to carry on beyond five years and to influence operated patients' relative risk of death until they reached 101 years. The log HRs for PVP vs. OPM were sampled from a normal distribution using the pooled data from the meta-analysis on the log scale directly into the model.

6.4.1.3 Utility valuation

The utilities associated with each health state were dependent on a number of factors including patient's age and gender, as well as the augmentation procedure undertaken and the time elapsed since the procedure [Stevenson 2012]. The quality of life metric recommended by NICE is the EQ-5D. Five out of 9 RCTs collected EQ-5D data but all the studies included in the effectiveness assessment reported a VAS score as a measure of pain [Buchbinder 2009, Kallmes 2009, Wardlaw 2009, Klazen 2010, Rousing 2010]. Initial VAS scores were extracted from both treatment arms for each included trial and analysed by Stevenson *et al.* [2012]; the mean initial VAS score was 7.36 (95%CI 0.58;0.88). Based on the visual inspection of VAS scores from the RCT data, it was assumed that VAS scores would stabilise one month post-operation for PVP and BKP, and 3 months post-treatment with OPM²⁸. An MTC of mean differences in VAS scores

²⁸Note that the following assumption made by the technology assessment group was respected in the economic model:

during the stable period was performed by Dias and Ades in appendix to the technology assessment report [Stevenson 2012]; however, due to concerns of backward calculations of the academic-in-confidence VAS data from Buchbinder *et al.* [2009], pooled results were not reported. As an alternative, stable VAS scores were extracted from VERTOS II [Klazen 2010] and Lui *et al.* [2010], PVP and BKP scores one month after the procedure were assumed to return to 2.30 (95%CI 2.01;2.59) and 2.60 (95%CI 2.52;2.68), respectively, and OPM scores stabilised to 3.60 (95%CI 3.28;3.92) 3 months after treatment. Absolute EQ-5D and VAS scores were mapped for the 5 studies that reported both outcomes; Stevenson *et al.* found the plot and resultant formula (8) to provide a “*relatively good fit*” to the data [2012]:

$$\text{EQ-5D} = 0.8053 - 0.0674 \times \text{VAS} \quad (8)$$

This formula (8) was used in the economic model to estimate for each health state the QALYs per cycle. A normal distribution was applied to initial and post-treatment VAS scores in the PSA; whilst a Beta distribution was applied to the EQ-5D scores using the variance on the intercept (0.00216) and the variance on the slope (0.00008) [Stevenson 2012]. A detailed account of the VAS data and the mapping exercise undertaken to obtain utilities is provided in the technology assessment report for TA279 [Stevenson 2012].

For patients having sustained an additional vertebral fracture and remaining alive, a utility decrement was incorporated in the model to account for the associated pain in the year of the fracture—multiplier of 0.626—and in the subsequent years—multiplier of 0.909 [Stevenson 2009, 2012]. Similarly, the assumed multipliers following a hip fracture were 0.792 in the first year and 0.813 in subsequent years [Stevenson 2009, 2012]. Utility multipliers were applied to the general population norm matched for age and gender from Ara and Brazier [2010]. No utility decrements were considered for adverse events, but a discount rate of 3.5% was applied yearly to QALYs.

6.4.1.4 Resource use and costs

The economic model incorporated the following cost inputs: costs associated with the initial osteoporotic VCF, acquisition costs of the procedure, as well as, operation and

“It is assumed that the stable utility following an active intervention remains constant until either the patient moves to another health state, or this value is greater than the underlying population norm value at the patient’s age adjusted for the impact of a vertebral fracture. (...) In the latter circumstance the utility was set equal to the adjusted population value for the given age.” [Stevenson 2012: p186]

hospitalisation costs. The one-off costs of a vertebral fracture and a hip fracture were estimated by Stevenson *et al.* [2009] and inflated to 2010/11 prices²⁹; these costs were assumed for each additional fracture. Ongoing costs of £239 were also assigned per annum following a fracture. The acquisition costs for PVP with high viscosity cement were obtained from the Johnson and Johnson manufacturer's submission; however, the assessment group down weighted these costs for low-viscosity cement PVP procedures [Stevenson 2012]. The lower estimate of £800 was used in my model. The list price of BKP and cement (i.e. £2,639) was used in the model, whilst no acquisition costs were associated with OPM. The costs incurred in preliminary phase, operating phase and post-operative phase totalled £1,311 per operation and included clinician visits, surgery, any required tests/X-rays, etc. Hospitalisation costs were calculated based on length of stay in days for each procedure estimated by Medtronic in their manufacturer submission from Hospital Episode Statistics data [2012] and a cost per day of £232³⁰ presented by Johnson and Johnson [2012]. Table 32 summarises all the relevant costs inputted into the model; a discount rate of 3.5% was applied yearly.

Table 32 Cost inputs in the economic model

	Cost	Source
VCF costs		
Vertebral fracture costs	£3,081	Stevenson <i>et al.</i> [2009]
Hip fracture costs	£7,536	
Post-VCF ongoing costs	£229 per year	
Acquisition costs		
PVP – high viscosity cement	£1,546	Johnson and Johnson [2012]
PVP – low viscosity cement	£800	
BKP	£2,639	List price [Stevenson 2012]
OPM	£0	
Operation costs		
Preliminary costs	£540	Johnson and Johnson [2012]
Operating costs	£528	
Post-operative costs	£243	
Hospitalisation costs		
PVP	£1,438	Johnson and Johnson [2012] Medtronic [2012]
BKP	£1,183	
OPM	£2,204	

²⁹Hospital and Community Health Services inflation indices reported by Curtis *et al.* [2011]

³⁰Payment by results national tariff price for an excess bed day associated with PVP/BKP and OPM health resource group codes [Stevenson 2012]

6.4.2 Results

Table 33 presents the PSA means for total costs, total QALYs and ICERs under the three assumptions of mortality benefit. Results for both scenarios including a mortality benefit are reported for analyses using a fully discounted HR (with $\alpha = 0$) and a ‘face value’ HR (with $\alpha = 1$) based on the observational data (cf. section 6.3.3). Under all scenarios, PVP was found to be cost-effective vs. OPM. Since no differential mortality effect was applied to BKP compared to PVP, the cost-effectiveness of BKP vs. OPM is likely to be underestimated.

Table 33 Cost-effectiveness results (PSA means) for PVP, BKP, and OPM

	Total costs	Total QALYs	ICERs (£/QALYs)
1) No mortality benefit			
OPM	£4,211	5.50	-
PVP	£5,581	5.56	£22,833
BKP	£7,165	5.56	<i>Dominated</i>
2a) Mortality benefit up to 5 years – fully discounted observational data, $\alpha = 0$			
OPM	£4,212	5.51	-
PVP	£5,617	5.67	£8,781
BKP	£7,205	5.68	£158,800
2b) Mortality benefit up to 5 years – ‘face value’ observational data, $\alpha = 1$			
OPM	£4,210	5.50	-
BKP	£7,178	5.60	£29,680
PVP	£5,594	5.61	<i>Dominant</i>
3a) Mortality benefit over lifetime - fully discounted observational data, $\alpha = 0$			
OPM	£4,211	5.49	-
BKP	£7,426	6.09	£5,358
PVP	£5,856	6.13	<i>Dominant</i>
3b) Mortality benefit over lifetime - ‘face value’ observational data, $\alpha = 1$			
OPM	£4,212	5.51	-
PVP	£5,663	5.75	£6,046
BKP	£7,247	5.75	<i>Dominated</i>

The total costs and QALYs obtained in Table 33 align with the findings from Stevenson *et al.* [2012] in terms of direction and magnitude. Results were not exactly matched due to slight differences in the modelling of the utilities, as previously noted in section 6.4.1.3 and because Stevenson *et al.* did not stipulate all the costs considered in their base case but evaluated multiple scenarios [2012]. In addition, it is hazardous to compare the ICERs with the assessment’s group’s findings for scenarios including a mortality benefit as information on the mortality HRs used by the assessment group was redacted. The

Markov traces for scenario 1 (i.e. no mortality benefit) are listed in Appendix Tables D6, D7, and D8 for reference.

Figure 13 illustrates the CEACs for PVP vs. OPM for a £20,000 willingness to pay threshold against ascending values of α and σ^2 for scenario 2 for the pooled mortality HRs from Model 1 and Model 2, respectively. Figure 14 presents full CEACs for PVP alone over a range of thresholds for different values of α and σ^2 assuming a mortality benefit for operated patients up to five years. Figures D9 in the Appendix present the CEACs for a £20,000 willingness to pay threshold against ascending values of α and σ^2 for scenario 3 assuming a mortality benefit over patients' lifetime.

Figure 13 Cost-effectiveness acceptability curves according to α and σ^2 for scenario 2 assuming a mortality benefit up to 5 years

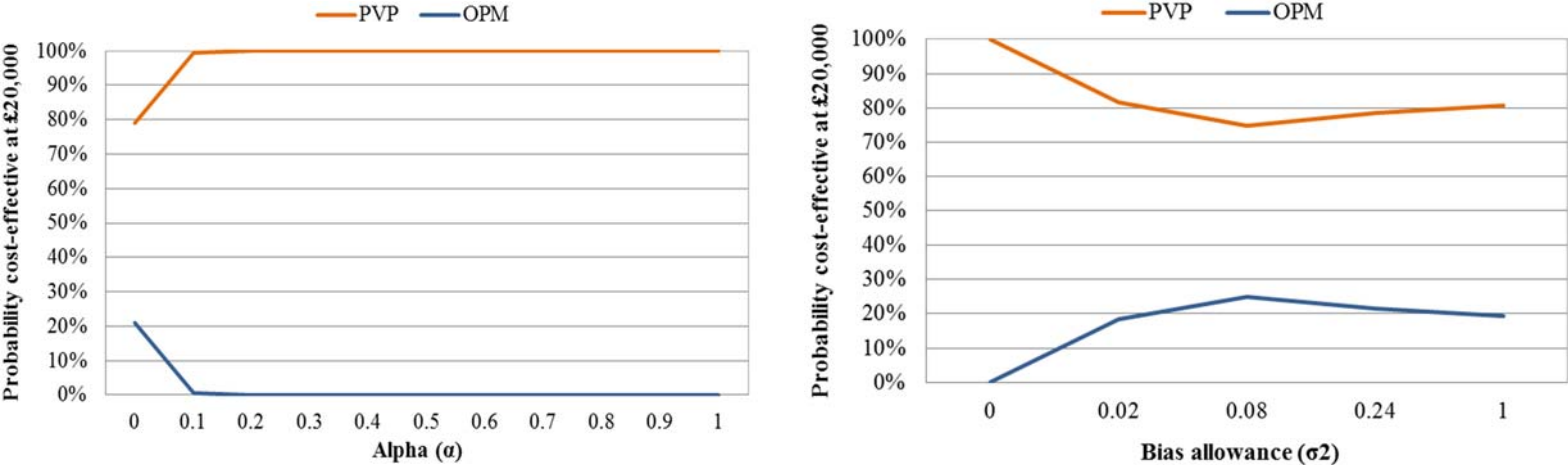
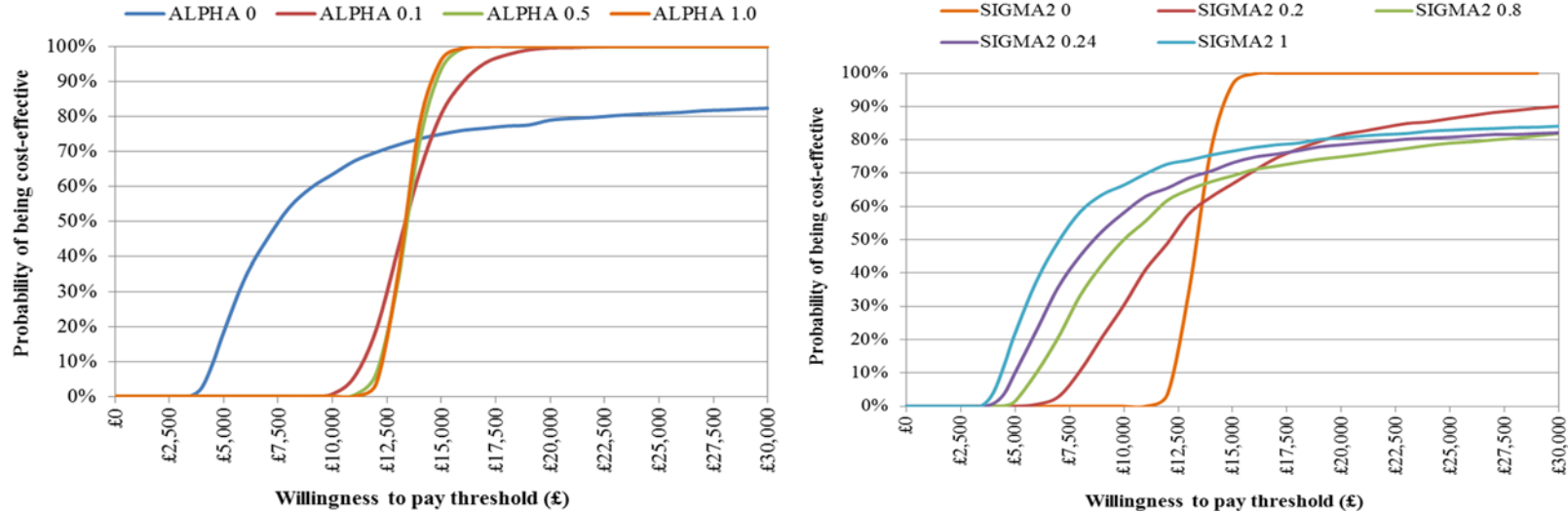
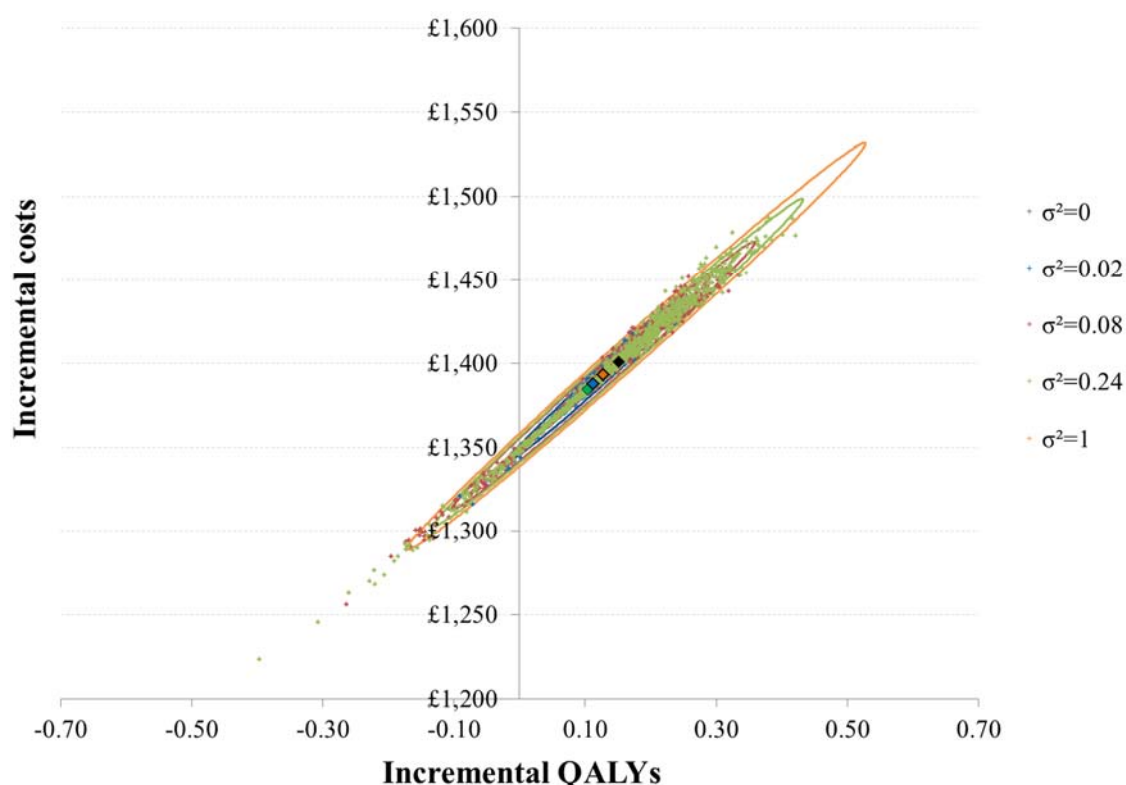


Figure 14 Cost-effectiveness acceptability curves for PVP at different values of α and σ^2 for scenario 2 assuming a mortality benefit up to 5 years



Both Figure 13 and Figure 14 show that PVP is more likely to be cost-effective the greater the weight given to the observational evidence in the meta-analysis models, i.e. the higher the α or the lower the σ^2 . Although the mean PSA results appear to favour the scenarios fully discounting the observational data in both models, it is the greater uncertainty in the pooled mortality HRs from the RCTs alone that is favouring PVP. Figure 15 shows the cost-effectiveness plane for PVP vs. OPM for different values of σ^2 for scenario 2 assuming a mortality benefit lasting up to five years for PVP patients. When the observational evidence is taken at ‘face value’, the prior distribution put on the fixed effect models substantially narrows the credible intervals around the point estimate HR for PVP vs. OPM reducing the overall uncertainty in the ICERs as shown in the probabilities of being cost-effective. Moreover, when testing the assumption of the duration of the mortality effect for operated patients, results in Table 33 reveal that extending the mortality effect over a lifetime can considerably improve the cost-effectiveness of PVP vs. OPM. Note that for all scenarios, PVP remained a cost-effective treatment option for patients with painful osteoporotic VCFs at a willingness to pay threshold between £20,000 and £30,000 per additional QALY.

Figure 15 Cost-effectiveness plane for different values of σ^2 for scenario 2



6.5 Discussion

Existing bias adjustment methods are often study-specific, relying on the availability of IPD, and tend to be applied prior to any evidence synthesis. Bayesian methods can provide a flexible framework to consider the inclusion of all evidence, at a minimum, as a sensitivity analysis to estimate the degree of relative bias in observational studies that would be required to change inferences and/or decisions primarily based on randomised evidence. The two bias adjustment models presented in this chapter were used to evaluate the impact of long-term registry data on the relative effectiveness and subsequently the cost-effectiveness of PVP vs. OPM for painful osteoporotic VCFs.

The results in section 6.4.2 demonstrated the sizeable influence that a differential mortality effect for operated patients could have on the cost-effectiveness of PVP and the key role that observational evidence could play in reducing the uncertainty in the ICERs. Especially, when the RCT evidence available is sparse and wide confidence intervals can drive model outputs.

In this particular example, PVP was cost-effective compared to OPM under all scenarios evaluated, when the mortality benefit was removed, limited to five years following surgery, or extended to a lifetime. Therefore the results presented would have most likely not changed NICE's recommendation, but allowing for observational evidence to be considered—even accounting for a varying degree of bias—noticeably reduced the uncertainty around the PVP vs. OPM ICERs. The power transform prior model and the bias allowance model also permitted the explicit and transparent consideration of long-term claims data in the economic evaluation.

6.5.1 Caveats and model limitations

The power prior approach can act as sensitivity analysis by testing different values of α , modelling the bias allowance requires empirical evidence or elicitation of prior beliefs regarding the appropriate degree of potential bias in the observational studies. The latter implies a degree of subjectivity and critical application of these methods, and well as the careful interpretation of results.

More sophisticated methods of adjusting observed effects in evidence synthesis were not examined in this chapter, such as a generalised evidence synthesis framework or hierarchical modelling, as well as the statistical techniques proposed by Turner *et al.*

[2009] and Welton *et al.* [2009]. The bias adjustment model by Turner *et al.* was considered but the elicitation of expert opinion on the internal and external validity of the observational studies included in the analysis proved very challenging.

In addition, I focused on PVP vs. OPM given the randomised and observational data availability; however, this did not allow for a simultaneous comparison of BKP, PVP, and OPM. Further work is needed to investigate how the models proposed could be extended to NMA in order to provide a more comprehensive assessment of osteoporotic VCF treatments in the UK. One methodological approach has been put forward by Schmitz *et al.* and evaluated in an adjusted MTC for rheumatoid arthritis [2013].

Since undertaking this analysis, an updated report on the Medicare claims data was published by Edidin *et al.* [2015] which provides even more compelling and long-term evidence of the mortality benefit of BKP and PVP compared to the non-invasive management of vertebral fractures in the USA. However, Stevenson *et al.* caution the use of observational studies in their concluding remarks of the TA279 technology assessment report:

“It is possible that BKP and PVP may lead to longer-term reductions in mortality and at different levels of effect; however, this possibility was derived from registry data and without information on the causes of death in these cohorts, and in the absence of randomisation, it was not possible to conclusively establish a causal link.” [2012: p8]

Indeed, exploring the association between treatment and mortality does not prove a causal link and issues of selection bias and confounding, despite adjustment, could be misleading. Stevenson *et al.* add:

“Ideally, this outcome would be explored in a well controlled RCT. However, the sample size and length of follow-up required to detect meaningful differences would make such a trial difficult to perform.” [2012:p9]

Therefore, optimising the use of observational data in HTA may be the only achievable option when an RCT is not feasible.

Chapter 7

Simulating early clinical evidence: cost-effectiveness of ticagrelor over time

7.1 Background

Recent literature extolling the merits of accelerated or conditional drug approval, as summarised in Chapter 2, focuses on the importance of enabling early patient access to valuable medicines based on promising clinical results. By fast-tracking the assessment and availability of ‘breakthrough’ drugs, both regulators and payers must consider the potential clinical benefit of a product, over that of other therapies on the market, to outweigh the inherent heightened uncertainty with respect to final outcomes. That is to say, the otherwise unrealised health gains of a new drug should be worth the additional risk incurred by patients when granting early market approval and reimbursement.

Using the example of ticagrelor in ACS presented in Chapter 4, I examine whether data simulation can help demonstrate the relative effectiveness and cost-effectiveness of a drug prior to the completion of a pivotal study; and what impact this may have on ‘early’ decision-making.

7.1.1 Ticagrelor in acute coronary syndromes

In October 2011, NICE recommended the use of ticagrelor for the prevention of atherothrombotic events in adult patients presenting with ACS in England and Wales [2011a]. As described in Chapter 4, the relative effectiveness and cost-effectiveness of ticagrelor compared to clopidogrel—the current standard of care in the NHS—were largely based on the results from one Phase III multicentre randomised study: PLATO [Wallentin 2009].

The PLATO study was designed to evaluate ticagrelor vs. clopidogrel in a broad patient population with ACS after encouraging results from a Phase IIb dose-guiding safety trial—DISPERSE2—demonstrated no significant difference in bleeding rates between treatment groups [Cannon 2007]. Moreover, initial efficacy results from DISPERSE2

revealed ticagrelor was associated with a favourable trend toward a lower risk of MI [James 2009]. Enrolment for the pivotal PLATO trial began in October 2006 and ended in July 2008, at which time 18,624 patients had been recruited from approximately 800 sites in 43 countries. Randomised treatment was scheduled to continue for a minimum of 6 months to a maximum of 12 months; the follow-up period ended in February 2009 and primary results were published by Wallentin *et al.* in September 2009 [James 2009, Wallentin 2009].

Following the Committee for Medicinal Products for Human Use (CHMP) advice, the EMA was one of the first regulators worldwide to grant ticagrelor a marketing authorisation valid throughout the European Union in December 2010 [EMA 2010]³¹. The NICE consultation and appraisal process span from September 2010, with the drafting of the scope and the matrix of consultees and commentators, to October 2011 when the technology appraisal was published and made available on the NICE website [NICE 2011a]. As summarised in Table 4.1 in Chapter 4, by the end of 2011, other HTA agencies such as CVZ (now ZIN), IQWiG, PBAC, SMC, and TLV had found ticagrelor to be highly cost-effective compared to clopidogrel and recommended its use combined with low-dose aspirin for up to a year as a possible treatment for some people with ACS.

The primary efficacy outcome in the PLATO trial was time from randomisation to first occurrence of any event in the composite measure of MI, stroke, or death from vascular causes [James 2009, Wallentin 2009]. The study protocol provided for a number of follow-up visits at 1, 3, 6, 9, and 12 months after index ACS event; however, Wallentin *et al.* solely reported event rates at 12 months [2009]. In addition, the consistency of treatment effects was to be assessed by comparing relative risk ratios from randomisation to 30 days, and from 31 to 360 days; a single interim analysis was also planned after approximately 1,200 primary events occurred [James 2009]. Unfortunately, although ticagrelor treatment showed early benefits compared to clopidogrel within the first 30 days—hazard ratio (HR) of 0.88 (95% CI 0.77, 1.00) and absolute risk ratio of 0.6%—

³¹ The USA Food and Drug Administration (FDA) approved ticagrelor in ACS patients in July 2011, one year after the FDA's Cardiovascular and Renal Drugs Advisory Committee recommended its approval [FDA 2011]. The FDA's protracted approval process which led to its postponed decision about ticagrelor is thought to be due to a regional interaction in North American patients with the co-administered aspirin maintenance dose which reduced ticagrelor's efficacy, as measured in the PLATO trial, in North America compared to the rest of the world [Mahaffey 2011].

initial findings were not documented in the medical literature and were only presented by AstraZeneca in a meeting with the USA FDA Cardiovascular and Renal Drugs Advisory Committee on July 28, 2010 [2010a, 2010b]. Likewise, the interim analysis was reviewed confidentially by an independent external data and safety monitoring board and results were never made public.

7.2 Objectives

The traditional timescales to complete and report on large RCTs, such as the PLATO study; as well as, the current marketing authorisation and HTA ‘entry hurdles’ can considerably delay the date at which a product reaches the market and impede patients’ access to new valuable treatments. Under new market access schemes, such as accelerated approval in the USA and adaptive licensing/CMA in Europe, REAs of new medicines are often based on early and immature but promising clinical evidence from Phase II and Phase III interim trial analyses.

A simulation study was conducted to investigate the impact of earlier decision-making and to examine the relative effectiveness and cost-effectiveness of a drug prior to the completion of a registrational study. Using the example of ticagrelor, this simulation study formally assesses the efficacy of ticagrelor for the treatment of ACS at different follow-up times not observed and/or reported in the PLATO study. In particular, it allows the clinical trial results to be retrospectively modelled from recruitment to final analysis and the evaluation of the relative effectiveness and cost-effectiveness of ticagrelor vs. clopidogrel under ‘virtual’ time constraints for ‘early’ approval/HTA. Simulation is the technique of closely mirroring the ‘real-world’ based on a delimited set of conditions or factors, sampled from some probability distribution, in order to answer a particular research question [Welton 2012, Lambert 2014]. For example, simulation can be used to recreate IPD for a given trial design based on reported aggregate data when IPD is not available either at the time of analysis or is unpublished.

The specific chapter objectives are:

- i. to simulate IPD for 18,000 patients based on the PLATO study design and report Kaplan-Meier estimates for three outcomes of interest: MI, stroke, and all-cause death;

- ii. to adapt the economic model described in Chapter 4 to incorporate simulated results and evaluate the impact of different efficacy input parameters obtained at different time points on the cost-effectiveness of ticagrelor vs. clopidogrel; and
- iii. to extend the simulation to a subgroup of ACS patients identified at randomisation by investigators as appropriate for an invasive strategy (cf. Chapter 4.4.1.3.2), and apply simulated results from the PLATO-INVASIVE substudy in an ITC to evaluate ticagrelor vs. prasugrel in ACS patients intended for PCI [Cannon 2010].

7.3 Simulation study

7.3.1 Methods

The simulation study was designed to reflect data from the pivotal PLATO trial. The latter RCT formed the core evidence submitted by the manufacturer for both the clinical and cost effectiveness assessment of ticagrelor by NICE [AstraZeneca 2010a]. The size and case-mix of hospitalised ACS patients, with or without ST-segment elevation, included in PLATO were deemed to be representative of UK clinical practice [Bagust 2011]. IPD was simulated using a stepwise approach for the three outcomes of interest—MI, stroke, and all-cause death—needed to populate the cost-effectiveness model. This section provides a detailed account of the data sources and design of the simulation study. All simulations and analyses were carried out in Stata 13 [StataCorp 2013] and Excel 2010.

7.3.1.1 Data sources

Two methods were employed to extract the observed data from the PLATO study. First, where available, the Kaplan-Meier curves for each outcome of interest were digitized and data values for each treatment arm were extracted using the digitizer software; DigitizeIt [Bormann 2012]. Once the data had been extracted from the digitized curves, it was exported in CSV format into Excel 2010 for analysis. A parametric time-to-event survival model was used to determine the baseline risk, i.e. the risk of events in the clopidogrel arm, for MI, stroke and death, respectively. A Weibull distribution was fitted to the ‘replicated’ Kaplan-Meier estimates in order to obtain scale (λ) and shape (γ) parameters for use in the simulation study to predict values of baseline risk. The shape and scale parameters were derived from the Weibull survival function (S_t) using two different time points (t_1) and (t_2)—randomly selected in Excel 2010 so that $0 < t_1 \leq 180$ days and $181 \leq t_2 \leq 360$ days—and the corresponding event or failure rates from the ‘replicated’ dataset:

$$S_t = e^{-(\lambda t^\gamma)}$$

so $S_1 = e^{-(\lambda t_1^\gamma)}$ and $S_2 = e^{-(\lambda t_2^\gamma)}$

with $\lambda = -\frac{\log(S_1)}{(t_1)^\gamma}$

and $\gamma = \log\left(\frac{\log(S_2)/\log(S_1)}{\log(t_2)/\log(t_1)}\right)$

for $t > 0$, $\lambda > 0$ (scale), and $\gamma > 0$ (shape)

Treatment effects for ticagrelor were modelled by applying published HRs from the PLATO study to the baseline risk and predicting Kaplan-Meier estimates for the ticagrelor treatment arm. Figure 16a, 16b, and 16c illustrate this step-by-step process for the Kaplan-Meier all-cause mortality estimates plotted for the PLATO-INVASIVE population in Cannon *et al.* [2010]. As illustrated in Figure 16c, the ‘replicated’ Kaplan-Meier estimates for time to all-cause mortality in the PLATO-INVASIVE subgroup can be plotted against the predicted values from the Weibull survival model.

Figure 16 Extracting data from Kaplan-Meier curves

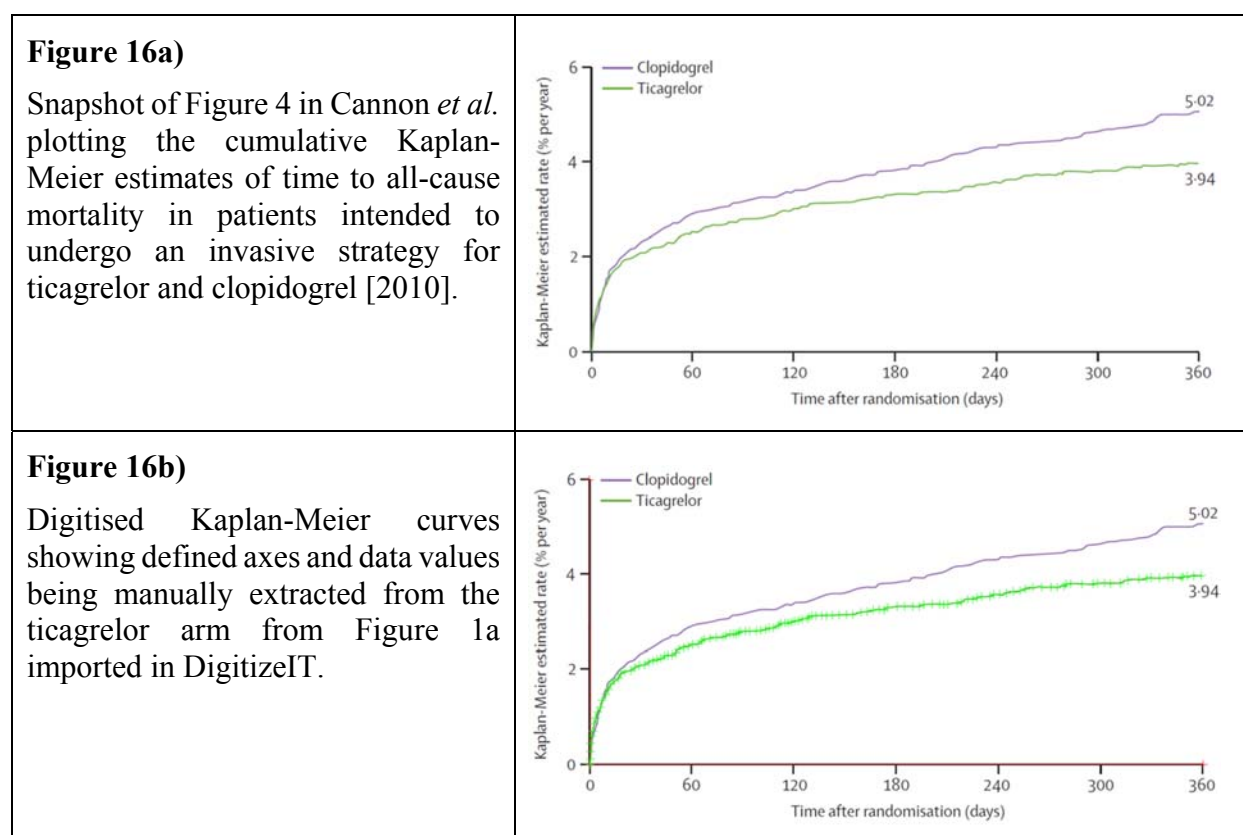
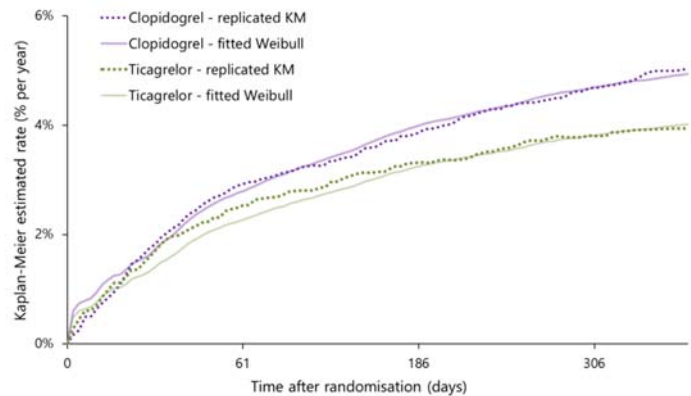


Figure 16c)

Replicated Kaplan-Meier estimates based on exported data files from DigitizeIT for clopidogrel and ticagrelor; and predicted Kaplan-Meier estimates of time to all-cause mortality obtained from fitting a Weibull distribution to 'replicated' dataset.



Unfortunately, for the full PLATO population, Kaplan-Meier estimates were only available for time to the first occurrence of the primary efficacy endpoint (i.e. the composite of MI, stroke and vascular death) and time to first major bleed [Wallentin 2009]. When Kaplan-Meier curves were not reported for an outcome of interest, I relied on the results from the survival analysis performed by AstraZeneca to transform crude proportions based on count data from the PLATO study into risks of events using a Weibull regression [2010a]. Table 6.6 in the manufacturer's submission to NICE summarises the results of the Weibull regression equations for MI, stroke, and death from any cause on the logarithmic scale [AstraZeneca 2010a]. Table 34 includes both the log and exponentiated coefficients.

Table 34 Results from the AstraZeneca Weibull regression for the PLATO study [2010a]

Variable	Coefficient (log)	95% CI (log)	Coefficient	95% CI
MI				
Treatment effect	-0.151	(-0.282, -0.020)	0.860	(0.754, 0.980)
Scale (constant)	-5.202	(-5.373, -5.032)	0.006	(0.005, 0.007)
Shape (Ln Gamma)	-0.907	(-0.971, -0.843)	0.404	(0.379, 0.430)
Stroke				
Treatment effect	0.086	(-0.230, 0.401)	1.090	(0.795, 1.493)
Scale (constant)	-7.392	(-7.852, -6.931)	0.001	(0.000, 0.001)
Shape (Ln Gamma)	-0.791	(-0.945, -0.637)	0.453	(0.389, 0.529)
Death any cause				
Treatment effect	-0.243	(-0.374, -0.112)	0.784	(0.688, 0.894)
Scale (constant)	-5.374	(-5.555, -5.192)	0.005	(0.004, 0.006)
Shape (Ln Gamma)	-0.830	(-0.894, -0.766)	0.436	(0.409, 0.465)

For the PLATO-INVASIVE subgroup, Kaplan-Meier estimates were given for MI, cardiovascular death, stent thrombosis, major bleed, and all-cause mortality [Cannon 2010]. Figures from Cannon *et al.* were digitized as in Figure 16b, before Kaplan-Meier estimates for MI and all-cause death were extracted and converted in (x,y)-data. Calculations were carried out in Excel 2010, as previously described, and the fit of the Weibull distribution was assessed by visual inspection for different values of (t_1) and (t_2). In addition, a simple Weibull regression was run in Stata 13 on the ‘replicated’ datasets for the baseline risk of MI and all-cause death using the code in the Appendix E1 to validate results and obtain a single consistent estimate for the shape and scale parameters. Since Kaplan-Meier estimates were not reported for the rate of stroke in patients intended to be managed invasively and subgroup-specific Weibull coefficients were not provided by AstraZeneca. I used the full population parameter estimates in the simulation study. This assumption seemed reasonable as there was no effect observed on the rate of stroke in either the PLATO or the PLATO-INVASIVE studies at one year [AstraZeneca 2010a].

7.3.1.2 Simulation study design

Independent datasets were simulated for the PLATO and PLATO-INVASIVE populations; i.e. the base case analysis and subgroup analysis, respectively, with the true difference between each treatment’s known effect for each outcome of interest. The starting point for the simulation was to generate a number of patients with an underlying baseline risk [Latimer 2013, 2014]. For the base case, I generated 18,000 observations and assigned each a recruitment time drawn from a uniform distribution between time 1 and 669 days. This interval approximately corresponds to the recruitment period for the PLATO trial between October 2006 and July 2008 [Wallentin 2009]. Similarly, 13,500 patients were generated for the subgroup analysis to reflect the sample size in the PLATO-INVASIVE substudy and recruitment times were assigned using the same distribution [Cannon 2010]. Treatment allocation between ticagrelor and clopidogrel was based on a Bernoulli distribution with probability $p=0.5$ to mimic the 1:1 randomisation ratio of the PLATO trial. Four different simulation scenarios were considered and patients were censored at the end of each follow-up period:

- Scenario 1: 30 days follow-up;
- Scenario 2: 180 days follow-up;
- Scenario 3: 1 year follow-up; and
- Scenario 4: 1 year follow-up only for patients recruited in the first 365 days.

I used the Stata command SURVSIM to simulate survival and time-to event data for both treatment groups [Crowther 2012]. A survival model was fitted to the generated data using the stpm2 command from Lambert and Royston [2009]. Using the ‘simplified’ syntax for stpm2, I fitted a Weibull (option df(1)) proportional hazards model with treatment effect (trt) on the log-cumulative hazard scale (option scale(h)). No time-dependency of effects or transformation of splines were included in this model. Using the predict command, I estimated the survival probabilities and associated 95% CIs at different time points (e.g. timevar(time365)) for clopidogrel (trt 0) and ticagrelor (trt 1). E2 in the Appendix provides the complete ado file for the ‘simplato’ program written to conduct the simulation and survival analysis in Stata; for illustrative purposes E2 includes the all-cause death data for the base case analysis, i.e. in the full PLATO population.

In order to limit sampling variation, a Monte Carlo simulation was run over 1000 iterations and the estimated values were stored after each replication. The results were summarised over the total number of iterations. Summary results can be compared to the ‘true’ values used to simulate the data; for the PLATO study, the published HRs and Kaplan–Meier estimates of the rate of an endpoint at one-year are considered to be the ‘truth’. According to Burton *et al.* this comparison provides “a *measure of the performance and associated precision of the simulation process*” [2006].

In addition, the performance of the simulation was assessed across the four different scenarios in terms of bias, accuracy, and coverage as recommended in Burton *et al.*’s checklist for reporting simulation studies in medical statistics [2006]. I calculated the percentage bias, mean square error (MSE), and coverage and for the scale, shape and log HR (trt) for each simulation scenario in Stata.

The bias is the deviation in a simulated estimate from the ‘true’ value; it can be assessed as the percentage difference between the average simulation estimate ($\bar{\beta}$) and the known estimate (β) [Burton 2006]:

$$\text{Percentage bias} = \left(\frac{\bar{\beta} - \beta}{\beta} \right) * 100$$

The MSE measures the accuracy of simulation and incorporates bias and variability. Using the notation from Burton *et al.*, the MSE is calculated as below:

$$\text{MSE} = \left(\bar{\hat{\beta}} - \beta \right)^2 + \left(SE(\hat{\beta}) \right)^2$$

Where $SE(\hat{\beta})$ is the empirical standard error of the estimate of interest over all simulations. The width of the CIs and the coverage statistics determine the variability in the simulation. The coverage of a CI is the proportion of times the simulated CIs contain—or *cover*—the ‘true’ parameter value; in other words, does the interval achieve the nominal level of coverage? For example, if one assumes normality of the samples, the coverage should approximately equal 95%; i.e. 95% of the samples of 95%CIs should include the ‘true’ value for the estimate of interest β [Burton 2006]. The *inrange* function was used in Stata to obtain the coverage probabilities for key parameters in each simulation scenario, see Appendix E2. Over-coverage—coverage rates greater than 95%—suggests that the standard errors are too large and the results are too conservative. Under-coverage—coverage rates less than 95%—suggests that the standard errors are too small and indicates over-confidence in the results [Burton 2006, Lambert 2014].

7.3.1.3 Subgroup analysis

A second simulation study was performed for a subgroup of ACS patients enrolled in the PLATO study who were identified at randomisation with investigator intent for early invasive strategy, i.e. PLATO-INVASIVE [Cannon 2010]. As previously mentioned in section 7.3.1.2, 13,500 observations were generated for the simulation to reflect the sample of 13,408 (72%) of the 18,624 patients included in the PLATO-INVASIVE substudy.

As discussed in section 4.4.1.3.2 in Chapter 4, NICE considered the invasive subpopulation to be the most appropriately matched to that of the prasugrel phase III trial—TRITON-TIMI 38 [Wiviott 2007]. Therefore, the subgroup analysis was conducted in order to indirectly compare ticagrelor to prasugrel under each simulation scenario; and extend the cost-effectiveness analysis to evaluate ticagrelor vs. prasugrel in ACS patients planned for invasive management. The same simulation design and Weibull survival model were used in the subgroup analysis as in the base case, Appendix E3 adapts the ‘simplato’ program for the PLATO-INVASIVE subgroup—‘simplatoinv’—for all-cause death.

7.3.1.3.1 Indirect treatment comparison

The ITC for ticagrelor vs. prasugrel was solely based on two studies, PLATO-INVASIVE and TRITON-TIMI 38 [Wiviott 2007, Cannon 2010], in which ticagrelor and prasugrel were both respectively compared to clopidogrel. Indirect estimates were calculated in Stata using the following equations:

$$\ln HR_{PT} = \ln HR_{CT} - \ln HR_{CP}$$

$$SE(\ln HR_{PT}) = \sqrt{(SE(\ln HR_{CT})^2 + SE(\ln HR_{CP})^2)}$$

with C: clopidogrel, P: prasugrel, and T: ticagrelor

I used the simulated treatment effect for ticagrelor vs. clopidogrel ($\ln HR_{CT}$) from the invasive subgroup analysis and the log of the HRs reported in Wiviott *et al.* for prasugrel vs. clopidogrel ($\ln HR_{CP}$) for MI, stroke, and all-cause death [2007], see Table 35.

Table 35 Results from Wiviott *et al.* for the TRITON-TIMI 38 study [2007]

Variable	Hazard ratio (prasugrel vs. clopidogrel)	95% CI	Hazard Ratio (log) (prasugrel vs. clopidogrel)	Standard error (log)
MI	0.76	(0.67, 0.85)	-0.274	0.061
Stroke	1.02	(0.71, 1.45)	0.020	0.182
Death any cause	0.95	(0.78, 1.16)	-0.0513	0.101

7.3.2 Results

7.3.2.1 Base case analysis

Baseline survival probabilities of MI, stroke and all-cause death for the clopidogrel group were obtained based on the Weibull coefficients and HRs in Table 34 in section 7.3.1.; these are presented in Appendix E4. The average HRs for ticagrelor vs. clopidogrel and 95% CI from the 1000s replications are presented in Table 36 for the three outcomes of interest under all four simulation scenarios.

Table 36 Summary HRs and 95% CIs for the base case simulation

Variable	Hazard ratio (ticagrelor vs. clopidogrel)	Lower 95%CI	Upper 95% CI
MI			
(1) 30 days follow-up	0.864	0.555	1.157
(2) 180 days follow-up	0.861	0.686	1.064
(3) 1 year follow-up	0.859	0.701	1.062
(4) after 1year recruitment	0.861	0.636	1.120
Stroke			
(1) 30 days follow-up	1.150	0.415	2.789
(2) 180 days follow-up	1.113	0.562	2.232
(3) 1 year follow-up	1.104	0.646	1.987
(4) after 1year recruitment	1.117	0.578	2.170
Death any cause			
(1) 30 days follow-up	0.792	0.559	1.150
(2) 180 days follow-up	0.786	0.621	1.006
(3) 1 year follow-up	0.786	0.651	0.936
(4) after 1year recruitment	0.787	0.606	1.013

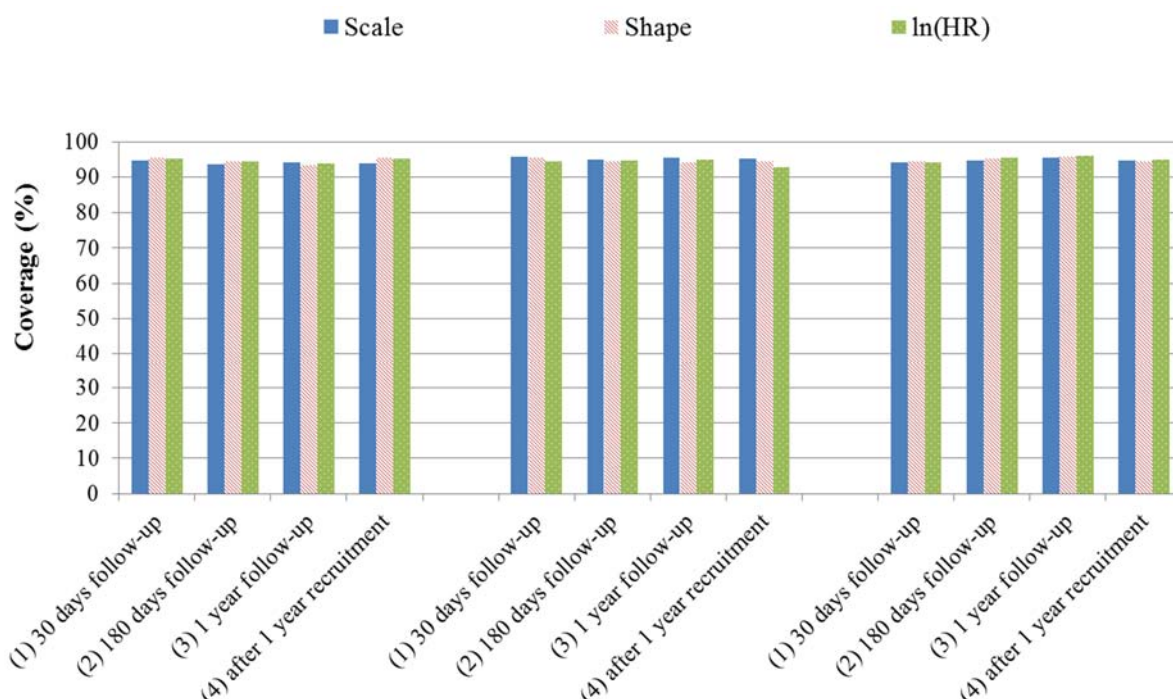
Across all scenarios, the point estimates for the HRs were stable and their associated uncertainty decreased as the length of follow-up increased from 30 days, to 180 days, and up to 1 year. Similarly, the 95%CI were slightly wider for HRs estimated after only 1 year of patient recruitment (scenario 4) compared to the simulated results from the full population at 1 year follow-up (scenario 3).

The percentage biases and MSE are summarised in Table 37 and Figure 17 illustrates the coverage for the three outcomes of interest under all four simulation scenarios. Overall, bias was considered low (< 2%) and there was a good coverage of consistently over or slightly under 95% for all outcomes and simulation scenarios. The only exception was for the estimate of treatment effect for stroke at 30 days follow-up which showed a 16.34% bias and 0.08 MSE, but 94.5% coverage.

Table 37 Summary percentage bias and MSE from the base case simulation

Variable	MI		Stroke		Death any cause	
	% bias	MSE	% bias	MSE	% bias	MSE
Scale						
(1) 30 days follow-up	-0.07	0.00	-1.08	0.00	-0.73	0.00
(2) 180 days follow-up	0.10	0.00	0.19	0.00	-0.30	0.00
(3) 1 year follow-up	0.27	0.00	0.61	0.00	-0.34	0.00
(4) after 1 year recruitment	0.19	0.00	1.28	0.00	-0.36	0.00
Shape						
(1) 30 days follow-up	0.27	0.00	1.55	0.00	0.61	0.00
(2) 180 days follow-up	0.18	0.00	0.81	0.00	0.29	0.00
(3) 1 year follow-up	0.08	0.00	0.52	0.00	0.28	0.00
(4) after 1 year recruitment	0.29	0.00	1.13	0.00	0.41	0.00
ln(HR)						
(1) 30 days follow-up	0.54	0.01	16.34	0.08	-1.33	0.01
(2) 180 days follow-up	1.11	0.01	4.50	0.03	0.17	0.01
(3) 1 year follow-up	1.68	0.00	0.31	0.03	-0.29	0.00
(4) after 1 year recruitment	1.80	0.01	-0.03	0.05	-0.01	0.01

Figure 17 Coverage across simulation scenarios for shape, scale, and treatment effect parameters



7.3.2.2 Subgroup analysis

The Weibull shape and scale parameters for the ‘replicated’ Kaplan-Meier estimates of the baseline rate of MI and all-cause death in the PLATO-INVASIVE population

obtained from the Weibull regression are presented in Table 38. Table 38 also includes the HRs extracted from Cannon *et al.* for ticagrelor vs. clopidogrel in the invasive population and the assumed scale and shape parameters for the stroke analysis taken from the base case.

Table 38 Results from the Weibull regression for the PLATO-INVASIVE study

Variable	Coefficient (log)	95% CI (log)	Coefficient	95% CI
MI				
Treatment effect	-0.22	(-0.37, -0.08)	0.80	(0.69, 0.92)
Scale (constant)	-4.993	(-5.139, -4.849)	0.007	(0.006, 0.008)
Shape (Ln Gamma)	-0.899	(-0.979, -0.826)	0.407	(0.376, 0.438)
Stroke				
Treatment effect	0.07	(-0.25, 0.41)	1.08	(0.78, 1.50)
Scale (constant)	-7.392	(-7.852, -6.931)	0.001	(0.000, 0.001)
Shape (Ln Gamma)	-0.791	(-0.945, -0.637)	0.453	(0.389, 0.529)
Death any cause				
Treatment effect	-0.21	(-0.39, -0.05)	0.81	(0.68, 0.95)
Scale (constant)	-5.322	(-5.418, -5.226)	0.005	(0.004, 0.005)
Shape (Ln Gamma)	-0.893	(-0.944, -0.844)	0.409	(0.389, 0.430)

The fit of the Weibull distribution was assessed for MI and all-cause death by visual inspection and found to be appropriate, see Figure 18a and 18b. In addition, AstraZeneca justified the use of the Weibull survival model as it allows the hazard rates to change as time elapses from randomisation in a similar fashion as the risk of events declined over time in the PLATO study [2010a].

Figure 18a) MI

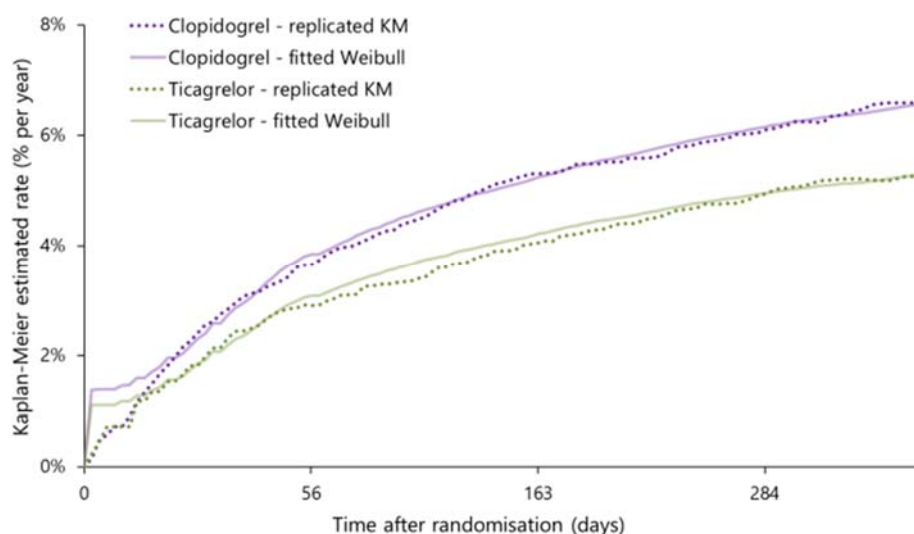


Figure 18b) Death any cause

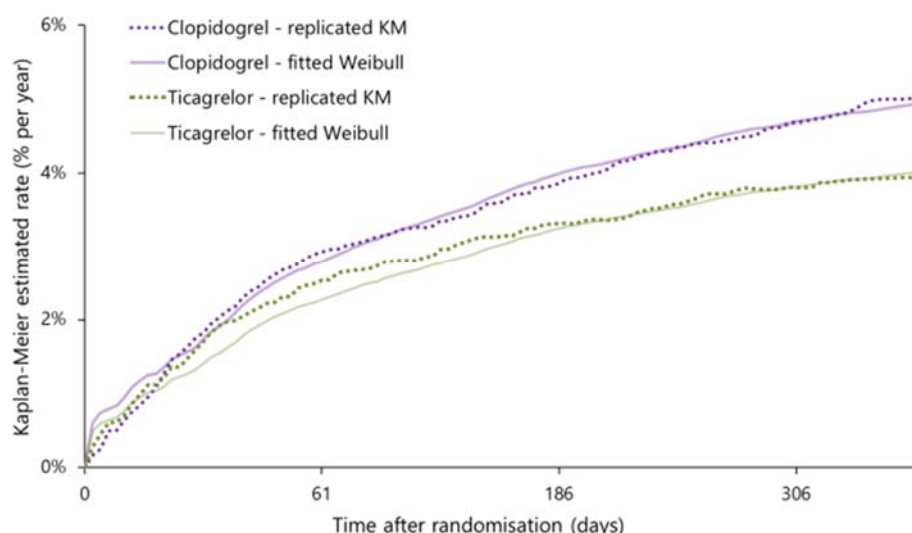


Table 39 summarises the HRs and 95% CI for ticagrelor vs. prasugrel averaged over 1000 iterations for the subgroup analysis including the ITC for the three outcomes of interest under all four simulation scenarios. Indirect estimates obtained from the simulation can be compared to the published results from Biondi-Zoccai *et al.* [2011] described in section 4.4.1.3.2 in Chapter 4. For example, the simulated all-cause death HR for ticagrelor vs. prasugrel at 1 year follow-up—0.85 (0.68-1.04)—was found to be similar to the HR reported by Biondi-Zoccai *et al.* of 0.82 (0.65, 1.04) [2011].

Table 39 Summary HRs and 95% CIs for the subgroup simulation

Variable	Hazard ratio (ticagrelor vs. prasugrel)	Lower 95%CI	Upper 95% CI
MI			
(1) 30 days follow-up	0.841	0.590	1.180
(2) 180 days follow-up	0.842	0.683	1.024
(3) 1 year follow-up	0.842	0.703	0.992
(4) after 1year recruitment	0.846	0.637	1.096
Stroke			
(1) 30 days follow-up	1.196	0.479	3.754
(2) 180 days follow-up	1.172	0.653	2.296
(3) 1 year follow-up	1.164	0.749	1.999
(4) after 1year recruitment	1.174	0.621	2.906
Death any cause			
(1) 30 days follow-up	0.861	0.634	1.193
(2) 180 days follow-up	0.855	0.694	1.104
(3) 1 year follow-up	0.853	0.677	1.036
(4) after 1year recruitment	0.854	0.599	1.170

Simulation results were largely unbiased with equal or less than 5% bias in the scale, shape, and treatment effect parameters across all four simulation scenarios, see Table E2 in the Appendix. Coverage was also considered to be satisfactory for all scenarios as illustrated in Figure E1 in the Appendix.

7.4 Cost-effectiveness analysis

7.4.1 Methods

In order to assess the impact of the simulated efficacy endpoints on the cost-effectiveness of ticagrelor at different time-points, I extended the economic model specified in section 4.4 of Chapter 4. The model was reconstructed in Excel 2010 based on the evidence review group's critique of the manufacturer's economic evaluation of ticagrelor as a treatment option for ACS patients in England and Wales [Bagust 2011]. The model adopted an NHS/PSS perspective and included a full population analysis evaluating ticagrelor vs. clopidogrel and a subgroup analysis vs. prasugrel. The following efficacy inputs were used to populate the model for the four different simulation scenarios holding all other parameters constant:

- baseline probabilities (and associated 95%CI) of having a non-fatal MI, of having a non-fatal stroke, or dying from any cause (see Table E1 in the Appendix);
- HRs for ticagrelor vs. clopidogrel (and associated standard errors) for non-fatal MI, non-fatal stroke, and all-cause death (see Table 36); and
- HRs for ticagrelor vs. prasugrel (and associated standard errors) for non-fatal MI, non-fatal stroke, and all-cause death (see Table 39).

Section 4.4.1 provides a detailed description of the economic evaluation including the model structure, input data, and model assumptions. An additional macro was created to automate a PSA for each simulated replication over 1000 iterations, i.e. the 1000 simulation replications were each run 1000 times to obtain summary cost-effectiveness outputs. The cost-effectiveness analysis time horizon was 40 years for all the simulation scenarios.

7.4.2 Results

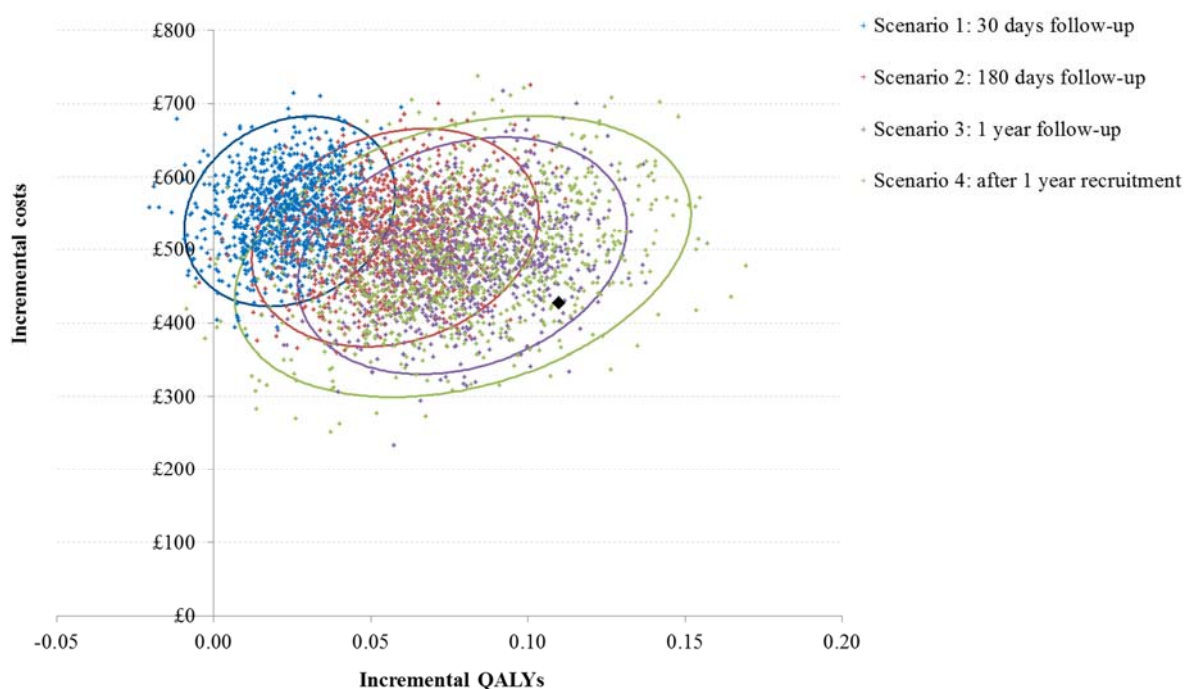
7.4.2.1 Base case analysis

The mean results from the PSA are summarised in Table 39 for the four simulation scenarios in the base case. The ICERs for ticagrelor vs. clopidogrel decreased from £22,853 to £6,228 between using the simulated efficacy estimates at 30 days follow-up and 1 year follow-up, respectively. At one year, the cost-effectiveness results projected from the simulated data can be compared to the trial-based ICERs reported in Section 4.4.2. The base case ICER at 1 year follow-up for the full ACS population was £6,228 in the simulated example compared to £3,443 in the original cost-effectiveness analysis. The simulation study allowed for a wider sampling variation in the clinical model inputs than the model built-in PSA, this could explain the marginally higher mean total life year (LYs) and QALYs seen in scenario 3 in Table 40 compared to the trial-based results, see Table 15 in Chapter 4. Figure 19 plots the base case PSA results for the four simulation scenarios; the ellipses delimit the region of 95% confidence.

Table 40 ICERs for ticagrelor vs. clopidogrel from the reconstructed economic model for the base case simulation

Scenario	Total costs	Total LYs	Total QALYs	ICER (£/QALYs)
(1) 30 days follow-up				
Clopidogrel	£17,709	8.06	6.71	-
Ticagrelor	£18,262	8.09	6.72	£22,853
(2) 180 days follow-up				
Clopidogrel	£17,854	7.87	6.55	-
Ticagrelor	£18,371	7.95	6.61	£8,947
(3) 1 year follow-up				
Clopidogrel	£17,952	7.75	6.45	-
Ticagrelor	£18,444	7.85	6.53	£6,228
(4) after 1 year recruitment				
Clopidogrel	£17,951	7.75	6.45	-
Ticagrelor	£18,443	7.85	6.53	£6,200

Figure 19 Cost-effectiveness plane for ticagrelor vs. clopidogrel for all simulation scenario



The economic model results show that ticagrelor was a cost-effective and acceptable alternative to clopidogrel at 180 days follow-up; as well as, after only one year recruitment, at a £20,000 WTP threshold. Figure 20 illustrates the probability of ticagrelor being cost-effective for a range of WTP ceiling ratio up to £40,000; Table 41 summarises these cost-effectiveness probabilities for ticagrelor at £10K, £20K, and £30K.

Figure 20 Cost-effectiveness acceptability curves for ticagrelor and clopidogrel for all simulation scenarios

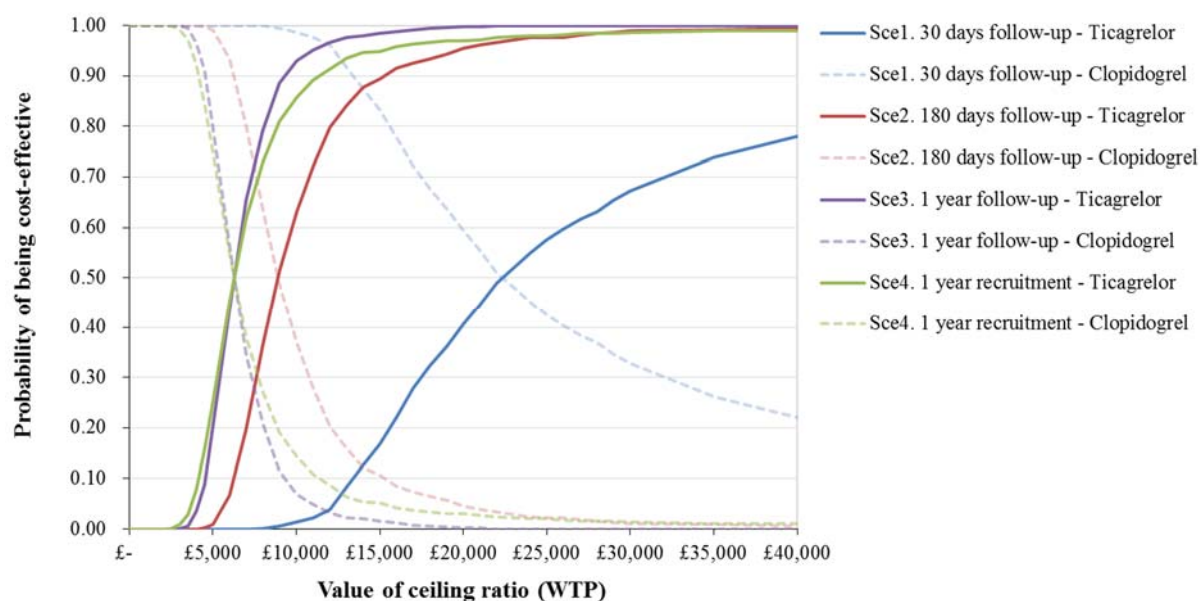


Table 41 Cost-effectiveness probabilities for ticagrelor for the base case simulation

Scenario	Willingness to pay threshold		
	£10,000	£20,000	£30,000
(1) 30 days follow-up	0.01	0.41	0.67
(2) 180 days follow-up	0.63	0.96	0.99
(3) 1 year follow-up	0.93	1.00	1.00
(4) after 1year recruitment	0.87	0.97	0.99

7.4.2.2 Subgroup analysis

Ticagrelor was also found to be highly cost-effective vs. prasugrel under all four simulation scenarios considered for the subgroup of ACS patients intended at randomisation for invasive management. Regardless of the length of trial follow-up, the probability of ticagrelor being cost-effective compared to prasugrel was higher than 90% at a £20K ceiling WTP threshold. Table 42 summarises the PSA means for the total costs, LYs, QALYs, and ICERs in the subgroup analysis. Similarly to the base case, ICERs based on the simulation can be compared for reference to the original cost-effectiveness analysis for ticagrelor vs. prasugrel based on the ITC estimates from Biondi-Zoccai *et al.* [2011]. Table 18 in Chapter 4 presents an ICER for the PLATO-INVASIVE subgroup at one year follow-up of £3,882 which is analogous to the £2,241 ICER estimated in scenario 3.

Table 42 ICERs for ticagrelor vs. prasugrel from the reconstructed economic model for the subgroup analysis simulation

Scenario	Total costs	Total LYs	Total QALYs	ICER (£/QALYs)	Probability cost-effective at £20,000
(1) 30 days follow-up					
Prasugrel	£23,131	11.13	9.20	-	-
Ticagrelor	£23,288	11.20	9.26	£2,474	0.91
(2) 180 days follow-up					
Prasugrel	£23,131	11.13	9.20	-	-
Ticagrelor	£23,283	11.21	9.26	£2,297	0.97
(3) 1 year follow-up					
Prasugrel	£23,135	11.13	9.20	-	-
Ticagrelor	£23,386	11.21	9.26	£2,241	0.99
(4) after 1 year recruitment					
Prasugrel	£23,133	11.13	9.20	-	-
Ticagrelor	£23,287	11.21	9.26	£2,318	0.96

7.5 Discussion

Data simulation based on the large-scale PLATO trial was used to predict relative efficacy estimates at different follow-up times and for different recruitment strategies in order to assess HTA outcomes for ticagrelor vs. clopidogrel and ticagrelor vs. prasugrel, in their respective indicated population, under ‘virtual’ time constraints for early drug approval. Ultimately, the aim of the simulation study was to explore whether a trial design with a shorter follow-up, smaller sample size, and/or the publication of an interim analysis could have resulted in a positive NICE recommendation for ticagrelor in ACS at an earlier date.

Latimer *et al.* highlight that the values for the model parameters should be selected “*in order to ensure that the simulated data suitably represent[s] the type of dataset that the study was designed to replicate*” [2013]. The simulation study was designed to reflect clinical practice in England and observations were generated matched on the PLATO and PLATO-INVASIVE ACS patient populations [Wallentin 2009, Cannon 2010]. IPD was simulated from published Kaplan-Meier estimates and fitted with a Weibull survival model to predict treatment effects over time. The goodness-of-fit of the Weibull distribution was assessed for the three outcomes of interest by visual inspection and was justified by AstraZeneca given the decreasing risk of events in the PLATO trial as time elapses from randomisation [2010a].

The simulation study performed well in terms of bias, accuracy, and coverage. Burnham and Anderson comment that the ‘best’ model is achieved by:

“properly balancing the errors of over-fitting and under-fitting [so that] bias and variance are controlled to achieve confidence interval coverage at approximately the nominal level [0.95] and where interval width is at a minimum” [1998: p25]

For both the base case and subgroup analysis, all four simulation scenarios produced low percentage biases and provided good coverage for shape and scale parameters, as well as the log HRs, implying the simulation design and Weibull model were appropriate. This was to be expected as the same Weibull model was used to generate and fit the data.

Despite increased uncertainty around the HRs and baseline risks of rate of MI, stroke, and all-cause death estimated from immature trial data; results from the economic evaluation demonstrated ticagrelor was a cost-effective and acceptable treatment option both vs. clopidogrel in a broad ACS population and vs. prasugrel in invasively managed ACS patients. In retrospect, a submission based on interim trial results prior to the original planned analysis at one year follow-up could have led to an earlier NICE recommendation for ticagrelor in England and Wales. For example, a clinical and cost-effectiveness analysis based on simulated time-to-event data at 180 days follow-up revealed ticagrelor to be 95.50% cost-effective at £20K in the base case and 97.30% cost-effective at £20K in the subgroup analysis. Furthermore, the results from the simulation scenario 4, which evaluated the impact of shortening recruitment times to 365 days instead of the 669 days in the PLATO study, also suggest that a favourable recommendation for ticagrelor treatment in ACS patients could have been achieved prior to 2011.

7.5.1 Caveats and model limitations

A general limitation of simulation studies is that the prediction estimates are likely to always be linked in some way to the chosen data generating process [Latimer 2013]. A number of assumptions were made in the design of this simulation study that could undermine both the relative effectiveness and cost-effectiveness assessment of ticagrelor. First, as pointed out by Bagust *et al.*, the assumption of proportional hazards and a common Weibull function for all the outcomes of interest and treatments investigated may not be sufficient to accurately represent the PLATO trial data [2011]. A more flexible parametric survival model could have been used to analyse the time-to-event data but this

was not thought to be warranted to meet the specific chapter objectives. Recruitment times were assumed to be normally distributed and administrative censoring at the end of the follow-up period was also assumed for each simulation scenario. Such assumptions could in part explain the greater uncertainty in the simulated results at one year follow-up (scenario 3) than those reported in the PLATO trial with a non-significant HR (95%CI) for MI of 0.86 (0.70, 1.06) compared to 0.84 (0.75, 0.95), respectively [Wallentin 2009]. I also assumed that the treatment effect was fixed and did not incorporate uncertainty around the point estimate extracted from the PLATO trial. This assumption was justified as I sought to recreate ‘realistic’ IPD for the PLATO trial; however, uncertainty would need to be taken into account if simulation was used to predict outcomes from future trials in ACS anchored around the PLATO results. Additional assumptions include the number of observations generated and the number of replications in the simulation study; as well as the modelling assumptions in the cost-effectiveness analysis already explained in Chapter 4’s discussion. Testing these assumptions and increasing the number of replications could help reduce the uncertainty and sampling error in simulated results.

Section 4.5 in Chapter 4 also highlights the contention in indirectly comparing ticagrelor to prasugrel due to the differences in patient populations in PLATO-INVASIVE and TRITON-TIMI 38. Namely, although the PLATO-INVASIVE substudy included patients identified at randomisation with investigator intent for an invasive strategy and undergoing early angiography; only 77% of this cohort actually underwent PCI, whereas TRITON-TIMI 38 represented a pure PCI-only patient cohort [AstraZeneca 2010a]. Although ticagrelor was compared to prasugrel in the manufacturer’s economic evaluation following NICE’s scope, AstraZeneca and the evidence review group did not endorse a formal ITC [2010a, Bagust 2011]. Patient heterogeneity as well as differences in the loading dose of clopidogrel and assessment of MI between the two trials synthesised in the subgroup analysis may have biased model results and subsequent claims of ticagrelor’s superiority. However, given the high probabilities of ticagrelor being cost-effective in invasively managed ACS patients across a range of WTP thresholds, its recommendation as an alternative treatment remains robust for all four simulation scenarios considered.

Overall the simulation showed that ticagrelor had a high probability of being cost-effective, and thus receiving a positive recommendation for reimbursement by NICE, based on the results from a shorter trial (i.e. 180 days follow-up instead of one year)—or

an interim analysis at 6 months—or a smaller trial (i.e. with an approximate sample size of 9,800 patients recruited in the first year). Further research would be needed to determine whether a combination of both shorter follow-up time and smaller sample size would have been a feasible study design without compromising ticagrelor's probability of being cost-effective for a willingness to pay threshold between £20,000 and £30,000 in the UK.

Chapter 8

Conclusions and further work

8.1 Summary of thesis

Assessing the value of a medical intervention has become routine practice and in several countries, including the UK, it is a gold standard for market access and coverage decisions. HTA is used to demonstrate the value of a health technology in terms of cost-effectiveness by combining both the incremental health benefits gained by a new product and the incremental costs. Recent developments in the field of HTA have focused on improving the scope, methods, and reporting of HTA to ensure it can address real-world decision problems [Hailey 2003, Draborg 2005, Sorenson 2008]. However, little has been published on how adaptive HTA is to real-world changes. Indeed, ongoing policy changes to the regulation and reimbursement of medicines in the UK, and beyond, are increasing pressures on health decision-makers to ensure timely patient access to safe and effective drugs at a price that still guarantees the best value for money. In my thesis, I aimed to examine how facilitated regulatory pathways for health technologies, such as accelerated approval or adaptive licensing, may influence the practice of HTA, including health policy decision-making on the basis of economic decision models. In particular, I explored how new evidential and methodological requirements for early approval could impact the relative effectiveness and cost-effectiveness assessment of a product earlier in its development cycle.

Chapter 1 highlighted my research objectives and defined the layout of the thesis. I introduced the concept of HTA and described the recent regulatory schemes proposed by the FDA and EMA to promote early access for patients to promising and innovative therapies in areas of unmet need. I also briefly presented initiatives developed in parallel by reimbursement authorities and payers, such as MEAs and CED, to ensure that regulatory efforts are matched by national health systems and access is not delayed.

In Chapter 2, I discussed key learnings from the literature on the evidential issues associated with early drug evaluations and the statistical methods available to potentially

address these issues in the context of HTA. The literature reviews highlighted a number of evidentiary trade-offs implicit to the assessment of products earlier in their development cycle. In particular, issues related to trial designs, non-randomised evidence, subgroup analyses and surrogate endpoints. It also demonstrated that under the potential new remit of HTA (i.e. ‘early’ assessments), assessors will have to be more sensitive to changing parameters and environment and more willing to accept greater levels of uncertainty. Following a comprehensive search of HTA and pharmacoeconomic guidelines, I also examined the current discrepancies and gaps in HTA ‘best practice’ worldwide to identify specific research areas most likely to improve the responsiveness of HTA to regulatory changes. I considered three challenges to present and future HTA practice: i) the selection of relevant patient subgroups and comparators for appraisal, ii) the use of specific search strategies to identify indirect evidence for NMA; and iii) bias adjustment techniques to include observational data in evidence synthesis.

Chapter 3 provided a brief overview of the methods used in HTA including meta-analysis, NMA, and economic evaluation. Bayesian methods and their application to HTA were also described, including the flexibility that they bring to often complex decision problems. In the chapter discussion, I concluded that these methods have now become standard practice for HTA worldwide and highlighted recent research efforts to further develop these methods in the context of future assessments.

Using the example of ticagrelor for the treatment of ACS, Chapter 4 assessed the impact on the relative effectiveness and cost-effectiveness of ticagrelor when two different approaches to the selection of relevant patient subgroups and comparators were taken. Based on the conflicting recommendations that NICE and IQWiG made in 2011 regarding the reimbursement of ticagrelor in a subgroup of ACS patients undergoing primary or delayed PCI, I considered two different subgroup analyses to allow for the comparison of ticagrelor vs. prasugrel. Both subgroup analyses were modelled using an NHS/PSS perspective to allow for the evaluation of a UK-based ‘counterfactual’ scenario, not undertaken by NICE, but centred on Germany’s IQWiG interpretation of the clinical evidence. The underlying issue was that at the time of ticagrelor’s assessment there was little evidence to support a comparison between ticagrelor and prasugrel, and the use of indirect evidence was contentious. Nonetheless, prasugrel was considered to be a relevant comparator for HTA and its inclusion in the economic model was required by NICE. My analysis showed that contextual factors, seemingly unrelated to evidential and

methodological issues, can also impose constraints on the evidence base and significantly influence HTA outcomes. This example demonstrated that defining a decision problem and simply interpreting the available evidence can be highly subjective, especially in the presence of insufficient or ambiguous clinical findings, and that this should be acknowledged.

Chapter 5 presented a second example to explore the impact of study identification methods and network size on NMA. Building on the latest NICE VTE technology appraisal for apixaban [2012a], I re-analysed the relative effectiveness of all recommended pharmacological VTE prophylaxis for adult patients undergoing TKR in England and Wales. Different network sizes were based on successive searches and compared to a replicated base case obtained from an adjusted ITC of apixaban, dabigatran etexilate, enoxaparin and rivaroxaban. The resulting comparative estimates were also inputted in an economic model emulating the cost-utility analysis performed by manufacturers for the apixaban NICE submission to investigate the potential impact of network sizes on decision-making. Overall, this chapter aimed to use a specific search methodology to identify indirect evidence to test whether increasing the network size could strengthen the NMA and reduce the uncertainty in the pooled effect estimates. This method could be applied in an early evaluation context, for example, if only a limited number of studies were initially included for comparison, could extending the network of evidence be a potential solution to maximise the decision space? The conclusions from this example in VTE remain very case-specific and little additional information was introduced to the NMA beyond first-order searches. However, paradoxical results were obtained by Hawkins *et al.* in their application of these methods to NSCLC [2009a]; thus, further work is needed to demonstrate the added value of different network sizes and shapes in NMA and to generalise findings across disease areas.

A final case study was conducted in Chapter 6 to investigate Bayesian statistical adjustment methods to combine randomised and non-randomised data when comparing PVP vs. OPM to treat osteoporotic VCFs. Long-term claims data from the USA and Germany were used to inform prior beliefs in a Bayesian meta-analysis of mortality risks following vertebral augmentation surgery. Two methods were considered—the power transform prior and the bias allowance—to adjust a fixed effect meta-analysis model of sparse RCT data. In this example, PVP was cost-effective compared to OPM irrespective of a mortality benefit if the willingness to pay threshold was between £20,000 and

£30,000 per additional QALY gained. However, allowing for observational evidence to be considered—even accounting for a varying degree of bias—noticeably reduced the uncertainty around the PVP vs. OPM ICERs. Chapter 6 illustrated how Bayesian methods can be used in HTA to explicitly account for prior knowledge of a treatment effect, in this case the reduced mortality risk of operated patients following an osteoporotic VCF. Such work, and potential extensions to NMA or multivariate analysis, may be very relevant to the early assessment of new technologies when RCT data is insufficient and observational data is available. Nonetheless, it remains unclear how the addition of a subjective judgement in the assessment of relative effectiveness, in the form of a bias allowance, will be perceived by decision-makers.

Lastly, Chapter 7 extended the analytical example of ticagrelor in ACS using a simulation approach to predict the impact of different effect sizes at different time-points during the drug development process on HTA outcomes. The rationale for the simulation was to reconstruct an individual patient dataset for the large registration trial comparing ticagrelor to clopidogrel, which I then truncated at different follow-up and recruitment times to mimic the possible conditions of early approval and HTA. The resulting relative treatment effects were then fed into an economic model to ultimately assess whether a trial design with a shorter follow-up, and/or smaller sample size, and/or the publication of an interim analysis could have resulted in a positive NICE recommendation for ticagrelor in ACS at an earlier date. In the case of these re-analyses, it could have been, but with a number of caveats.

8.2 Caveats and limitations

This thesis makes a contribution to the ongoing research agenda on the role of HTA and HTA methods in healthcare decision-making driven by a wide range of stakeholders on the local, national, and international scene. Over the past two decades, European and international initiatives have increased in this area, as the role of HTA continues to expand and evolve, and the demand for high quality yet affordable healthcare intensifies. For example, collaborations such as EUnetHTA seek to provide a supportive and sustainable network of European HTA organisations and develop a framework for joint assessments to facilitate the efficient use of resources and promote best practice [EUnetHTA 2016]. Moreover, Berntgen *et al.* also summarise recent collaborative efforts between EUnetHTA and the EMA to improve the contributions of regulatory assessments to REA and HTA in Europe [2014]. The Innovative Medicine Initiatives, in partnership with the

European Commission and the European Federation of Pharmaceutical Industries and Associations (EFPIA), aims to speed up the development of better and safer medicines for patients and stimulate innovation [IMI 2016]. Additionally, national programmes such as the NIHR HTA programme and the NICE DSU in the UK are funded to research the effectiveness, costs and broader impact of healthcare treatments, as well as to foster new methodological developments in HTA. Given recent policy changes, the scope of HTA is no longer confined to post-market assessments of product value, but extends from pipeline to clinical practice with HTA agencies becoming key stakeholders in the early dialogue with industry as well as in price negotiations and implementation strategies.

My work aimed to understand and evaluate the evidential and methodological challenges facing HTA in the context of early assessments of new health technologies. I focused on three issues which I deemed both relevant to the present and future practice of HTA in the UK and beyond. However, this research area remains relatively unexplored and many issues are still to be considered. For example, although identified in my literature review, I did not consider issues associated with surrogate outcomes nor did I choose to evaluate extrapolation as a potential methodological solution to limited clinical evidence. One reason for this is the existing and ongoing research projects on both topics [Taylor 2009, Latimer 2011], and their apparent acceptance by international HTA agencies.

Each chapter provides a summary of the caveats and limitations specific to each analysis; but one overarching limitation of the work presented in this thesis is that it is largely case-study specific. Most of the analyses were retrospective and UK-centric using examples from past NICE appraisals to explore new or unresolved issues. Examples were limited in their scope and used to answer a targeted research question; therefore, it is difficult to generalise any chapter conclusions. The HTA simulation attempted to address this limitation and provide a more flexible approach to analyse the data under different scenarios of early assessment. Unfortunately, I was unable to obtain IPD which could have been helpful to validate the simulation results, as well as to explore a ‘real-life’ example and how decisions could have been influenced by evidence generation and evidence synthesis throughout the drug development process.

In addition, each example required making some simplistic assumptions and in most instances, case studies could have been extended to explore additional issues or more sophisticated methods. Note that methodological development was not *per se* prioritised

in this thesis but rather the translation and application of existing methods to address issues raised by early HTA.

8.3 Further work

Case studies are necessarily stylised to an extent; however, in practice many of the issues I explored would occur simultaneously within a single HTA. Further work could investigate how a combination of methods performs to assess a new product with several issues associated with immature and incomplete data. For example, an NMA using observational data to maximise the network evidence base, synthesising multiple outcomes and including intermediate and final endpoints. Although not currently widely used in HTA research, more complex simulation studies could be used to address such challenges.

Furthermore, Walker *et al.* point out that early HTA should not only seek to assess the expected value of a new health technology, but should consider the value of further research, the anticipated effect of coverage decisions on further research, and the costs associated with reversing such decisions [2012]. Undeniably, by allowing earlier access to medicines, decision-makers are implicitly agreeing to more ‘bad’ decisions. Proponents of early or conditional drug approval will claim that this is a calculated risk that patients are often willing to take particularly if no alternative treatment is available [Eichler 2012]. Recent experience with CMAs and approvals under ‘exceptional circumstances’ by the EMA has shown that the risk to patients may not be substantiated. Arnadottir *et al.* compared these expedited schemes with standard procedures and did not find drugs receiving ‘accelerated’ approval to be associated with a higher probability of serious safety issue and none were withdrawn from the market [2011]. However, whether this is a risk worth taking by decision-makers is a question that requires additional research.

As highlighted in discussion section 2.2.4 of the early drug evaluations literature review, expert groups and key opinion leaders have debated the feasibility and practicality of early HTA and MEA/CED schemes [Douw 2004, Trueman 2010, Baird 2013, Husereau 2014]. Amongst the concerns voiced were how all relevant stakeholders would interact and collaborate to achieve both conditional and subsequently full approval/reimbursement; how early HTA recommendations would be implemented in clinical practice—as most schemes until now have focused on conditional coverage ‘only

in research’; how products should be identified and prioritised for early assessment; and how a decision could be reversed following re-appraisal.

Value of information research could shed light on the added value of further research as well as monetise the risk of ‘false positive’ and ‘false negative’ decisions. Further research is needed on how health outcomes should be monitored in post-marketing authorisation studies and whether similar safeguards to those currently used for drug risk surveillance could be applied to effectiveness measures [Tubach 2011]. In addition, there is the related issue of who pays for and conducts post-reimbursement data gathering. New guidelines for the Cancer Drug Fund in the UK and the potential for linked electronic health records may offer a more efficient way of monitoring outcomes. The UK could also draw on other countries experience under MEAs, such as in the Netherlands, where industry funds cancer registries to do data collection post-market. Lastly, if withdrawing a product for poor cost-effectiveness is unlikely; pricing could be a way of controlling the market. Beyond current MEA schemes that offer price rebates or financial settlements, more work could be done to evaluate value based pricing and what it entails for HTA.

8.3.1 Recommendations for future HTA practice

NICE and the UK have always been, and remain, at the forefront of HTA research, continuously aiming to provide pioneering guidance on how to develop and apply new methodologies in HTA to address areas of uncertainty. I would recommend additional research be conducted on the role of HTA in promoting timely access for patients to new medicines and on the HTA methods required to tackle evidentiary trade-offs inherent to the early assessment of product value. Further work should build on existing initiatives by NICE and current research, such as Claxton *et al.* report on the use of health technologies only in the context of an appropriately designed programme of evidence development, sponsored by the Health Technology Assessment NIHR HTA programme [2012]. Indeed, several parallels can be drawn from Claxton *et al.* work on OIR/AWR recommendations by NICE [2013], amongst others [Jönsson 2015, Longworth 2013, Claxton 2011, Briggs 2010, Chalkidou 2007], both in terms of the specific data considerations for the assessment of new technologies likely to benefit from ‘conditional’ access and additional evidence generation, and in understanding how best to measure the potential added value, risks and costs of decisions made under uncertainty.

8.4 Conclusion

This thesis has explored the concept of early HTA and what impact evidential and methodological requirements for the ‘accelerated’ assessment of new health technologies could have on relative effectiveness and cost-effectiveness analysis. I have examined how the selection of relevant patient subgroups and comparators for appraisal, specific search strategies to identify indirect evidence for NMA, and bias adjustment techniques to include observational data in evidence synthesis could influence HTA recommendations through the impact that these might have on economic decision models. I simulated a trial-based cost-effectiveness analysis to predict the impact of immature and incomplete clinical evidence on HTA outcomes. Despite stated caveats and limitations, my work highlights that HTA can adapt to meet new evidentiary standards and build on existing methodologies to inform public health policy decision-making.

Appendix A

Table A1 Search strategy for OVID (Medline®/Medline-In-Process®/EMBASE) based on preliminary searches

#	Search terms	Hits
1	exp Guideline/	21,778
2	guidelines as topic.sh.	27,193
3	guideline.pt,sh.	15,181
4	(guid* or (good adj2 practice\$1) or (best adj2 example\$1) or methodolog* or recommendation\$1 or tool* or check* or handbook\$1 or standard\$1 or principle\$1).ab,ti.	3,258,304
5	1 or 2 or 3 or 4	3,277,922
6	*Biomedical Technology/ or *Medical Technology/	19,419
7	*Economics, Pharmaceutical/ or *pharmacoeconomics/	3,327
8	(pharmaceutical\$1 or drug\$1 or medicine\$1 or pharmacoeconomic\$1 or pharmaco-economic\$1).ab,ti.	2,730,064
9	((health or healthcare or health care or medical or single or multiple) adj (technolog\$3 or intervention\$1)).ab,ti.	33,698
10	(assessment\$1 or appraisal\$1 or evaluation\$1).ab,ti.	2,510,923
11	(8 or 9) adj 10	7,732
12	(6 or 7) and 10	2,383
13	(HTA or HTAs).ab,ti.	2,541
14	*Technology Assessment, Biomedical/	7,849
15	(8 or 9) and 14	2,278
16	11 or 12 or 13 or 15	11,965
17	5 and 16	4,381
18	(addresses or autobiography or bibliography or biography or case reports or clinical trial or comment or controlled clinical trial or dictionary or directory or editorial or in vitro or interview or letter or multicenter study or note or randomized controlled trial or series or video audio media or webcasts).pt,sh.	6,274,891
19	17 not 18	3,704
20	remove duplicates from 19	2,554
21	limit 20 to yr="2006 -Current"	1,408

Table A2 Search strategy for the Cochrane Library based on preliminary searches

#	Search terms	Hits
1	MeSH descriptor Guidelines as Topic explode all trees	1,566
2	(guideline):pt	29
3	(guid* or (good NEAR/2 practice?) or (best NEAR/2 example?) or methodolog* or recommendation? or tool* or check* or handbook? or standard* or principle?):ti,ab,kw	124,421
4	(#1 or #2 or #3)	124,427
5	MeSH descriptor Biomedical Technology explode all trees	53
6	MeSH descriptor Technology, Medical explode all trees	41
7	MeSH descriptor Economics, Pharmaceutical explode all trees	202
8	(pharmaceutical? or drug? or medicine? or pharmaco-economic? or pharmacoeconomic?):ti,ab,kw	37,142
9	(health or healthcare or (health NEXT care) or medical or single or multiple) NEXT (technolog* or intervention?):ti,ab,kw	1,177
10	(assessment? or appraisal? or evaluation?):ti,ab,kw	15,902
11	(8 or 9) NEXT 10	79
12	(HTA or HTAs):ti,ab,kw	338
13	(#5 or #6 or #7) AND #10	9
14	MeSH descriptor Technology Assessment, Biomedical, this term only	491
15	(#8 or #9) AND #14	139
16	(#11 or #12 or #13 or #15)	519
17	(#4 AND #16)	369

Table A3 Data Extraction Form (in Excel) for HTA and pharmacoeconomic guidelines

Top-line
<i>Guideline type</i>
<i>Guideline title</i>
<i>Publication date</i>
<i>Authors</i>
<i>Authors affiliation</i>
<i>Country</i>
<i>Location</i>
<i>Version number</i>
<i>Availability of previous version(s)</i>
<i>Language</i>
<i>Number of pages</i>
<i>Link to document</i>
<i>Contact website/email</i>
Foreword
<i>Summary of foreword or brief introduction of document</i>
<i>Financial disclosure/Conflicts of interest</i>
Comment by extractor
<i>Personal notes on the documents that were of most interest during data extraction</i>
Table of contents
<i>Structure of the guideline</i>
Background
<i>Stated purpose of document</i>
<i>Was a standard reporting format included in the guideline?</i>
<i>Target audience of funding/ author's interests</i>
<i>Timing of HTA (how long should an assessment take?)</i>
<i>List of any associated/supporting documents referenced in the guideline</i>
Context
<i>Stated indication or criteria for research question</i>
<i>Are subgroup analysis recommended in the guideline?</i>
<i>Choice of comparator (preferred)</i>
<i>Choice of comparator (other)</i>
Evidence identification - clinical trials
<i>Is a pre-analyses required?</i>
<i>Is a systematic search of clinical evidence required?</i>
<i>Search strategy</i>
<i>Databases to be searched</i>
<i>Can unpublished data be considered, if available?</i>
<i>Is the reporting of a search strategy required?</i>
<i>Definition of selection criteria (i.e. PICOS, other)</i>
<i>Minimum number of reviewers recommended</i>
<i>Level of evidence considered (i.e. order of preference for study types included)</i>
<i>Minimum quality score for trials included</i>
<i>How is quality/validity of studies assessed?</i>
<i>Preference for effectiveness over efficacy</i>

Evidence identification - supplementary clinical data
<i>Statement regarding evidence synthesis and supplementary clinical data</i>
<i>Inclusion of meta-analysis in the review of evidence</i>
<i>Inclusion of indirect/mixed treatment comparison in the review of evidence</i>
Data extraction
<i>What are the outcomes of interest for extraction, if any stated?</i>
<i>Are any data adjustments recommended?</i>
Evidence synthesis - clinical efficacy/effectiveness
<i>Preferred statistical software</i>
<i>Are Bayesian methods discussed?</i>
Meta-analysis
<i>Preferred outcome measures</i>
<i>Timepoints</i>
<i>Meta-analysis results</i>
<i>Pictorial representation of studies</i>
<i>Pictorial representation of results</i>
<i>Dichotomous data</i>
<i>Continuous data</i>
<i>Ordinal or categorical data</i>
<i>Time-to-event data</i>
<i>Subset analysis</i>
<i>Cross-over trial analysis</i>
<i>Patient level data analysis</i>
<i>Treatment effect modifiers/meta-regression</i>
<i>Covariate analysis</i>
<i>Statement on alternative methods of combining data</i>
<i>Fixed effects model</i>
<i>Assessment of heterogeneity</i>
<i>Criteria/method to assess heterogeneity</i>
<i>Random effects model</i>
<i>Assessment of goodness of fit</i>
<i>Criteria/method to assess goodness of fit</i>
<i>Sensitivity analysis</i>
<i>Consistency of evidence</i>
Indirect/mixed treatment comparison
<i>Common comparator ("anchor treatment")</i>
<i>Additional study identification</i>
<i>Inclusion criteria for trials in M/ITC</i>
<i>Assumptions required</i>
<i>Preferred outcome measures</i>
<i>Timepoints</i>
<i>I/MTC results</i>
<i>Pictorial representation of studies</i>
<i>Pictorial representation of results</i>
<i>Statement on alternative methods of combining data</i>
<i>Fixed effects model</i>
<i>Assessment of heterogeneity</i>
<i>Criteria/method to assess heterogeneity</i>

<i>Random effects model</i>
<i>Assessment of goodness of fit</i>
<i>Criteria/method to assess goodness of fit</i>
<i>Sensitivity analysis</i>
<i>Consistency of evidence</i>
Evidence synthesis - safety
<i>Statement regarding adverse event data (collection and synthesis)</i>
<i>Should an assessment of comparative harms be performed?</i>
Evidence synthesis - limitations
<i>Limitations with regards to safety data (inclusion, interpretation, etc.)</i>
<i>Limitations with regards to bias in randomised trials</i>
<i>What is recommended in the guideline in the absence of direct randomised comparison or indirect comparison?</i>
Translating the clinical evaluation for the inclusion in the economic evaluation
<i>Is a pre-modelling study required?</i>
<i>What translation issues should be considered, if any?</i>
<i>How should translation issues be addressed?</i>
Economic modelling - clinical effectiveness
<i>Is an economic evaluation required?</i>
<i>What is the recommended time horizon for the economic evaluation?</i>
<i>How is the baseline risk included in the economic model (data source)?</i>
<i>How is the effectiveness/ treatment effect included in the economic model (data source, outcomes, etc.)?</i>
<i>Are and how are adverse events included in the economic model?</i>
<i>What scenario analyses related to clinical effectiveness are recommended in the guideline?</i>
References to 'external' guidelines and key papers
<i>References included within guideline</i>

CD A1 – Database of all licensing and HTA bodies in the world, based on review conducted in section 2.3 in Chapter 2 (October 2011).

Appendix B

Table B1 Adjusted interim life tables for the UK weighted by gender for ACS population

Based on data for the years 2007-2009, available from ONS [2013]

Age	Mortality rate for Males	Mortality rate for Females	Average mortality rate based on %male (overall) in population	Average mortality rate based on %male (intensive) in population	Average mortality rate based on %male (STEMI) in population
0	0.005232	0.004244	0.49%	0.50%	0.49%
1	0.000365	0.000301	0.03%	0.03%	0.03%
2	0.000219	0.000192	0.02%	0.02%	0.02%
3	0.000156	0.000162	0.02%	0.02%	0.02%
4	0.000120	0.000121	0.01%	0.01%	0.01%
5	0.000125	0.000099	0.01%	0.01%	0.01%
6	0.000115	0.000087	0.01%	0.01%	0.01%
7	0.000095	0.000084	0.01%	0.01%	0.01%
8	0.000120	0.000076	0.01%	0.01%	0.01%
9	0.000101	0.000099	0.01%	0.01%	0.01%
10	0.000091	0.000092	0.01%	0.01%	0.01%
11	0.000108	0.000096	0.01%	0.01%	0.01%
12	0.000115	0.000103	0.01%	0.01%	0.01%
13	0.000143	0.000112	0.01%	0.01%	0.01%
14	0.000168	0.000120	0.02%	0.02%	0.02%
15	0.000253	0.000148	0.02%	0.02%	0.02%
16	0.000333	0.000180	0.03%	0.03%	0.03%
17	0.000503	0.000229	0.04%	0.04%	0.04%
18	0.000590	0.000263	0.05%	0.05%	0.05%
19	0.000628	0.000254	0.05%	0.05%	0.05%
20	0.000666	0.000238	0.05%	0.06%	0.05%
21	0.000647	0.000270	0.05%	0.06%	0.05%
22	0.000647	0.000248	0.05%	0.05%	0.05%
23	0.000679	0.000250	0.05%	0.06%	0.06%
24	0.000699	0.000271	0.05%	0.06%	0.06%
25	0.000716	0.000286	0.06%	0.06%	0.06%

Age	Mortality rate for Males	Mortality rate for Females	Average mortality rate based on %male (overall) in population	Average mortality rate based on %male (intensive) in population	Average mortality rate based on %male (STEMI) in population
26	0.000803	0.000319	0.06%	0.07%	0.07%
27	0.000779	0.000314	0.06%	0.07%	0.06%
28	0.000859	0.000344	0.07%	0.07%	0.07%
29	0.000849	0.000398	0.07%	0.07%	0.07%
30	0.000940	0.000417	0.08%	0.08%	0.08%
31	0.000964	0.000430	0.08%	0.08%	0.08%
32	0.001012	0.000507	0.08%	0.09%	0.09%
33	0.001097	0.000519	0.09%	0.10%	0.09%
34	0.001179	0.000606	0.10%	0.10%	0.10%
35	0.001326	0.000619	0.11%	0.11%	0.11%
36	0.001284	0.000622	0.10%	0.11%	0.11%
37	0.001332	0.000716	0.11%	0.12%	0.12%
38	0.001466	0.000790	0.12%	0.13%	0.13%
39	0.001541	0.000881	0.13%	0.14%	0.13%
40	0.001667	0.001006	0.14%	0.15%	0.15%
41	0.001779	0.001062	0.15%	0.16%	0.16%
42	0.001917	0.001123	0.16%	0.17%	0.17%
43	0.002031	0.001243	0.18%	0.18%	0.18%
44	0.002148	0.001346	0.19%	0.19%	0.19%
45	0.002408	0.001528	0.21%	0.22%	0.22%
46	0.002561	0.001631	0.22%	0.23%	0.23%
47	0.002779	0.001769	0.24%	0.25%	0.25%
48	0.002956	0.001959	0.26%	0.27%	0.27%
49	0.003250	0.002100	0.28%	0.30%	0.29%
50	0.003581	0.002490	0.32%	0.33%	0.33%
51	0.003960	0.002546	0.35%	0.36%	0.35%
52	0.004295	0.002779	0.38%	0.39%	0.39%
53	0.004813	0.003126	0.42%	0.44%	0.43%
54	0.005173	0.003549	0.46%	0.48%	0.47%
55	0.005888	0.003720	0.51%	0.53%	0.53%
56	0.006317	0.004043	0.55%	0.57%	0.57%
57	0.006748	0.004344	0.59%	0.61%	0.60%
58	0.007478	0.004692	0.65%	0.68%	0.67%
59	0.008071	0.005239	0.71%	0.74%	0.72%
60	0.008680	0.005650	0.76%	0.79%	0.78%

Age	Mortality rate for Males	Mortality rate for Females	Average mortality rate based on %male (overall) in population	Average mortality rate based on %male (intensive) in population	Average mortality rate based on %male (STEMI) in population
61	0.009606	0.006293	0.84%	0.88%	0.86%
62	0.010559	0.006637	0.92%	0.96%	0.94%
63	0.011930	0.007634	1.04%	1.08%	1.07%
64	0.013085	0.008254	1.14%	1.19%	1.17%
65	0.014319	0.008966	1.24%	1.30%	1.28%
66	0.015791	0.009858	1.37%	1.43%	1.41%
67	0.017132	0.010837	1.49%	1.55%	1.53%
68	0.019357	0.011919	1.67%	1.75%	1.72%
69	0.021078	0.013256	1.83%	1.91%	1.88%
70	0.022566	0.014876	1.98%	2.06%	2.03%
71	0.025002	0.016044	2.18%	2.27%	2.24%
72	0.027841	0.017827	2.43%	2.53%	2.49%
73	0.030718	0.020214	2.70%	2.81%	2.77%
74	0.033759	0.022527	2.98%	3.09%	3.05%
75	0.037991	0.024893	3.34%	3.47%	3.42%
76	0.042295	0.028209	3.73%	3.87%	3.82%
77	0.046721	0.031449	4.13%	4.29%	4.23%
78	0.051565	0.035339	4.58%	4.75%	4.68%
79	0.058170	0.040520	5.19%	5.37%	5.30%
80	0.064932	0.045924	5.82%	6.01%	5.94%
81	0.072472	0.051195	6.49%	6.71%	6.63%
82	0.080207	0.056994	7.20%	7.44%	7.34%
83	0.088476	0.064608	8.00%	8.25%	8.15%
84	0.098993	0.072630	8.97%	9.23%	9.13%
85	0.109296	0.081427	9.94%	10.23%	10.12%
86	0.120497	0.090915	11.00%	11.30%	11.19%
87	0.131106	0.101717	12.07%	12.37%	12.25%
88	0.137116	0.111061	12.79%	13.06%	12.95%
89	0.148716	0.123014	13.96%	14.22%	14.12%
90	0.159607	0.134853	15.08%	15.34%	15.24%
91	0.179491	0.155472	17.10%	17.34%	17.25%
92	0.200188	0.177071	19.20%	19.44%	19.34%
93	0.216959	0.193418	20.86%	21.10%	21.01%
94	0.231609	0.212269	22.48%	22.67%	22.60%
95	0.261066	0.232289	25.09%	25.38%	25.27%

Age	Mortality rate for Males	Mortality rate for Females	Average mortality rate based on %male (overall) in population	Average mortality rate based on %male (intensive) in population	Average mortality rate based on %male (STEMI) in population
96	0.275355	0.251371	26.69%	26.93%	26.84%
97	0.299192	0.265039	28.71%	29.06%	28.92%
98	0.311452	0.294414	30.54%	30.72%	30.65%
99	0.327639	0.311356	32.19%	32.35%	32.29%
100	0.346497	0.335427	34.26%	34.37%	34.33%

Table B2 Markov trace for clopidogrel in overall ACS population (PLATO, base case)

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
start (0)	847			63	11	79	£9,748,059	£9,748,059	921	921	789	789
1	771	27	9	60	11	123	£453,817	£453,817	847	847	708	708
2	697	24	8	79	17	174	£430,273	£442,045	771	809	643	675
3	627	22	7	95	23	227	£404,037	£417,155	698	734	581	612
4	559	20	6	107	27	281	£375,763	£389,900	627	662	521	551
5	495	18	6	115	30	337	£345,716	£360,740	558	592	463	492
6	433	16	5	119	31	395	£314,289	£330,003	492	525	407	435
7	375	14	4	120	32	454	£282,264	£298,276	429	461	355	381
8	322	12	4	118	32	512	£250,231	£266,247	370	400	306	330
9	271	10	3	113	31	571	£218,221	£234,226	314	342	259	282
10	225	9	3	105	29	629	£186,910	£202,566	263	289	216	238
11	183	7	2	95	27	685	£157,064	£171,987	215	239	177	197
12	147	6	2	84	24	738	£129,348	£143,206	173	194	142	160
13	114	5	1	72	21	787	£104,129	£116,738	136	155	112	127
14	87	4	1	60	17	831	£81,565	£92,847	105	121	85	99
15	64	3	1	48	14	870	£62,045	£71,805	78	91	64	75
16	46	2	1	38	11	903	£45,709	£53,877	56	67	46	55
17	32	1	0	28	8	930	£32,560	£39,135	39	48	32	39
18	21	1	0	21	6	951	£22,612	£27,586	27	33	22	27
19	14	1	0	14	4	966	£15,155	£18,884	17	22	14	18

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
20	9	0	0	10	3	978	£9,772	£12,463	11	14	9	12
21	5	0	0	6	2	987	£5,895	£7,834	6	9	5	7
22	3	0	0	3	1	993	£3,278	£4,587	3	5	3	4
23	1	0	0	2	1	996	£1,693	£2,486	2	3	1	2
24	1	0	0	1	0	998	£812	£1,252	1	1	1	1
25	0	0	0	0	0	999	£345	£579	0	1	0	0
26	0	0	0	0	0	1,000	£132	£239	0	0	0	0
27	0	0	0	0	0	1,000	£45	£88	0	0	0	0
28	0	0	0	0	0	1,000	£13	£29	0	0	0	0
29	0	0	0	0	0	1,000	£3	£8	0	0	0	0
30	0	0	0	0	0	1,000	£1	£2	0	0	0	0
31	0	0	0	0	0	1,000	£0	£0	0	0	0	0
32	0	0	0	0	0	1,000	£0	£0	0	0	0	0
33	0	0	0	0	0	1,000	£0	£0	0	0	0	0
34	0	0	0	0	0	1,000	£0	£0	0	0	0	0
35	0	0	0	0	0	1,000	£0	£0	0	0	0	0
36	0	0	0	0	0	1,000	£0	£0	0	0	0	0
37	0	0	0	0	0	1,000	£0	£0	0	0	0	0
38	0	0	0	0	0	1,000	£0	£0	0	0	0	0
39	0	0	0	0	0	1,000	£0	£0	0	0	0	0
40	0	0	0	0	0	1,000	£0	£0	0	0	0	0

Table B3 Markov trace for ticagrelor in overall ACS population (PLATO, base case)

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
start (0)	872			54	12	62	£10,009,068	£10,009,068	938	938	797	797
1	793	27	9	51	12	107	£466,573	£466,573	863	863	721	721
2	718	25	8	72	18	159	£442,393	£454,483	785	824	655	688
3	645	23	7	89	24	212	£415,440	£428,917	710	748	591	623
4	576	20	7	102	28	268	£386,388	£400,914	638	674	530	561
5	509	18	6	111	31	325	£355,509	£370,948	568	603	471	501
6	446	16	5	116	33	384	£323,206	£339,357	501	534	415	443
7	386	14	5	117	34	444	£290,286	£306,746	437	469	361	388
8	331	12	4	116	33	504	£257,355	£273,820	377	407	311	336
9	279	10	3	111	32	564	£224,444	£240,900	320	349	264	287
10	232	9	3	103	30	623	£192,251	£208,347	267	294	220	242
11	189	7	2	94	28	680	£161,560	£176,905	219	243	180	200
12	151	6	2	83	25	733	£133,058	£147,309	176	198	145	162
13	118	5	2	72	21	783	£107,122	£120,090	139	158	114	129
14	90	4	1	60	18	828	£83,915	£95,519	106	123	87	100
15	66	3	1	48	15	867	£63,838	£73,877	79	93	65	76
16	47	2	1	37	11	901	£47,034	£55,436	57	68	47	56
17	33	1	0	28	9	929	£33,507	£40,271	40	48	32	39
18	22	1	0	21	6	950	£23,273	£28,390	27	33	22	27
19	14	1	0	14	5	966	£15,599	£19,436	18	22	14	18

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
20	9	0	0	10	3	978	£10,060	£12,830	11	14	9	12
21	5	0	0	6	2	986	£6,070	£8,065	7	9	5	7
22	3	0	0	3	1	992	£3,376	£4,723	4	5	3	4
23	1	0	0	2	1	996	£1,744	£2,560	2	3	1	2
24	1	0	0	1	0	998	£837	£1,290	1	1	1	1
25	0	0	0	0	0	999	£356	£596	0	1	0	0
26	0	0	0	0	0	1000	£136	£246	0	0	0	0
27	0	0	0	0	0	1000	£46	£91	0	0	0	0
28	0	0	0	0	0	1000	£14	£30	0	0	0	0
29	0	0	0	0	0	1000	£3	£8	0	0	0	0
30	0	0	0	0	0	1000	£1	£2	0	0	0	0
31	0	0	0	0	0	1000	£0	£0	0	0	0	0
32	0	0	0	0	0	1000	£0	£0	0	0	0	0
33	0	0	0	0	0	1000	£0	£0	0	0	0	0
34	0	0	0	0	0	1000	£0	£0	0	0	0	0
35	0	0	0	0	0	1000	£0	£0	0	0	0	0
36	0	0	0	0	0	1000	£0	£0	0	0	0	0
37	0	0	0	0	0	1000	£0	£0	0	0	0	0
38	0	0	0	0	0	1000	£0	£0	0	0	0	0
39	0	0	0	0	0	1000	£0	£0	0	0	0	0
40	0	0	0	0	0	1000	£0	£0	0	0	0	0

Table B4 Markov trace for prasugrel in invasively managed ACS patients (PLATO-INVASIVE, subgroup analysis 1)

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
start (0)	895			48	10	47	£10,009,417	£10,009,417	953	953	790	790
1	839	28	9	47	10	67	£474,865	£474,865	901	901	754	754
2	784	26	9	72	18	91	£466,793	£470,829	848	875	708	731
3	730	25	8	95	25	117	£456,061	£461,427	797	823	663	686
4	679	23	7	115	32	144	£443,106	£449,583	746	771	620	642
5	629	21	7	132	38	172	£428,089	£435,598	697	722	578	599
6	581	20	6	147	43	203	£411,349	£419,719	649	673	537	558
7	535	18	6	159	47	235	£392,774	£402,061	601	625	497	517
8	490	17	5	169	50	269	£372,781	£382,777	555	578	458	477
9	447	15	5	176	52	304	£351,898	£362,339	510	533	421	439
10	406	14	5	181	54	341	£330,105	£341,001	467	489	384	402
11	366	13	4	182	55	380	£307,383	£318,744	425	446	349	367
12	328	12	4	182	55	420	£283,981	£295,682	384	404	315	332
13	292	10	3	179	54	461	£260,254	£272,117	345	364	282	299
14	257	9	3	173	53	504	£236,175	£248,214	306	325	251	266
15	225	8	3	166	51	548	£212,013	£224,094	270	288	220	235
16	194	7	2	156	48	593	£188,225	£200,119	235	252	192	206
17	166	6	2	145	45	637	£165,097	£176,661	202	219	165	178
18	139	5	2	132	41	681	£142,533	£153,815	171	187	140	152
19	115	4	1	118	37	725	£120,969	£131,751	143	157	116	128

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
20	93	4	1	103	32	767	£100,784	£110,877	117	130	95	106
21	74	3	1	88	28	807	£82,339	£91,561	94	105	76	86
22	57	2	1	73	24	843	£65,808	£74,073	74	84	60	68
23	43	2	1	59	19	876	£51,196	£58,502	56	65	46	53
24	32	1	0	47	15	905	£38,713	£44,955	42	49	34	40
25	22	1	0	36	12	929	£28,373	£33,543	30	36	24	29
26	15	1	0	26	9	949	£20,137	£24,255	21	25	17	21
27	10	0	0	19	6	964	£13,975	£17,056	14	18	11	14
28	7	0	0	13	5	976	£9,367	£11,671	9	12	8	9
29	4	0	0	9	3	984	£6,054	£7,710	6	8	5	6
30	2	0	0	5	2	990	£3,665	£4,859	3	5	3	4
31	1	0	0	3	1	995	£2,055	£2,860	2	3	2	2
32	1	0	0	2	1	997	£1,077	£1,566	1	1	1	1
33	0	0	0	1	0	999	£528	£803	0	1	0	1
34	0	0	0	0	0	999	£230	£379	0	0	0	0
35	0	0	0	0	0	1000	£92	£161	0	0	0	0
36	0	0	0	0	0	1000	£33	£62	0	0	0	0
37	0	0	0	0	0	1000	£11	£22	0	0	0	0
38	0	0	0	0	0	1000	£3	£7	0	0	0	0
39	0	0	0	0	0	1000	£1	£2	0	0	0	0
40	0	0	0	0	0	1000	£0	£1	0	0	0	0

Table B5 Markov trace for ticagrelor in invasively managed ACS patients (PLATO-INVASIVE, subgroup analysis 1)

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
start (0)	896			53	12	39	£10,101,919	£10,101,919	961	961	799	799
1	840	28	9	52	11	59	£482,618	£482,618	909	909	760	760
2	785	26	9	77	20	84	£474,134	£478,376	856	882	714	737
3	731	25	8	100	27	109	£462,995	£468,565	803	829	669	691
4	680	23	7	120	33	136	£449,640	£456,318	753	778	625	647
5	630	21	7	137	39	165	£434,227	£441,934	703	728	583	604
6	582	20	6	152	44	196	£417,098	£425,663	654	678	542	562
7	535	18	6	164	48	229	£398,136	£407,617	606	630	501	521
8	490	17	5	173	51	263	£377,764	£387,950	560	583	462	481
9	447	15	5	180	54	298	£356,512	£367,138	515	537	424	443
10	406	14	5	184	55	335	£334,357	£345,435	471	493	387	406
11	367	13	4	186	56	374	£311,280	£322,819	429	450	352	370
12	329	12	4	185	56	415	£287,529	£299,405	387	408	318	335
13	292	10	3	182	55	456	£263,464	£275,496	348	367	285	301
14	258	9	3	176	54	500	£239,054	£251,259	309	328	253	269
15	225	8	3	168	52	544	£214,571	£226,813	272	290	222	237
16	194	7	2	158	49	589	£190,476	£202,524	237	254	193	208
17	166	6	2	147	46	634	£167,057	£178,767	204	221	166	180
18	139	5	2	134	42	679	£144,215	£155,636	173	189	141	153
19	115	4	1	119	37	723	£122,390	£133,302	144	159	117	129

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
20	93	4	1	104	33	765	£101,964	£112,177	118	131	96	107
21	74	3	1	89	28	805	£83,301	£92,632	95	106	77	86
22	57	2	1	74	24	841	£66,577	£74,939	74	85	60	69
23	43	2	1	60	20	875	£51,795	£59,186	57	66	46	53
24	32	1	0	47	16	904	£39,168	£45,481	42	50	34	40
25	22	1	0	36	12	928	£28,708	£33,938	30	36	24	29
26	15	1	0	26	9	948	£20,377	£24,543	21	26	17	21
27	10	0	0	19	7	964	£14,143	£17,260	14	18	12	14
28	7	0	0	13	5	975	£9,481	£11,812	9	12	8	10
29	4	0	0	9	3	984	£6,128	£7,804	6	8	5	6
30	2	0	0	5	2	990	£3,711	£4,920	4	5	3	4
31	1	0	0	3	1	994	£2,081	£2,896	2	3	2	2
32	1	0	0	2	1	997	£1,091	£1,586	1	1	1	1
33	0	0	0	1	0	999	£535	£813	0	1	0	1
34	0	0	0	0	0	999	£233	£384	0	0	0	0
35	0	0	0	0	0	1000	£93	£163	0	0	0	0
36	0	0	0	0	0	1000	£33	£63	0	0	0	0
37	0	0	0	0	0	1000	£11	£22	0	0	0	0
38	0	0	0	0	0	1000	£3	£7	0	0	0	0
39	0	0	0	0	0	1000	£1	£2	0	0	0	0
40	0	0	0	0	0	1000	£0	£0	0	0	0	0

Table B6 Markov trace for prasugrel in STEMI patients undergoing PCI (PLATO-STEMI, subgroup analysis 2)

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
start (0)	899			35	8	57	£9,924,791	£9,924,791	943	943	785	785
1	827	28	9	34	8	93	£464,053	£464,053	876	876	733	733
2	758	26	8	58	16	134	£446,943	£455,498	809	843	675	704
3	693	24	8	78	22	175	£427,424	£437,184	744	776	620	647
4	629	22	7	95	28	219	£405,705	£416,565	681	712	566	593
5	569	20	6	109	32	265	£381,831	£393,768	619	650	514	540
6	510	18	6	119	35	312	£356,149	£368,990	559	589	463	489
7	454	16	5	125	37	362	£329,202	£342,675	502	531	415	439
8	401	14	5	129	39	413	£301,112	£315,157	446	474	368	392
9	350	13	4	129	39	465	£272,236	£286,674	393	419	324	346
10	303	11	4	126	38	518	£243,233	£257,735	342	367	281	303
11	259	10	3	121	37	570	£214,571	£228,902	295	318	242	262
12	218	8	3	114	35	623	£186,225	£200,398	250	272	205	223
13	180	7	2	104	32	675	£158,775	£172,500	208	229	170	188
14	146	6	2	93	29	724	£132,817	£145,796	170	189	139	155
15	116	5	1	81	25	771	£108,887	£120,852	137	153	112	125
16	91	4	1	69	22	814	£87,278	£98,083	107	122	87	100
17	68	3	1	57	18	853	£68,065	£77,672	82	94	67	77
18	50	2	1	45	14	887	£51,556	£59,810	61	71	49	58
19	36	2	1	35	11	916	£37,822	£44,689	44	52	35	42

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
20	25	1	0	26	8	940	£26,841	£32,332	30	37	25	30
21	17	1	0	19	6	958	£18,594	£22,717	21	25	17	21
22	11	1	0	13	4	971	£12,434	£15,514	13	17	11	14
23	7	0	0	9	3	981	£8,006	£10,220	8	11	7	9
24	4	0	0	5	2	989	£4,826	£6,416	5	7	4	5
25	2	0	0	3	1	994	£2,686	£3,756	3	4	2	3
26	1	0	0	2	1	997	£1,391	£2,038	1	2	1	2
27	0	0	0	1	0	998	£671	£1,031	1	1	0	1
28	0	0	0	0	0	999	£287	£479	0	0	0	0
29	0	0	0	0	0	1000	£111	£199	0	0	0	0
30	0	0	0	0	0	1000	£38	£75	0	0	0	0
31	0	0	0	0	0	1000	£12	£25	0	0	0	0
32	0	0	0	0	0	1000	£3	£7	0	0	0	0
33	0	0	0	0	0	1000	£1	£2	0	0	0	0
34	0	0	0	0	0	1000	£0	£0	0	0	0	0
35	0	0	0	0	0	1000	£0	£0	0	0	0	0
36	0	0	0	0	0	1000	£0	£0	0	0	0	0
37	0	0	0	0	0	1000	£0	£0	0	0	0	0
38	0	0	0	0	0	1000	£0	£0	0	0	0	0
39	0	0	0	0	0	1000	£0	£0	0	0	0	0
40	0	0	0	0	0	1000	£0	£0	0	0	0	0

Table B7 Markov trace for ticagrelor in STEMI patients undergoing PCI (PLATO-STEMI, subgroup analysis 2)

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
start (0)	896			34	12	57	£10,029,709	£10,029,709	943	943	787	787
1	825	28	9	33	12	93	£474,975	£474,975	876	876	733	733
2	756	26	8	57	19	134	£457,037	£466,006	809	843	675	704
3	691	24	8	77	25	175	£436,726	£446,882	744	776	619	647
4	628	22	7	94	31	219	£414,241	£425,484	681	712	566	593
5	567	20	6	108	35	265	£389,623	£401,932	619	650	514	540
6	509	18	6	118	38	312	£363,223	£376,423	559	589	463	488
7	453	16	5	124	40	361	£335,585	£349,404	502	531	415	439
8	400	14	5	128	41	412	£306,827	£321,206	446	474	368	391
9	349	13	4	128	41	465	£277,309	£292,068	393	419	323	346
10	302	11	4	126	40	517	£247,696	£262,503	342	367	281	302
11	258	10	3	121	39	570	£218,459	£233,077	295	318	242	262
12	217	8	3	113	36	623	£189,564	£204,011	250	272	205	223
13	180	7	2	103	33	675	£161,602	£175,583	208	229	170	188
14	146	6	2	92	30	724	£135,173	£148,388	170	189	139	155
15	116	5	1	81	26	771	£110,818	£122,995	137	153	112	125
16	90	4	1	69	22	814	£88,829	£99,823	107	122	87	100
17	68	3	1	56	18	853	£69,282	£79,055	82	95	67	77
18	50	2	1	45	15	887	£52,486	£60,884	61	71	49	58
19	36	2	1	35	12	916	£38,515	£45,500	44	52	35	42

Cycle	No event	Non-fatal MI	Non-fatal stroke	Post-MI	Post-stroke	Death	Cost	Cost (1/2-cycle correction)	Life-years	Life-years (1/2-cycle correction)	QALYs	QALYs (1/2-cycle correction)
20	25	1	0	26	9	940	£27,342	£32,928	30	37	25	30
21	16	1	0	19	6	958	£18,949	£23,145	21	25	17	21
22	11	1	0	13	4	971	£12,678	£15,813	13	17	11	14
23	7	0	0	9	3	981	£8,168	£10,423	8	11	7	9
24	4	0	0	5	2	989	£4,927	£6,548	5	7	4	5
25	2	0	0	3	1	994	£2,744	£3,836	3	4	2	3
26	1	0	0	2	1	997	£1,423	£2,084	1	2	1	2
27	0	0	0	1	0	998	£688	£1,056	1	1	1	1
28	0	0	0	0	0	999	£295	£492	0	0	0	0
29	0	0	0	0	0	1000	£115	£205	0	0	0	0
30	0	0	0	0	0	1000	£40	£77	0	0	0	0
31	0	0	0	0	0	1000	£12	£26	0	0	0	0
32	0	0	0	0	0	1000	£3	£8	0	0	0	0
33	0	0	0	0	0	1000	£1	£2	0	0	0	0
34	0	0	0	0	0	1000	£0	£0	0	0	0	0
35	0	0	0	0	0	1000	£0	£0	0	0	0	0
36	0	0	0	0	0	1000	£0	£0	0	0	0	0
37	0	0	0	0	0	1000	£0	£0	0	0	0	0
38	0	0	0	0	0	1000	£0	£0	0	0	0	0
39	0	0	0	0	0	1000	£0	£0	0	0	0	0
40	0	0	0	0	0	1000	£0	£0	0	0	0	0

Appendix C

Table C1 Second order search strategy without placebo (cf. Table 20 search 2i)

#	Search terms	Hits
99	(acenocoumarol).mp.	5,653
100	(ancrod or viprinex).mp.	1,684
101	(aspirin or (ASA) or (acetylsalicylic ADJ acid)).mp.	239,630
102	(ave5026).mp.	11
103	(hirudin or cgp39393 or desirudin or revasc or iprivask).mp.	9,042
104	(betrixaban or prt054).mp	209
105	(dextran).mp.	74,581
106	(edoxaban or du176b).mp.	410
107	(dihydroergotamine or dhe or migranal).mp.	8,924
108	(rosuvastatin or crestor).mp.	8,576
109	(indomethacin or indomethacin).mp.	76,805
110	(LY517717 or TAK442 or TB402).mp.	33
111	(melagatran or ximelagatran or exanta or exarta).mp.	3,004
112	(lomoparan or org10172).mp.	85
113	(warfarin or coumadin or jantoven or marevan or lawarin or waran or warfant).mp.	76,514
114	((intermittent or pneumatic) and (compression\$1)).mp	5,647
115	((synchroni* adj flow adj technolog\$3) or (SFT or SCD)).mp	14,046
116	(foot adj pump\$1).mp	214
117	((continuous adj enhanced adj circulation adj therap\$3) or CECT).mp.	1,961
118	((garment\$1) or (graduated adj compression adj stocking\$1)).mp.	5,670
	placebo .mp.	
119	or/99-118	490,606
120	8 and 119 and 84 and 88	1,461
121	limit 120 to yr="2012 -Current"	77
122	120 not 121	1,384
123	remove duplicates from 122	1,049
124	123 not 92	237

Table C2 Second order search strategy placebo only (cf Table 20 search 2ii)

#	Search terms	Hits
125	placebo.mp.	444,434
126	8 and 125 and 84 and 88	684
127	limit 126 to yr="2012 -Current"	45
128	126 not 127	639
129	remove duplicates from 128	508
130	128 not 92	176
131	130 not 123	122

Table C3 Third order search strategy without steroid (cf Table 20 search 3i)

#	Search	Hits
132	(danaparoid or organan).mp	2,545
133	triflusal.mp	662
134	(methylprednisolone or medrol).mp	88,869
135	(tocainide or tonocard or lidocaine).mp	87,763
136	(dipyridamole or persantin\$1 or antistenocardin\$1).mp	32,607
137	((impulse) and (foot or system or AV)).mp	10,357
138	((inflation or sequential or pneumatic or plantar or intermittent) and (compression\$1)).mp	8,264
	(steroid*).mp	
139	or/132-138	227,692
140	8 and 139 and 84 and 88	334
141	limit 140 to yr="2012 -Current"	8
142	140 not 141	326
143	remove duplicates from 142	248
144	143 not 92	78

Table C4 Third order search strategy steroid only (cf. Table 20 search 3ii)

#	Search	Hits
145	(steroid*).mp	550,465
146	8 and 145 and 84 and 88	48
147	limit 146 to yr="2012 -Current"	2
148	146 not 147	46
149	remove duplicates from 148	43
150	148 not 92	33
151	150 not 144	30

Table C5 Study selection criteria for abstracts and full-text papers

	Inclusion criteria	Exclusion criteria
Population	Adult patients (≥ 18 years) undergoing elective knee or hip replacement surgery	Patients: <ul style="list-style-type: none"> • undergoing emergency hip or knee surgery • undergoing surgery for hip fracture repair • undergoing other types of surgery • treated under non-surgical indications; e.g. to prevent VTE in acute medical illness • treated only once a VTE event has occurred (i.e. active treatment of VTE event)
Outcomes	<ul style="list-style-type: none"> • Mortality (VTE-related, all cause) • Incidence of VTE • Post DVT complications including post thrombotic syndrome (PTS) • Length of hospital stay • Joint outcomes, including joint infection • Adverse events including bleeding events (intracranial bleeding, major bleeding, clinically relevant non-major bleeding) • Health-related quality of life 	
Study design	Prospective, randomised controlled trials, phase II-IV	Non-RCT studies
Language restrictions	Only abstracts in English were included	

Table C6 List of studies included in the NMAs by order

Study #	Author, Publication Date	Treatments	All VTE /death	All DVT	All bleeds
Base case (ITC)					
1	Eriksson 2007 (RE-MODEL)	enoxaparin 40mg.qd.sc	✓	✓	✓
		dabigatran etexilate 150mg.qd.o			
		dabigatran etexilate 220mg.qd.o			
2	Lassen 2010 (ADVANCE 2)	enoxaparin 40mg.qd.sc	✓	✓	✓
		apixaban 2.5mg.bid.o			
3	Lassen 2008 (RECORD 3)	enoxaparin 40mg.qd.sc	✓	✓	✓
		rivaroxaban 10mg.qd.o			
First order NMA					
4	Bauer 2001 (Pentamks)	enoxaparin 30mg.bid.sc	nr	✓	nr
		fondaparinux 2.5mg.qd.sc			
5	Blanchard 1999	fraxiparine (nadroparin calcium) 0.2-0.4mL.qd.sc	✓	✓	nr
		continuous intermittent pneumatic compression			
6	Chin 2009	placebo/control	✓	✓	nr
		enoxaparin 40mg.qd.sc			
		graduated compression stockings			
		intermittent pneumatic compression with each inflation/ deflation cycle lasting 1min and pressures up to 45 to 52 mmHg			
7	Colwell 1995	enoxaparin 30mg.bid.sc	nr	✓	nr
		unfractionated heparin 5000U.q8h.sc			
8	Fauno 1994	enoxaparin 40mg.qc.sc	nr	✓	nr
		unfractionated heparin 5000U.q8h.sc			
9	Fitzgerald 2001	enoxaparin 30mg.bid.sc	✓	✓	nr
		warfarin adj. INR2.0-3.0			
10	Fuji 2010	placebo/control	✓	✓	✓
		dabigatran etexilate 110mg.qd.o			
		dabigatran etexilate 150mg.qd.o			
		dabigatran etexilate			

Study #	Author, Publication Date	Treatments	All VTE /death	All DVT	All bleeds
		220mg.qd.o			
11	Fuji 2008a	placebo/control	✓	nr	✓
		fondaparinux 0.75mg.qd.sc			
		fondaparinux 1.5mg.qd.sc			
		fondaparinux 2.5mg.qd.sc			
		fondaparinux 3.0mg.qd.sc			
12	Fuji 2008b	placebo/control	nr	✓	✓
		enoxaparin 20mg.qc.sc			
		enoxaparin 40mg.qd.sc			
		enoxaparin 20mg.bid.sc			
13	Lassen 2009 (ADVANCE 1)	enoxaparin 30mg.bid.sc	✓	✓	✓
		apixaban 2.5mg.bid.o			
14	Lassen 2007 (APROPOS)	enoxaparin 30mg.bid.sc	✓	nr	✓
		warfarin titrated INR 1.8-3.0.o.			
		apixaban 2.5mg.bid.o			
		apixaban 5mg.qd.o			
		apixaban 5mg.bid.o			
		apixaban 10mg.qd.o			
		apixaban 10mg.bid.o			
15	REMOBILIZE 2009	apixaban 20mg.qd.o	✓	✓	nr
		enoxaparin 30mg.bid.sc			
		dabigatran etexilate 150mg.qd.o			
		dabigatran etexilate 220mg.qd.o			
16	Turpie 2009 (RECORD 4)	enoxaparin 30mg.bid.sc	✓	✓	✓
		rivaroxaban 10mg.qd.o			
17	Turpie 2005	enoxaparin 30mg.big.sc	✓	✓	✓
		rivaroxaban 2.5mg.qd.o			
		rivaroxaban 5mg.qd.o			
		rivaroxaban 10mg.qd.o			
		rivaroxaban 20mg.qd.o			
		rivaroxaban 30mg.qd.o			
18	Wang 2004	placebo/control	nr	✓	nr
		fraxiparine (nadroparin calcium) 0.2-0.4mL.qd.sc			
		indomethacin 25mg.bid.o			
Second order NMA					
19	Colwell 2005	warfarin adj. INR 2.5 (1.8-3.0)	nr	✓	nr
		ximelagatran 36mg.bid.o			
20	Francis 2003	warfarin adj. INR 2.5 (1.8-3.0)	✓	✓	✓

Study #	Author, Publication Date	Treatments	All VTE /death	All DVT	All bleeds
		ximelagatran 24mg.bid.o			
		ximelagatran 36mg.bid.o			
21	Francis 2002	warfarin adj. INR 2.5 (1.8-3.0)	nr	nr	✓
		ximelagatran 24mg.bid.o			
22	Heit 2001	enoxaparin 30mg.bid.sc	nr	✓	✓
		ximelagatran 8mg.bid.o			
		ximelagatran 12mg.bid.o			
		ximelagatran 18mg.bid.o			
		ximelagatran 24mg.bid.o			
23	Kaempffe 1991	warfarin, prothrombin time to 1.5 to 2.0 times normal	nr	✓	nr
		intermittent pneumatic compression with each inflation/ deflation cycle lasting 1min and pressures up to 35 to 55 mmHg			
24	Warwick 2002	enoxaparin 40mg.qc.sc	nr	✓	nr
		AV-impulse foot pump, activated every 20sec/30mmHG pressure			
25	Wilson 1992	placebo/control	nr	nr	nr
		AV-Impulse foot pump, activated every 20sec/30mmHG pressure			
Third order NMA					
26	McKenna 1980	placebo/control	nr	✓	nr
		aspirin 325mg.tid.o			
		aspirin 1300mg.tid.o			
		intermittend pneumatic compression device, max. pressure of 30 mm Hg in 5 sec			

Figure C1 First, second and third network orders for all identified studies (* indicates multiple dosages)

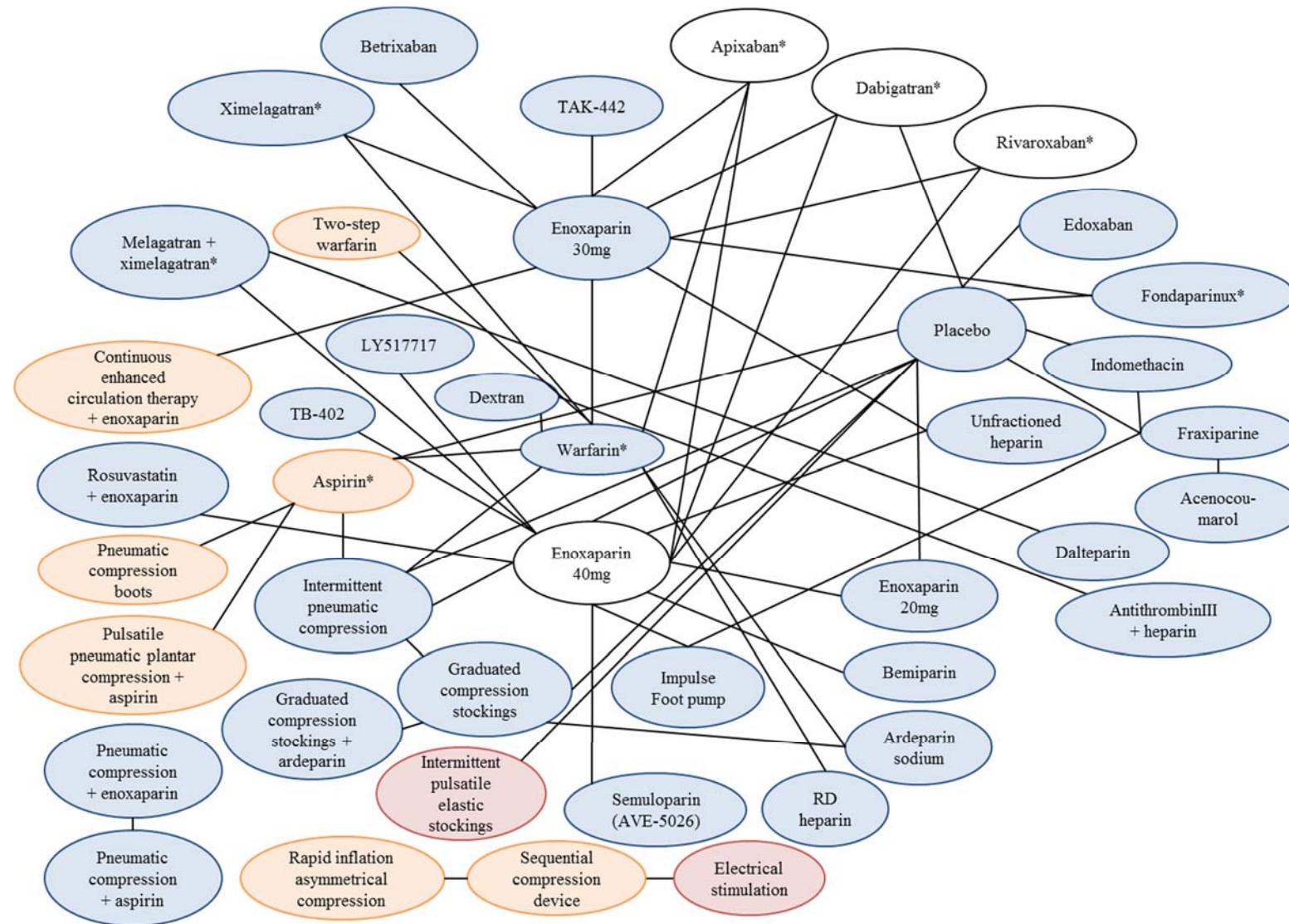


Figure C2 Random effects NMA model for 3rd network order (i.e. with the most complete dataset for total VTE/all-cause death, all DVT, and any bleeds)

```

# Random effects model for multi-arm trials (any number of arms)
# Taken from Ades et al. 2007 (http://www.bris.ac.uk/social-community-medicine/media/mpes/intro-to-mtc.pdf)
# NT=no. treatments, NS=no. studies

model{

d[1]<-0

for(i in 1:NS) {
w[i,1] <-0
delta[i,t[i,1]]<-0
delta.new[i,t[i,1]]<-0

for (k in 1:na[i]) {

r[i,k] ~ dbin(p[i,t[i,k]],n[i,k])          # binomial likelihood
logit(p[i,t[i,k]]) <- mu[i] + delta[i,t[i,k]] # model
rhat[i,k] <- p[i,t[i,k]] * n[i,k]          # predicted r values
dev[i,k] <- 2 *(r[i,k] *(log(r[i,k])-log(rhat[i,k]))+(n[i,k]-r[i,k])*(log(n[i,k]-r[i,k])-log(n[i,k]-rhat[i,k])))) }

# deviance contribution for study i (should be the same for each treatment arm)
resdev[i] <- sum(dev[i,1:na[i]])

for (k in 2:na[i]) {
delta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],taud[i,t[i,k]]) # trial-specific LOR distributions
md[i,t[i,k]] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]      # mean of LOR distributions
taud[i,t[i,k]] <- tau *2*(k-1)/k                    # precision of LOR distributions
w[i,k] <- (delta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]]) #adjustment, multi-arm RCTs
sw[i,k] <-sum(w[i,1:k-1])/(k-1)                    #cumulative adjustment for multi-arm trials

# Generate new set of contrasts for studies
delta.new[i,t[i,k]] ~ dnorm(md[i,t[i,k]],taud[i,t[i,k]])
} }

totresdev <- sum(resdev[]) # total residual deviance

# Prior distributions
for(i in 1:NS){ mu[i] ~ dnorm(0,0.01) } # vague priors for 10 trial baselines
for (k in 2:NT) { d[k] ~ dnorm(0,0.01) } # vague priors for basic parameters
sd~dunif(0,2) # vague prior for random effects standard deviation
tau<-1/pow(sd,2)

```

```

# Generate replicate observations & calculate mixed predictive p-values
for(i in 1:NS){
  for (k in 1:na[i]) {
    # generate new probability estimate
    logit(p.new[i,k]) <- mu[i] + delta.new[i,t[i,k]]
    r.mxd[i,k] ~ dbin(p.new[i,k],n[i,k])

                                                                    # generate
    predicted r
    p.mxd[i,k] <- step(r.mxd[i,k] - r[i,k]) - 0.5*equals(r.mxd[i,k],r[i,k]) # calculate p-
    value
  } }

# Using Treatment 1 as baseline
for (i in 1:NS) { mu2[i] <- mu[i] * equals(t[i,1],1)
  treat1.stud[i] <- equals(t[i,1],1) }

for (k in 1:NT) { logit(T[k])<- sum(mu2[])/sum(treat1.stud[]) +d[k] }

# Ranking and probability treatment k is best
for (k in 1:NT) {
  rk[k] <- rank(T[],k)
  best[k] <- equals(rk[k],1) }

# Pairwise Odd Ratios
for (c in 1:(NT-1)) {
  for (k in (c+1):NT) {
    lor[c,k] <- d[k] - d[c]
    log(or[c,k]) <- lor[c,k] } }

} # END

```

Data

#Total VTE/all-cause death

list(NT=20,NS=12)

r[,1]	n[,1] r[,7] t[,8]	r[,2] n[,7] na[]	n[,2] r[,8]	r[,3] n[,8]	n[,3] t[,1]	r[,4] t[,2]	n[,4] t[,3]	r[,5] t[,4]	n[,5] t[,5]	r[,6] t[,6]	n[,6] t[,7]
243	997 NA NA	147 NA 2	976 NA NA	NA NA NA	NA 1 NA	NA 4 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
57	101 NA NA	42 NA 4	106 NA NA	34 NA NA	104 2 NA	23 5 NA	96 6 NA	NA 7 NA	NA NA NA	NA NA NA	NA NA NA
100	1130 NA NA	104 NA 2	1157 NA NA	NA NA NA	NA 3 NA	NA 4 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
97	959 NA NA	67 NA 2	965 NA NA	NA NA NA	NA 3 NA	NA 8 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
163	643 NA NA	219 NA 3	649 NA NA	188 NA NA	604 3 NA	NA 6 7	NA 7 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
166	878 NA NA	79 NA 2	824 NA NA	NA NA NA	NA 1 NA	NA 8 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
5	109 3 14	2 110 8	109 2 NA	2 110 NA	111 3 NA	2 9 NA	97 4 NA	0 10 NA	105 11 NA	2 12 NA	105 13 NA
193	512 NA NA	213 NA 3	526 NA NA	183 NA NA	503 1 NA	NA 6 NA	NA 7 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
31	70 NA NA	20 NA 6	63 NA NA	23 NA NA	57 3 NA	14 15 NA	60 16 NA	20 10 NA	57 17 NA	15 18 NA	59 NA NA
44	173 NA NA	80 NA 2	176 NA NA	NA NA NA	NA 3 NA	NA 9 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
308	967 NA NA	221 NA 2	982 NA NA	NA NA NA	NA 13 NA	NA 20 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA
168	608 NA NA	153 NA 3	614 NA NA	128 NA NA	629 13 NA	NA 19 NA	NA 20 NA	NA NA NA	NA NA NA	NA NA NA	NA NA NA

END

Data

#All DVT

list(NT=33,NS=22)

r[,1]	n[,1] r[,7] t[,8]	r[,2] n[,7] na[]	n[,2] r[,8]	r[,3] n[,8]	n[,3] t[,1]	r[,4] t[,2]	n[,4] t[,3]	r[,5] t[,4]	n[,5] t[,5]	r[,6] t[,6]	n[,6] t[,7]
243	997 NA NA	142 NA 2	971 NA	NA NA	NA 1	NA 4	NA NA	NA NA	NA NA	NA NA	NA NA
57	101 NA NA	42 NA 4	106 NA	34 NA	104 2	23 5	96 6	NA 7	NA NA	NA NA	NA NA
24	110 NA NA	6 NA 4	110 NA	14 NA	110 2	9 1	110 8	NA 9	NA NA	NA NA	NA NA
92	1122 NA NA	89 NA 2	1142 NA	NA NA	NA 3	NA 4	NA NA	NA NA	NA NA	NA NA	NA NA
86	959 NA NA	61 NA 2	965 NA	NA NA	NA 3	NA 10	NA NA	NA NA	NA NA	NA NA	NA NA
163	643 NA NA	219 NA 3	649 NA	188 NA	604 3	NA 6	NA 7	NA NA	NA NA	NA NA	NA NA
160	878 NA NA	79 NA 2	824 NA	NA NA	NA 1	NA 10	NA NA	NA NA	NA NA	NA NA	NA NA
48	79 NA NA	34 NA 4	78 NA	25 NA	74 2	25 11	84 1	NA 12	NA NA	NA NA	NA NA
17	109 6 18	29 110 8	109 9	10 110	97 3	11 13	97 4	5 14	105 15	13 16	105 17
192	699 NA NA	211 NA 3	708 NA	182 NA	694 1	NA 6	NA 7	NA NA	NA NA	NA NA	NA NA
31	70 NA NA	20 NA 6	63 NA	21 NA	57 3	14 19	60 20	20 10	57 21	15 22	59 NA
44	173 NA NA	79 NA 2	176 NA	NA NA	NA 3	NA 13	NA NA	NA NA	NA NA	NA NA	NA NA
56	228 NA NA	77 NA 2	225 NA	NA NA	NA 3	NA 23	NA NA	NA NA	NA NA	NA NA	NA NA
21	92 NA NA	25 NA 2	93 NA	NA NA	NA 1	NA 23	NA NA	NA NA	NA NA	NA NA	NA NA
36	51 NA NA	25 NA 3	50 NA	22 NA	49 2	NA 24	NA 25	NA NA	NA NA	NA NA	NA NA
16	67 NA NA	34 NA 2	63 NA	NA NA	NA 24	NA 9	NA NA	NA NA	NA NA	NA NA	NA NA
48	89 NA NA	57 NA 2	99 NA	NA NA	NA 1	NA 26	NA NA	NA NA	NA NA	NA NA	NA NA
23	97 NA NA	27 NA 5	63 NA	20 NA	101 3	29 27	87 28	16 29	95 30	NA NA	NA NA
301	960 NA NA	214 NA 2	976 NA	NA NA	NA 13	NA 31	NA NA	NA NA	NA NA	NA NA	NA NA
166	606 NA NA	151 NA 3	612 NA	124 NA	625 13	NA 30	NA 31	NA NA	NA NA	NA NA	NA NA
4	21 NA NA	8 NA 2	25 NA	NA NA	NA 13	NA 9	NA NA	NA NA	NA NA	NA NA	NA NA
9	12 NA NA	7 NA 4	9 NA	1 NA	12 2	1 32	10 33	NA 9	NA NA	NA NA	NA NA

END

Data

#All bleeds

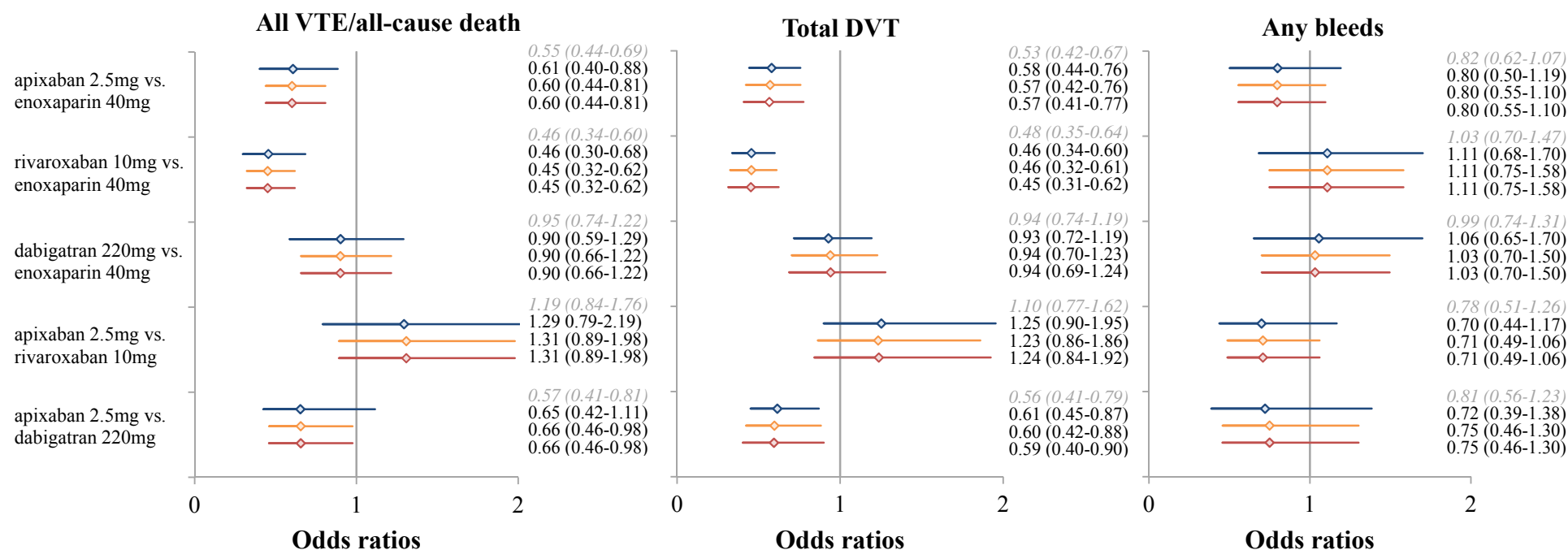
list(NT=29,NS=14)

r[,1]	n[,1] r[,7] t[,8]	r[,2] n[,7] na[]	n[,2] r[,8]	r[,3] n[,8]	n[,3] t[,1]	r[,4] t[,2]	n[,4] t[,3]	r[,5] t[,4]	n[,5] t[,5]	r[,6] t[,6]	n[,6] t[,7]
126	1508 NA NA	104 NA 2	1501 NA	NA NA	NA 2	NA 4	NA NA	NA NA	NA NA	NA NA	NA NA
10	124 NA NA	13 NA 4	133 NA	13 NA	126 1	14 5	129 6	NA 7	NA NA	NA NA	NA NA
108	1588 NA NA	85 NA 2	1596 NA	NA NA	NA 3	NA 4	NA NA	NA NA	NA NA	NA NA	NA NA
142	1508 NA NA	160 NA 2	1526 NA	NA NA	NA 3	NA 8	NA NA	NA NA	NA NA	NA NA	NA NA
60	1239 NA NA	30 NA 2	1220 NA	NA NA	NA 2	NA 8	NA NA	NA NA	NA NA	NA NA	NA NA
8	89 NA NA	5 NA 4	89 NA	7 NA	91 1	13 9	95 2	NA 10	NA NA	NA NA	NA NA
8	149 15 16	8 153 8	151 15	6 151	154 3	5 11	151 4	10 12	153 13	11 14	155 15
115	694 NA NA	116 NA 3	703 NA	111 NA	679 2	NA 6	NA 7	NA NA	NA NA	NA NA	NA NA
8	104 NA NA	9 NA 6	100 NA	9 NA	102 3	9 17	103 18	18 8	98 19	27 20	106 NA
3	125 NA NA	0 NA 5	85 NA	3 NA	134 3	2 21	126 22	3 23	130 24	NA NA	NA NA
44	1148 NA NA	58 NA 2	1151 NA	NA NA	NA 13	NA 25	NA NA	NA NA	NA NA	NA NA	NA NA
34	759 NA NA	36 NA 3	757 NA	41 NA	769 13	NA 24	NA 25	NA NA	NA NA	NA NA	NA NA
23	330 NA NA	31 NA 2	345 NA	NA NA	NA 13	NA 24	NA NA	NA NA	NA NA	NA NA	NA NA
4	87 NA NA	0 NA 5	86 NA	5 NA	85 1	3 26	84 27	4 28	84 29	NA NA	NA NA

END

Figure C3 Odds ratio for all VTE/all-cause death, total DVT, and any bleeds from random effects NMA models

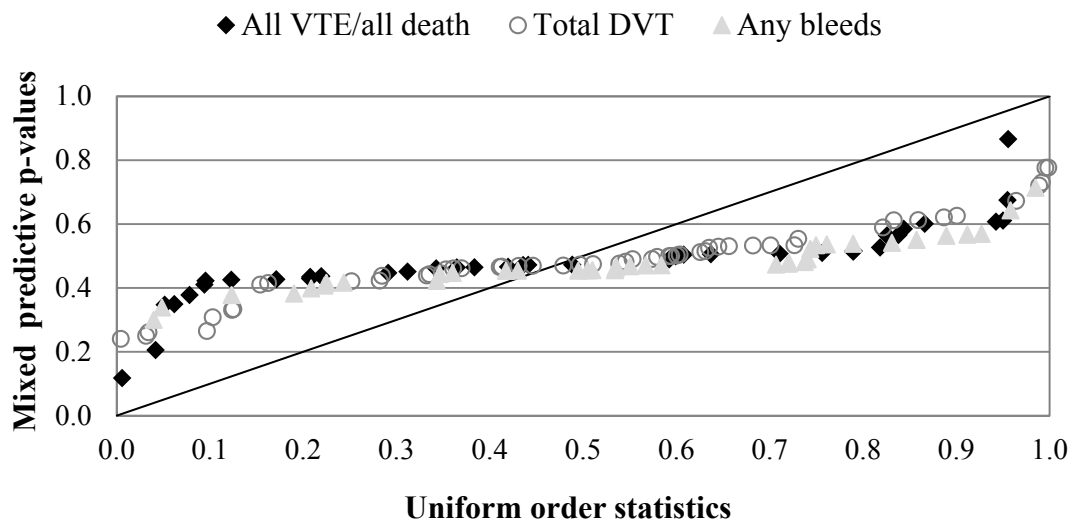
— 1st order — 2nd order — 3rd order



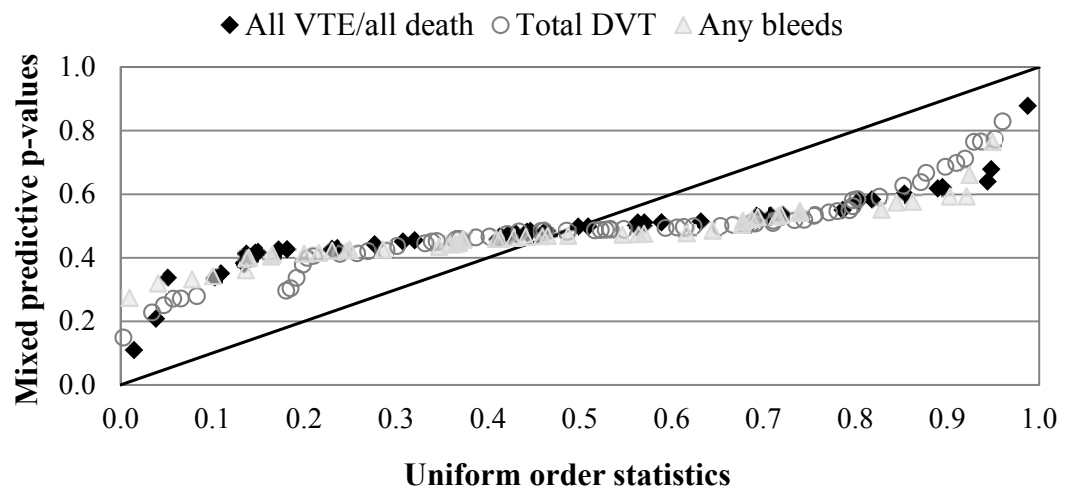
Results for the base case are not plotted on the graphs because a random effects model was not used for ITC; base case OR estimates are provided in italics next to the plots for reference only.

Figure C4 First, second and third order NMA inconsistency plots (mixed p-values)

a) 1st network order



b) 2nd network order



c) 3rd network order

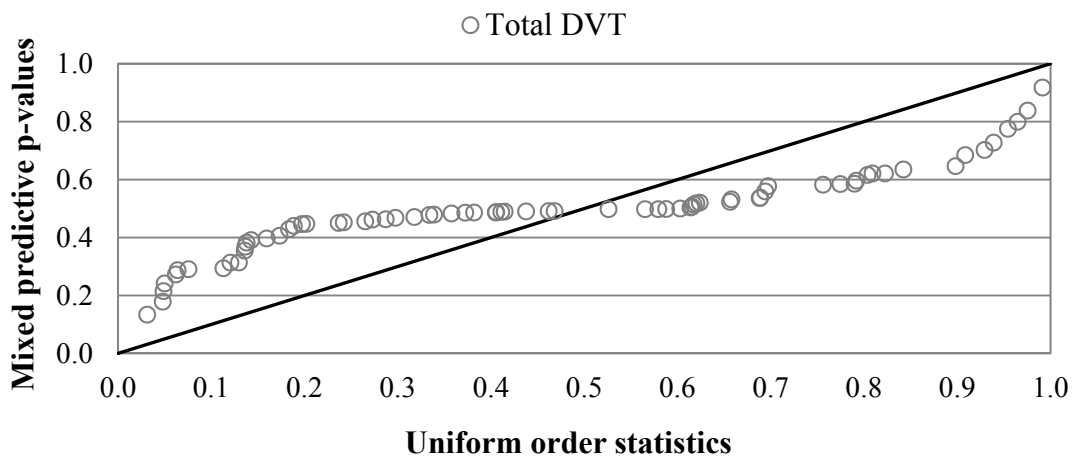


Table C7 Markov trace for apixaban (base case)

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
start(0)	1000															
1	855	128	11	-	-			-			-	-	6	994	773.59	£94,833
2	843	69	9	0	24	-	24	5	-	5	28	1	20	947	733.94	£105,641
3	832	57	38	0	25	24	2	8	5	3	6	0	33	903	695.38	£41,696
4	814	50	42	0	26	25	2	10	8	2	4	0	54	853	652.88	£32,980
5	796	45	44	0	27	26	1	11	10	1	2	0	75	806	612.88	£27,465
6	778	42	45	0	27	27	0	11	10	0	2	0	95	762	575.35	£25,339
7	761	40	45	0	27	26	0	11	11	0	1	0	115	720	540.03	£22,407
8	745	37	45	0	26	26	0	11	11	0	1	0	134	681	506.81	£21,499
9	716	35	44	0	26	25	0	11	11	0	1	0	167	633	467.78	£20,524
10	689	32	43	0	25	25	0	11	11	0	1	0	198	588	431.69	£19,521
11	663	30	42	0	24	24	0	11	11	0	1	0	229	547	398.32	£18,547
12	638	28	41	0	24	23	0	11	10	0	1	0	258	508	367.67	£17,602
13	614	26	40	0	23	23	0	11	10	0	1	0	286	472	339.66	£16,689
14	559	23	36	0	21	21	0	10	10	0	1	0	350	415	296.86	£15,067
15	509	20	34	0	19	19	0	9	9	0	1	0	408	365	259.44	£13,501
16	463	17	31	0	18	18	0	9	8	0	1	0	462	321	226.73	£12,088
17	422	15	28	0	16	16	0	8	8	0	0	0	510	283	198.13	£10,813
18	384	13	26	0	15	15	0	7	7	0	0	0	554	249	173.13	£9,666
19	349	12	24	0	14	14	0	7	7	0	0	0	594	219	151.27	£8,634
20	318	10	22	0	13	13	0	6	6	0	0	0	630	192	132.17	£7,708
21	289	9	20	0	12	12	0	6	6	0	0	0	663	169	115.47	£6,876
22	263	8	18	0	11	11	0	5	5	0	0	0	694	149	100.88	£6,131
23	240	7	17	0	10	10	0	5	5	0	0	0	721	131	88.13	£5,463
24	184	5	13	0	8	8	0	4	4	0	0	0	786	97	64.99	£4,177

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
25	142	4	10	0	6	6	0	3	3	0	0	0	835	72	47.92	£3,144
26	109	3	8	0	5	5	0	2	2	0	0	0	873	54	35.34	£2,365
27	84	2	6	0	4	4	0	2	2	0	0	0	903	40	26.06	£1,778
28	64	2	5	0	3	3	0	1	1	0	0	0	925	30	19.21	£1,336
29	49	1	4	0	2	2	0	1	1	0	0	0	942	22	14.16	£1,004
30	38	1	3	0	2	2	0	1	1	0	0	0	956	16	10.44	£753
31	29	1	2	0	1	1	0	1	1	0	0	0	966	12	7.70	£565
32	22	0	2	0	1	1	0	1	1	0	0	0	974	9	5.67	£424
33	17	0	1	0	1	1	0	0	0	0	0	0	980	7	4.18	£318
34	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£12
35	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£-

C8 Markov trace for enoxaparin (base case)

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
start(0)	1000															
1	737	232	19	-	-			-			-	-	11	988	768.44	£456,486
2	727	125	17	0	43	-	43	10	-	10	51	3	25	942	728.46	£191,063
3	717	103	68	0	46	43	3	15	9	6	12	1	38	898	688.86	£75,413
4	702	90	76	0	48	45	3	18	15	3	7	0	59	849	645.27	£59,649
5	686	82	80	0	49	47	3	19	17	2	4	0	79	802	604.27	£49,675
6	671	76	81	0	49	48	1	19	19	1	4	0	100	758	565.89	£45,828
7	657	71	82	0	48	48	1	20	19	1	2	0	119	716	529.81	£40,526
8	642	68	82	0	48	47	1	20	19	1	2	0	139	677	495.90	£38,883
9	618	63	80	0	47	46	1	20	19	1	2	0	171	629	456.44	£37,120
10	595	58	78	0	45	45	0	20	19	1	2	0	203	585	419.98	£35,307

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
11	572	54	76	0	44	44	0	20	19	1	1	0	233	544	386.33	£33,544
12	550	50	74	0	43	42	0	19	19	1	1	0	262	506	355.61	£31,835
13	529	47	72	0	42	41	0	19	19	0	1	0	290	470	327.86	£30,184
14	482	41	66	0	38	38	0	18	17	0	1	0	354	413	285.95	£27,250
15	439	36	61	0	35	35	0	17	16	0	1	0	412	363	249.38	£24,418
16	399	32	56	0	32	32	0	16	15	0	1	0	464	320	217.46	£21,862
17	364	28	51	0	30	29	0	14	14	0	1	0	512	281	189.61	£19,557
18	331	24	47	0	27	27	0	13	13	0	1	0	556	247	165.30	£17,482
19	301	21	43	0	25	25	0	12	12	0	1	0	596	217	144.10	£15,616
20	274	19	40	0	23	23	0	12	11	0	1	0	632	191	125.61	£13,940
21	250	16	36	0	21	21	0	11	11	0	1	0	665	168	109.48	£12,436
22	227	14	33	0	19	19	0	10	10	0	0	0	695	148	95.41	£11,088
23	207	13	30	0	18	18	0	9	9	0	0	0	723	130	83.14	£9,881
24	159	9	23	0	14	14	0	7	7	0	0	0	787	97	61.15	£7,554
25	122	7	18	0	11	11	0	6	6	0	0	0	836	72	44.97	£5,686
26	94	5	14	0	8	8	0	4	4	0	0	0	874	53	33.07	£4,277
27	72	4	11	0	6	6	0	3	3	0	0	0	903	40	24.32	£3,216
28	55	3	8	0	5	5	0	3	3	0	0	0	926	29	17.88	£2,417
29	43	2	6	0	4	4	0	2	2	0	0	0	943	22	13.14	£1,815
30	33	1	5	0	3	3	0	2	2	0	0	0	956	16	9.66	£1,363
31	25	1	4	0	2	2	0	1	1	0	0	0	966	12	7.10	£1,023
32	19	1	3	0	2	2	0	1	1	0	0	0	974	9	5.22	£767
33	15	1	2	0	1	1	0	1	1	0	0	0	980	7	3.83	£575
34	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£22
35	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£-

C9 Markov trace for rivaroxaban (base case)

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
start(0)	1000															
1	880	106	9	-	-			-			-	-	5	995	774.67	£110,004
2	868	57	8	0	20	-	20	4	-	4	23	1	19	920	735.09	£87,469
3	856	47	31	0	21	19	1	7	4	3	5	0	32	892	696.75	£34,524
4	837	41	35	0	22	20	2	8	7	1	3	0	53	844	654.49	£27,307
5	819	37	36	0	23	21	1	9	8	1	2	0	74	798	614.70	£22,741
6	801	35	37	0	22	22	0	9	9	0	2	0	94	754	577.34	£20,980
7	784	33	38	0	22	22	0	9	9	0	1	0	114	713	542.19	£18,553
8	766	31	37	0	22	22	0	9	9	0	1	0	133	673	509.12	£17,801
9	737	29	37	0	21	21	0	9	9	0	1	0	166	626	470.19	£16,993
10	709	27	36	0	21	21	0	9	9	0	1	0	198	581	434.17	£16,163
11	683	25	35	0	20	20	0	9	9	0	1	0	228	540	400.86	£15,356
12	657	23	34	0	20	19	0	9	9	0	1	0	257	502	370.22	£14,574
13	632	21	33	0	19	19	0	9	9	0	1	0	285	467	342.16	£13,818
14	575	19	30	0	18	17	0	8	8	0	1	0	349	410	299.18	£12,475
15	524	16	28	0	16	16	0	8	7	0	0	0	408	361	261.58	£11,179
16	477	14	25	0	15	15	0	7	7	0	0	0	461	317	228.70	£10,008
17	434	13	23	0	14	13	0	7	6	0	0	0	509	279	199.94	£8,953
18	395	11	22	0	13	12	0	6	6	0	0	0	553	245	174.79	£8,003
19	359	10	20	0	12	11	0	6	6	0	0	0	593	216	152.79	£7,149
20	327	9	18	0	11	10	0	5	5	0	0	0	630	190	133.56	£6,382
21	298	8	17	0	10	10	0	5	5	0	0	0	663	167	116.75	£5,693
22	271	7	15	0	9	9	0	5	4	0	0	0	693	147	102.04	£5,076
23	247	6	14	0	8	8	0	4	4	0	0	0	721	129	89.19	£4,523
24	190	4	11	0	6	6	0	3	3	0	0	0	785	96	65.80	£3,458

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
25	146	3	8	0	5	5	0	3	3	0	0	0	835	71	48.55	£2,603
26	112	2	6	0	4	4	0	2	2	0	0	0	873	53	35.82	£1,958
27	86	2	5	0	3	3	0	2	2	0	0	0	903	39	26.42	£1,472
28	66	1	4	0	2	2	0	1	1	0	0	0	925	29	19.49	£1,106
29	51	1	3	0	2	2	0	1	1	0	0	0	942	22	14.38	£831
30	39	1	2	0	1	1	0	1	1	0	0	0	956	16	10.61	£624
31	30	0	2	0	1	1	0	1	1	0	0	0	966	12	7.82	£468
32	23	0	1	0	1	1	0	0	0	0	0	0	974	9	5.77	£351
33	18	0	1	0	1	1	0	0	0	0	0	0	980	7	4.26	£263
34	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£10
35	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£-

C10 Markov trace for dabigatran etexilate (base case)

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
start(0)	1000															
1	749	221	19	-	-			-			-	-	11	989	768.97	£129,728
2	739	120	16	0	41	-	41	9	-	9	48	3	24	943	729.02	£182,313
3	729	99	65	0	43	41	3	14	9	5	11	1	37	899	689.52	£71,959
4	713	86	73	0	46	43	3	17	14	3	6	0	58	849	646.05	£56,917
5	698	78	76	0	47	45	2	18	17	2	4	0	79	803	605.15	£47,399
6	682	72	77	0	47	46	1	19	18	1	3	0	99	758	566.86	£43,729
7	667	68	79	0	46	46	1	19	18	1	2	0	119	717	530.86	£38,670
8	653	64	78	0	46	45	1	19	18	1	2	0	138	677	497.02	£37,102
9	628	60	76	0	44	44	1	19	18	1	2	0	171	630	457.60	£35,420
10	604	56	74	0	43	43	0	19	18	1	1	0	202	585	421.18	£33,690

Cycle	Well	Untreated VTE	Treated VTE	Disabled	PTs ynd	# old	# new	Severe PTsynd	# old	# new	DVT	PE	Death	Life-years	QALYs	Costs
11	581	52	72	0	42	42	0	19	18	1	1	0	232	544	387.56	£32,008
12	559	48	70	0	41	40	0	18	18	1	1	0	262	506	356.85	£30,377
13	538	45	68	0	40	39	0	18	18	0	1	0	289	470	329.07	£28,802
14	490	39	63	0	37	36	0	17	17	0	1	0	353	414	287.07	£26,002
15	446	34	58	0	34	33	0	16	16	0	1	0	411	364	250.41	£23,300
16	406	30	53	0	31	31	0	15	14	0	1	0	464	320	218.41	£20,860
17	369	26	49	0	28	28	0	14	13	0	1	0	512	281	190.48	£18,661
18	336	23	45	0	26	26	0	13	13	0	1	0	556	247	166.11	£16,682
19	306	20	41	0	24	24	0	12	12	0	1	0	596	218	144.84	£14,901
20	279	18	38	0	22	22	0	11	11	0	1	0	632	191	126.28	£13,302
21	254	16	35	0	20	20	0	10	10	0	1	0	665	168	110.09	£11,867
22	231	14	32	0	19	18	0	9	9	0	0	0	695	148	95.97	£10,580
23	210	12	29	0	17	17	0	9	9	0	0	0	722	130	83.65	£9,428
24	162	9	22	0	13	13	0	7	7	0	0	0	787	97	61.55	£7,208
25	124	7	17	0	10	10	0	5	5	0	0	0	836	72	45.28	£5,425
26	95	5	13	0	8	8	0	4	4	0	0	0	874	53	33.30	£4,081
27	73	4	10	0	6	6	0	3	3	0	0	0	903	40	24.49	£3,069
28	56	3	8	0	5	5	0	3	3	0	0	0	925	29	18.01	£2,306
29	43	2	6	0	4	4	0	2	2	0	0	0	943	22	13.25	£1,732
30	33	1	5	0	3	3	0	2	2	0	0	0	956	16	9.74	£1,300
31	26	1	4	0	2	2	0	1	1	0	0	0	966	12	7.16	£976
32	20	1	3	0	2	2	0	1	1	0	0	0	974	9	5.26	£732
33	15	1	2	0	1	1	0	1	1	0	0	0	980	7	3.87	£549
34	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£21
35	0	0	0	0	0	-	-	0	-	-	0	0	1000	-	-	£-

Appendix D

D1 Assumed effect on vertebral and hip fractures following an initial vertebral fracture for patients taking a bisphosphonate

	Relative risk for vertebral fractures	Relative risk for hip fractures
Year 1	0.580	0.580
Year 2	0.580	0.580
Year 3	0.580	0.580
Year 4	0.580	0.580
Year 5	0.580	0.720
Year 6	0.664	0.776
Year 7	0.748	0.832
Year 8	0.832	0.888
Year 9	0.916	0.944
Year 10	1.000	1.000

Calculated from technology assessment report for TA279 [Stevenson 2012]

D2 Fixed effect meta-analysis model for overall mortality at 12 months with power transform prior model

```
#Fixed effect meta-analysis for RCTs
#Careful input data for OBV adjusted

model{
  for (j in 1:alpha.n) {
    for (i in 1:n.RCT) {
      y.RCT.x[i,j] <- y.RCT[i]
      p.RCT.x[i,j] <- pow(sd.RCT[i],-2) }
    d.OBS.x[j] <- d.OBS
    sd.OBS.x[j]<- sd.OBS
    p.OBS.x[j] <- pow(sd.OBS,-2) }

    for (j in 1:alpha.n) {
      for (i in 1:n.RCT) {
        y.RCT.x[i,j] ~ dnorm(d.RCT[j],p.RCT.x[i,j]) }
        d.RCT[j] ~ dnorm(d.OBS.x[j],p.OBS.star[j])
        p.OBS.star[j] <- p.OBS.x[j]*alpha[j] }
    }
  }

#Data
list(n.RCT=3, alpha.n=11,
y.RCT=c(-0.733969175, -0.040821995, -0.186329578),
sd.RCT=c(0.693647969, 1.399887593, 0.59128892),
d.OBS=-0.09227,
sd.OBS=0.009699,
alpha=c(0.0001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1))

#Initials
list(d.RCT=c(0,0,0,0,0, 0,0,0,0,0, 0)
```

D3 Fixed effect meta-analysis model for overall mortality at 12 months with bias allowance model

```
#Fixed effect meta-analysis for RCTs
#Careful input data for OBV adjusted

model{
  for (i in 1:n.RCT) {
    p.RCT[i] <- pow(sd.RCT[i],-2)
    y.RCT[i] ~ dnorm(d.RCT,p.RCT[i]) }
    d.RCT ~ dnorm(d.OBS,p.OBS.adj)
    p.OBS.adj <- 1/(pow(sd.OBS,2) + sigma2)

  # where sd.OBS is from OBS FE model
  # and sigma2 are values from empirical evidence
  }

#Data adjusted
list(n.RCT=3,
y.RCT=c(-0.733969175, -0.040821995, -0.186329578),
sd.RCT=c(0.693647969, 1.399887593, 0.59128892),
d.OBS=-0.09227,
sd.OBS=0.009699,
sigma2=1)

#sigma2 was tested for 0, 0.02, 0.08, and 0.24, 1

#Initials
list(d.RCT=0)
```

D4 Annual all-cause mortality hazards for England and Wales

Based on data for the years 2010-2012, available from ONS [2014]

Age	Males	Females
60	0.008075	0.005307
61	0.008753	0.005765
62	0.009557	0.006224
63	0.010220	0.006593
64	0.011196	0.007272
65	0.012172	0.007953
66	0.013871	0.008949
67	0.015105	0.009741
68	0.016289	0.010575
69	0.018507	0.011966
70	0.020758	0.013537
71	0.022957	0.014549
72	0.025137	0.016410
73	0.027212	0.017830
74	0.030368	0.019995
75	0.033276	0.022213
76	0.037166	0.025052
77	0.041029	0.028359
78	0.046088	0.031973
79	0.050915	0.035927
80	0.057068	0.040905
81	0.064854	0.045894
82	0.072440	0.052108
83	0.080695	0.059757
84	0.090335	0.067561
85	0.100832	0.075578
86	0.112255	0.085610
87	0.124962	0.094929
88	0.137492	0.107177
89	0.156663	0.120614
90	0.169236	0.138698
91	0.183155	0.149474
92	0.196816	0.165912
93	0.211629	0.176483
94	0.235581	0.200225
95	0.259230	0.222489
96	0.281894	0.243500
97	0.303375	0.261584
98	0.322859	0.279844
99	0.344662	0.301601
100	0.360908	0.323829
101	1.000000	1.000000

D5 Power transform prior model results for PVP vs. OPM according to α

Alpha	HR (95% CrI)
0.0	0.71 (0.33;1.52)
0.1	0.91 (0.86;0.97)
0.2	0.91 (0.87;0.95)
0.3	0.91 (0.88;0.94)
0.4	0.91 (0.88;0.94)
0.5	0.91 (0.89;0.94)
0.6	0.91 (0.89;0.93)
0.7	0.91 (0.89;0.93)
0.8	0.91 (0.89;0.93)
0.9	0.91 (0.89;0.93)
1.0	0.91 (0.89;0.93)

D6 Markov trace for PVP (Scenario 1 - no mortality benefit)

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
(start) 0	1,000						£3,549,000		
1 month	998	0	1			1	£6,252	80	4.36
2 month	996	1	1	0	0	2	£6,475	80	4.35
3 month	993	1	2	0	0	4	£6,688	80	4.35
4 month	991	2	3	0	0	5	£6,892	80	4.34
5 month	989	2	3	0	0	6	£7,087	80	4.33
6 month	987	2	4	0	0	7	£7,275	80	4.33
7 month	985	3	4	0	0	9	£7,455	80	4.32
8 month	982	3	5	0	0	10	£7,628	80	4.31
9 month	980	3	5	0	0	11	£7,794	80	4.31
10 month	978	4	5	0	0	13	£7,954	79	4.30
11 month	976	4	6	0	0	14	£8,108	79	4.29
12 month	974	5	6	0	0	16	£8,256	79	4.29
13 month	971	5	6	0	0	17	£8,113	76	4.14
14 month	969	5	7	0	0	19	£8,245	76	4.13
15 month	967	6	7	0	0	20	£8,372	76	4.12
16 month	965	6	7	0	0	22	£8,495	76	4.11
17 month	962	6	7	0	0	24	£8,613	76	4.11
18 month	960	7	8	0	0	25	£8,728	76	4.10
19 month	958	7	8	0	0	27	£8,838	76	4.09
20 month	956	7	8	0	0	29	£8,945	76	4.08
21 month	954	8	8	0	0	30	£9,048	75	4.08
22 month	951	8	8	0	0	32	£9,148	75	4.07
23 month	949	8	9	0	0	34	£9,245	75	4.06

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
24 month	947	9	9	0	0	35	£9,339	75	4.05
25 month	945	9	9	0	0	37	£9,109	72	3.91
26 month	942	9	9	0	0	39	£9,192	72	3.90
27 month	940	10	9	0	0	41	£9,273	72	3.89
28 month	938	10	9	0	0	43	£9,351	72	3.88
29 month	935	10	9	0	0	45	£9,427	72	3.88
30 month	933	11	10	0	0	47	£9,501	72	3.87
31 month	931	11	10	0	0	49	£9,572	72	3.86
32 month	928	11	10	0	0	50	£9,642	71	3.85
33 month	926	12	10	0	0	52	£9,709	71	3.84
34 month	924	12	10	0	0	54	£9,775	71	3.84
35 month	922	12	10	0	0	56	£9,839	71	3.83
36 month	919	12	10	0	0	58	£9,901	71	3.82
4 year	891	16	17	0	0	76	£65,407	805	520.40
5 year	862	19	23	0	0	96	£63,340	761	490.82
6 year	826	23	32	1	0	119	£90,152	717	461.28
7 year	787	27	42	1	1	143	£92,553	674	431.92
8 year	747	31	51	1	1	170	£94,040	631	402.93
9 year	704	36	59	2	1	198	£94,687	589	374.48
10 year	660	41	67	2	2	227	£94,552	548	346.67
11 year	610	46	76	3	3	263	£117,365	505	317.64
12 year	560	50	83	4	3	301	£108,912	463	289.36
13 year	511	53	87	5	4	341	£100,350	422	261.84
14 year	462	54	90	5	5	383	£91,723	381	235.12
15 year	415	55	91	6	6	427	£83,182	342	209.51
16 year	362	54	97	8	6	472	£103,447	305	184.86

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
17 year	313	53	100	9	7	518	£91,297	269	161.60
18 year	267	50	101	10	8	563	£79,883	235	140.10
19 year	226	47	100	11	9	607	£69,182	204	120.26
20 year	188	43	96	12	9	652	£59,156	175	101.89
21 year	153	38	88	12	9	699	£49,476	146	84.26
22 year	124	34	80	12	9	742	£40,940	121	69.11
23 year	98	29	72	11	9	781	£33,432	99	55.92
24 year	77	24	63	11	9	817	£26,998	80	44.77
25 year	59	20	54	10	8	850	£21,325	64	34.99
26 year	44	16	45	9	7	880	£16,465	49	26.74
27 year	32	13	36	7	6	906	£12,443	37	20.01
28 year	23	10	29	6	5	928	£9,227	28	14.70
29 year	16	7	22	5	4	945	£6,717	20	10.60
30 year	11	5	17	4	4	959	£4,785	14	7.48
31 year	7	4	12	3	3	971	£3,333	10	5.16
32 year	0	0	0	0	0	1000	£557	-	-0.02
33 year	0	0	0	0	0	1000	£-	-	-
34 year	0	0	0	0	0	1000	£-	-	-
35 year	0	0	0	0	0	1000	£-	-	-
36 year	0	0	0	0	0	1000	£-	-	-
37 year	0	0	0	0	0	1000	£-	-	-
38 year	0	0	0	0	0	1000	£-	-	-
39 year	0	0	0	0	0	1000	£-	-	-
40 year	0	0	0	0	0	1000	£-	-	-
41 year	0	0	0	0	0	1000	£-	-	-
42 year	0	0	0	0	0	1000	£-	-	-

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
43 year	0	0	0	0	0	1000	£-	-	-
44 year	0	0	0	0	0	1000	£-	-	-
45 year	0	0	0	0	0	1000	£-	-	-
46 year	0	0	0	0	0	1000	£-	-	-
47 year	0	0	0	0	0	1000	£-	-	-
48 year	0	0	0	0	0	1000	£-	-	-
49 year	0	0	0	0	0	1000	£-	-	-
50 year	0	0	0	0	0	1000	£-	-	-

D7 Markov trace for BKP (Scenario 1 - no mortality benefit)

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
(start) 0	1,000						£5,133,000		
1 month	998	0	1			1	£6,252	80	4.22
2 month	996	1	1	0	0	2	£6,475	80	4.22
3 month	993	1	2	0	0	4	£6,688	80	4.21
4 month	991	2	3	0	0	5	£6,892	80	4.21
5 month	989	2	3	0	0	6	£7,087	80	4.20
6 month	987	2	4	0	0	7	£7,275	80	4.19
7 month	985	3	4	0	0	9	£7,455	80	4.19
8 month	982	3	5	0	0	10	£7,628	80	4.18
9 month	980	3	5	0	0	11	£7,794	80	4.18
10 month	978	4	5	0	0	13	£7,954	79	4.17
11 month	976	4	6	0	0	14	£8,108	79	4.16

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
12 month	974	5	6	0	0	16	£8,256	79	4.16
13 month	971	5	6	0	0	17	£8,113	76	4.01
14 month	969	5	7	0	0	19	£8,245	76	4.00
15 month	967	6	7	0	0	20	£8,372	76	3.99
16 month	965	6	7	0	0	22	£8,495	76	3.99
17 month	962	6	7	0	0	24	£8,613	76	3.98
18 month	960	7	8	0	0	25	£8,728	76	3.97
19 month	958	7	8	0	0	27	£8,838	76	3.97
20 month	956	7	8	0	0	29	£8,945	76	3.96
21 month	954	8	8	0	0	30	£9,048	75	3.95
22 month	951	8	8	0	0	32	£9,148	75	3.94
23 month	949	8	9	0	0	34	£9,245	75	3.94
24 month	947	9	9	0	0	35	£9,339	75	3.93
25 month	945	9	9	0	0	37	£9,109	72	3.91
26 month	942	9	9	0	0	39	£9,192	72	3.90
27 month	940	10	9	0	0	41	£9,273	72	3.89
28 month	938	10	9	0	0	43	£9,351	72	3.88
29 month	935	10	9	0	0	45	£9,427	72	3.88
30 month	933	11	10	0	0	47	£9,501	72	3.87
31 month	931	11	10	0	0	49	£9,572	72	3.86
32 month	928	11	10	0	0	50	£9,642	71	3.85
33 month	926	12	10	0	0	52	£9,709	71	3.84
34 month	924	12	10	0	0	54	£9,775	71	3.84
35 month	922	12	10	0	0	56	£9,839	71	3.83
36 month	919	12	10	0	0	58	£9,901	71	3.82
4 year	891	16	17	0	0	76	£65,407	805	520.40

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
5 year	862	19	23	0	0	96	£63,340	761	490.82
6 year	826	23	32	1	0	119	£90,152	717	461.28
7 year	787	27	42	1	1	143	£92,553	674	431.92
8 year	747	31	51	1	1	170	£94,040	631	402.93
9 year	704	36	59	2	1	198	£94,687	589	374.48
10 year	660	41	67	2	2	227	£94,552	548	346.67
11 year	610	46	76	3	3	263	£117,365	505	317.64
12 year	560	50	83	4	3	301	£108,912	463	289.36
13 year	511	53	87	5	4	341	£100,350	422	261.84
14 year	462	54	90	5	5	383	£91,723	381	235.12
15 year	415	55	91	6	6	427	£83,182	342	209.51
16 year	362	54	97	8	6	472	£103,447	305	184.86
17 year	313	53	100	9	7	518	£91,297	269	161.60
18 year	267	50	101	10	8	563	£79,883	235	140.10
19 year	226	47	100	11	9	607	£69,182	204	120.26
20 year	188	43	96	12	9	652	£59,156	175	101.89
21 year	153	38	88	12	9	699	£49,476	146	84.26
22 year	124	34	80	12	9	742	£40,940	121	69.11
23 year	98	29	72	11	9	781	£33,432	99	55.92
24 year	77	24	63	11	9	817	£26,998	80	44.77
25 year	59	20	54	10	8	850	£21,325	64	34.99
26 year	44	16	45	9	7	880	£16,465	49	26.74
27 year	32	13	36	7	6	906	£12,443	37	20.01
28 year	23	10	29	6	5	928	£9,227	28	14.70
29 year	16	7	22	5	4	945	£6,717	20	10.60
30 year	11	5	17	4	4	959	£4,785	14	7.48

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
31 year	7	4	12	3	3	971	£3,333	10	5.16
32 year	0	0	0	0	0	1000	£557	-	-0.02
33 year	0	0	0	0	0	1000	£-	-	-
34 year	0	0	0	0	0	1000	£-	-	-
35 year	0	0	0	0	0	1000	£-	-	-
36 year	0	0	0	0	0	1000	£-	-	-
37 year	0	0	0	0	0	1000	£-	-	-
38 year	0	0	0	0	0	1000	£-	-	-
39 year	0	0	0	0	0	1000	£-	-	-
40 year	0	0	0	0	0	1000	£-	-	-
41 year	0	0	0	0	0	1000	£-	-	-
42 year	0	0	0	0	0	1000	£-	-	-
43 year	0	0	0	0	0	1000	£-	-	-
44 year	0	0	0	0	0	1000	£-	-	-
45 year	0	0	0	0	0	1000	£-	-	-
46 year	0	0	0	0	0	1000	£-	-	-
47 year	0	0	0	0	0	1000	£-	-	-
48 year	0	0	0	0	0	1000	£-	-	-
49 year	0	0	0	0	0	1000	£-	-	-
50 year	0	0	0	0	0	1000	£-	-	-

D8 Markov trace for OPM (Scenario 1 - no mortality benefit)

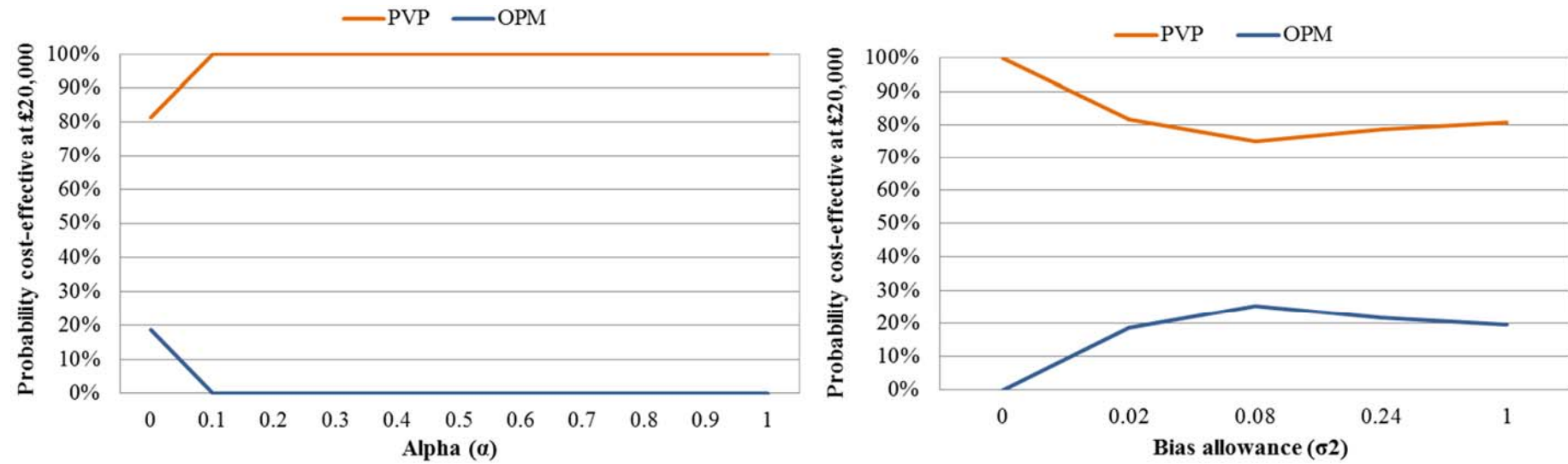
Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
(start) 0	1,000						£2,204,000		
1 month	998	0	1			1	£6,252	80	3.64
2 month	996	1	1	0	0	2	£6,475	80	3.63
3 month	993	1	2	0	0	4	£6,688	80	3.63
4 month	991	2	3	0	0	5	£6,892	80	3.62
5 month	989	2	3	0	0	6	£7,087	80	3.62
6 month	987	2	4	0	0	7	£7,275	80	3.61
7 month	985	3	4	0	0	9	£7,455	80	3.61
8 month	982	3	5	0	0	10	£7,628	80	3.60
9 month	980	3	5	0	0	11	£7,794	80	3.60
10 month	978	4	5	0	0	13	£7,954	79	3.59
11 month	976	4	6	0	0	14	£8,108	79	3.59
12 month	974	5	6	0	0	16	£8,256	79	3.58
13 month	971	5	6	0	0	17	£8,113	76	3.46
14 month	969	5	7	0	0	19	£8,245	76	3.45
15 month	967	6	7	0	0	20	£8,372	76	3.45
16 month	965	6	7	0	0	22	£8,495	76	3.44
17 month	962	6	7	0	0	24	£8,613	76	3.43
18 month	960	7	8	0	0	25	£8,728	76	3.43
19 month	958	7	8	0	0	27	£8,838	76	3.42
20 month	956	7	8	0	0	29	£8,945	76	3.42
21 month	954	8	8	0	0	30	£9,048	75	3.41
22 month	951	8	8	0	0	32	£9,148	75	3.40
23 month	949	8	9	0	0	34	£9,245	75	3.40

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
24 month	947	9	9	0	0	35	£9,339	75	3.39
25 month	945	9	9	0	0	37	£9,109	72	3.91
26 month	942	9	9	0	0	39	£9,192	72	3.90
27 month	940	10	9	0	0	41	£9,273	72	3.89
28 month	938	10	9	0	0	43	£9,351	72	3.88
29 month	935	10	9	0	0	45	£9,427	72	3.88
30 month	933	11	10	0	0	47	£9,501	72	3.87
31 month	931	11	10	0	0	49	£9,572	72	3.86
32 month	928	11	10	0	0	50	£9,642	71	3.85
33 month	926	12	10	0	0	52	£9,709	71	3.84
34 month	924	12	10	0	0	54	£9,775	71	3.84
35 month	922	12	10	0	0	56	£9,839	71	3.83
36 month	919	12	10	0	0	58	£9,901	71	3.82
4 year	891	16	17	0	0	76	£65,407	805	520.40
5 year	862	19	23	0	0	96	£63,340	761	490.82
6 year	825	23	32	1	0	119	£90,132	717	461.16
7 year	787	27	42	1	1	144	£92,506	673	431.68
8 year	746	31	51	1	1	170	£93,959	630	402.55
9 year	703	36	59	2	1	199	£94,565	588	373.95
10 year	659	41	67	2	2	229	£94,379	547	345.97
11 year	608	46	76	3	3	265	£117,060	504	316.74
12 year	557	50	83	4	3	303	£108,527	461	288.25
13 year	508	52	87	5	4	344	£99,877	419	260.51
14 year	459	54	90	5	5	387	£91,152	379	233.53
15 year	411	54	91	6	6	432	£82,506	339	207.66
16 year	357	54	97	8	6	478	£102,303	301	182.75

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
17 year	307	52	100	9	7	524	£90,004	265	159.23
18 year	261	49	100	10	8	571	£78,455	231	137.52
19 year	219	46	99	11	9	616	£67,631	200	117.48
20 year	181	41	95	12	9	662	£57,502	170	98.96
21 year	146	37	87	12	9	710	£47,738	141	81.20
22 year	116	32	79	11	9	753	£39,163	116	66.02
23 year	91	27	70	11	9	793	£31,659	94	52.88
24 year	70	22	62	10	8	828	£25,277	75	41.86
25 year	52	18	52	9	7	861	£19,693	59	32.26
26 year	38	14	43	8	6	891	£14,952	45	24.24
27 year	26	10	35	7	5	916	£11,078	33	17.78
28 year	18	8	27	6	5	937	£8,033	24	12.77
29 year	12	5	21	5	4	953	£5,704	17	8.99
30 year	8	4	16	4	3	966	£3,952	12	6.17
31 year	5	3	11	3	2	976	£2,669	8	4.13
32 year	0	0	0	0	0	1000	£420	-	-0.01
33 year	0	0	0	0	0	1000	£-	-	-
34 year	0	0	0	0	0	1000	£-	-	-
35 year	0	0	0	0	0	1000	£-	-	-
36 year	0	0	0	0	0	1000	£-	-	-
37 year	0	0	0	0	0	1000	£-	-	-
38 year	0	0	0	0	0	1000	£-	-	-
39 year	0	0	0	0	0	1000	£-	-	-
40 year	0	0	0	0	0	1000	£-	-	-
41 year	0	0	0	0	0	1000	£-	-	-
42 year	0	0	0	0	0	1000	£-	-	-

Cycle	Post-op VCF	Additional vertebral fracture (VF)	Additional hip fracture	VF + Hip fractures	Hip + VF fractures	Dead	Total costs	Total life-years	Total QALYs
43 year	0	0	0	0	0	1000	£-	-	-
44 year	0	0	0	0	0	1000	£-	-	-
45 year	0	0	0	0	0	1000	£-	-	-
46 year	0	0	0	0	0	1000	£-	-	-
47 year	0	0	0	0	0	1000	£-	-	-
48 year	0	0	0	0	0	1000	£-	-	-
49 year	0	0	0	0	0	1000	£-	-	-
50 year	0	0	0	0	0	1000	£-	-	-

D9 Cost-effectiveness acceptability curves according to α and σ^2 assuming mortality benefit over a lifetime



Appendix E

E1 Stata do-file—Weibull regression—subgroup analysis (PLATO-INVASIVE population)

```
Weibull fit 0.1.do* - Printed on 20/11/2015

1 rename var1 time
2 rename var2 fail
3 gen surv = 1 - fail
4 gen failp = fail*100
5 *create log cumulative hazard
6 *Weibull function
7 gen lnch = ln(-ln(surv))
8 gen lnt = ln(time)
9 *scatter lnch lnt
10 regress lnch lnt
11 di "lambda = `=exp(_b[_cons])', gamma = `=_b[lnt]'"
12 local lambda = exp(_b[_cons])
13 local gamma = _b[lnt]
14 predict predict
15 replace predict = exp(-exp(predict))
16 *tway scatter surv time || line predict time, sort
17 list if predict==.
18 gen failure = 1 - predict
19 gen failurep = failure*100
20 twoway scatter failp time || line failurep time, sort
21 *if statement lists transprop only once at top of column
22 gen transprop = 1 - exp((`lambda'*(365-365)^`gamma')-
(`lambda'*(365)^`gamma')) if _n==1
```

E2 Stata do-file—‘simplato’—for all-cause death, base case analysis (PLATO population)

```
simulation (deathFULL 0.6) for chapter.do - Printed on 22/10/2015

1 capture program drop simplato
2 program define simplato, rclass
3 *death, parameters from NICE Manufacturer Submission (p131)
4 syntax [, LAMBdas(real 0.0046) GAMmas(real 0.4360) TRT(real -
0.243) OBS(integer 18000) ]
5 clear
6 cd "\\uol.le.ac.uk\root\staff\home\p\pd135\My
Documents\Projects\Ticagrelor\Simulation"
7 set obs `obs'
8 gen recruit = 1+int((669-1+1)*runiform())
9 gen trt = rbinomial(1, 0.5)
10 survsim stime1 died, lambdas(`lambdas') gammas(`gammas')
cov(trt `trt') maxt(365)
11
12 stset stime1, f(died=1)
13 stpm2 trt, df(1) scale(h) noorthog
14 gen time365=365 in 1
15 predict surv365_trt0, surv timevar(time365) at(trt 0) ci
16 predict surv365_trt1, surv timevar(time365) at(trt 1) ci
17 return scalar trt365=_b[trt]
18 return scalar hr365=exp(_b[trt])
19 return scalar gamma365=_b[_rcs1]
20 return scalar lambda365=exp(_b[_cons])
21 return scalar trt365se=_se[trt]
22 return scalar gamma365se=_se[_rcs1]
23 return scalar lnlambda365se=_se[_cons]
24 return scalar surv365_trt0=surv365_trt0[1]
25 return scalar surv365_trt1=surv365_trt1[1]
26 return scalar surv365_trt0_lci=surv365_trt0_lci[1]
27 return scalar surv365_trt0_uci=surv365_trt0_uci[1]
28 return scalar surv365_trt1_lci=surv365_trt1_lci[1]
29 return scalar surv365_trt1_uci=surv365_trt1_uci[1]
30 return scalar death365_trt0=1-surv365_trt0[1]
31 return scalar death365_trt0_lci=1-surv365_trt0_uci[1]
32 return scalar death365_trt0_uci=1-surv365_trt0_lci[1]
33
34 *30 days analysis
35 stset stime1, f(died=1) exit(time 1*(30))
36 stpm2 trt, df(1) scale(h) noorthog
37 gen time30=30 in 1
38 predict surv30_trt0, surv timevar(time30) at(trt 0) ci
39 predict surv30_trt1, surv timevar(time30) at(trt 1) ci
40 return scalar trt30=_b[trt]
41 return scalar hr30=exp(_b[trt])
42 return scalar gamma30=_b[_rcs1]
43 return scalar lambda30=exp(_b[_cons])
44 return scalar trt30se=_se[trt]
45 return scalar gamma30se=_se[_rcs1]
46 return scalar lnlambda30se=_se[_cons]
47 return scalar surv30_trt0=surv30_trt0[1]
48 return scalar surv30_trt1=surv30_trt1[1]
49 return scalar surv30_trt0_lci=surv30_trt0_lci[1]
50 return scalar surv30_trt0_uci=surv30_trt0_uci[1]
51 return scalar surv30_trt1_lci=surv30_trt1_lci[1]
```

```

52 return scalar surv30_trt1_uci=surv30_trt1_uci[1]
53 return scalar death30_trt0=1-surv30_trt0[1]
54 return scalar death30_trt0_lci=1-surv30_trt0_uci[1]
55 return scalar death30_trt0_uci=1-surv30_trt0_lci[1]
56
57 *180 days analysis (i.e. 6 months)
58 stset stime1, f(died=1) exit(time 1*(180))
59 stpm2 trt, df(1) scale(h) noorthog
60 gen time180=180 in 1
61 predict surv180_trt0, surv timevar(time180) at(trt 0) ci
62 predict surv180_trt1, surv timevar(time180) at(trt 1) ci
63 return scalar trt180=_b[trt]
64 return scalar hr180=exp(_b[trt])
65 return scalar gamma180=_b[_rcs1]
66 return scalar lambda180=exp(_b[_cons])
67 return scalar trt180se=_se[trt]
68 return scalar gamma180se=_se[_rcs1]
69 return scalar lnlambda180se=_se[_cons]
70 return scalar surv180_trt0=surv180_trt0[1]
71 return scalar surv180_trt1=surv180_trt1[1]
72 return scalar surv180_trt0_lci=surv180_trt0_lci[1]
73 return scalar surv180_trt0_uci=surv180_trt0_uci[1]
74 return scalar surv180_trt1_lci=surv180_trt1_lci[1]
75 return scalar surv180_trt1_uci=surv180_trt1_uci[1]
76 return scalar death180_trt0=1-surv180_trt0[1]
77 return scalar death180_trt0_lci=1-surv180_trt0_uci[1]
78 return scalar death180_trt0_uci=1-surv180_trt0_lci[1]
79
80 *Recruitment time 1 year
81 stset stime1 if rec<365, f(died=1)
82 stpm2 trt, df(1) scale(h) noorthog
83 gen timerec1=365 in 1
84 predict survrec1_trt0, surv timevar(timerec1) at(trt 0) ci
85 predict survrec1_trt1, surv timevar(timerec1) at(trt 1) ci
86 return scalar trtrecl=_b[trt]
87 return scalar hrrec1=exp(_b[trt])
88 return scalar gammarec1=_b[_rcs1]
89 return scalar lambdarec1=exp(_b[_cons])
90 return scalar trtreclse=_se[trt]
91 return scalar gammareclse=_se[_rcs1]
92 return scalar lnlambdaareclse=_se[_cons]
93 return scalar survrec1_trt0=survrec1_trt0[1]
94 return scalar survrec1_trt1=survrec1_trt1[1]
95 return scalar survrec1_trt0_lci=survrec1_trt0_lci[1]
96 return scalar survrec1_trt0_uci=survrec1_trt0_uci[1]
97 return scalar survrec1_trt1_lci=survrec1_trt1_lci[1]
98 return scalar survrec1_trt1_uci=survrec1_trt1_uci[1]
99 return scalar deathrec1_trt0=1-survrec1_trt0[1]
100 return scalar deathrec1_trt0_lci=1-survrec1_trt0_uci[1]
101 return scalar deathrec1_trt0_uci=1-survrec1_trt0_lci[1]
102
103 local simlist
104 local simlisttot
105 foreach sec in 30 180 365 rec1 {
106 local simlist trt`sec'=r(trt`sec') hr`sec'=r(hr`sec')
trt`sec'se=r(trt`sec'se) ///
107 lambda`sec'=r(lambda`sec') gamma`sec'=r(gamma`sec') ///
108 lnlambda`sec'se=r(lnlambda`sec'se)
gamma`sec'se=r(gamma`sec'se) ///

```

```

109 death`sec'_trt0=r(death`sec'_trt0)
death`sec'_trt0_lci=r(death`sec'
_trt0_lci) ///
110 death`sec'_trt0_uci=r(death`sec'_trt0_uci)
111 local simlisttot `simlisttot' `simlist'
112 }
113
114 simulate `simlisttot', reps(1000): simplato
115 su *
116
117 foreach var in lambda365 lambda30 lambda180 lambdarec1 {
118 su `var'
119 local m`var'=r(mean)
120 local v`var'=r(Var)
121 local reallambda=0.0046
122 gen pb`var'=((`m`var'`'-`reallambda')/`reallambda')*100
123 gen mse`var'=((`m`var'`'-`reallambda')^2)+ `v`var''
124 gen in`var'=inrange(ln(`reallambda'), ln(`var')-
1.96*ln`var'se, ln(`var')+1.96*ln`var'
se)
125 }
126
127 foreach var in gamma365 gamma30 gamma180 gammarec1 {
128 su `var'
129 local m`var'=r(mean)
130 local v`var'=r(Var)
131 local realgamma=0.4360
132 gen pb`var'=((`m`var'`'-`realgamma')/`realgamma')*100
133 gen mse`var'=((`m`var'`'-`realgamma')^2)+ `v`var''
134 gen in`var'=inrange(`realgamma', (`var')-1.96*`var'se,
(`var')+1.96*`var'se)
135 }
136
137 foreach var in trt365 trt30 trt180 trtrecl {
138 su `var'
139 local m`var'=r(mean)
140 local v`var'=r(Var)
141 local realtrt=-0.243
142 gen pb`var'=((`m`var'`'-`realtrt')/`realtrt')*100
143 gen mse`var'=((`m`var'`'-`realtrt')^2)+ `v`var''
144 gen in`var'=inrange(`realtrt', `var'-1.96*`var'se,
`var'+1.96*`var'se)
145 }
146
147 su pb*
148 su mse*
149 tab inlambda365
150 tab ingamma365
151 tab intrt365
152 tab inlambda30
153 tab ingamma30
154 tab intrt30
155 tab inlambda180
156 tab ingamma180
157 tab intrt180
158 tab inlambdarec1
159 tab ingammarec1
160 tab intrtrecl
161 save simplato_death1000reps, replace
162 end

```

E3 Stata ado-file—‘simplatoinv’—for all-cause death, subgroup analysis (invasive population)

```
simulation (deathINV 0.6) for chapter.do - Printed on 22/10/2015

1 capture program drop simplatoinv
2 program define simplatoinv, rclass
3 *death, parameters from Weibull fit for baseline and treatment
  effect from Cannon et al. 2010
4 syntax [, LAMbdas(real 0.0048822993962951) GAMmas(real
  0.4093665719972181) TRT(real -
  0.210721031315653) OBS(integer 13500) ]
5 clear
6 cd "\\uol.le.ac.uk\root\staff\home\p\pd135\My
  Documents\Projects\Ticagrelor\Simulation"
7 set obs `obs'
8 gen recruit = 1+int((669-1+1)*runiform())
9 gen trt = rbinomial(1, 0.5)
10 survsim stime1 died, lambdas(`lambdas') gammas(`gammas')
  cov(trt `trt') maxt(365)
11
12 stset stime1, f(died=1)
13 stpm2 trt, df(1) scale(h) noorthog
14 *INVASIVE subgroup
15 gen time365=365 in 1
16 predict surv365_trt0, surv timevar(time365) at(trt 0) ci
17 predict surv365_trt1, surv timevar(time365) at(trt 1) ci
18 return scalar trt365=_b[trt]
19 return scalar hr365=exp(_b[trt])
20 return scalar gamma365=_b[_rcs1]
21 return scalar lambda365=exp(_b[_cons])
22 return scalar trt365se=_se[trt]
23 return scalar gamma365se=_se[_rcs1]
24 return scalar lnlambda365se=_se[_cons]
25 return scalar surv365_trt0=surv365_trt0[1]
26 return scalar surv365_trt1=surv365_trt1[1]
27 return scalar surv365_trt0_lci=surv365_trt0_lci[1]
28 return scalar surv365_trt0_uci=surv365_trt0_uci[1]
29 return scalar surv365_trt1_lci=surv365_trt1_lci[1]
30 return scalar surv365_trt1_uci=surv365_trt1_uci[1]
31 return scalar death365_trt0=1-surv365_trt0[1]
32 return scalar death365_trt0_lci=1-surv365_trt0_uci[1]
33 return scalar death365_trt0_uci=1-surv365_trt0_lci[1]
34 *Indirect comparison with Prasugrel
35 merge 1:1 _n using "\\uol.le.ac.uk\root\staff\home\p\pd135\My
  Documents\Projects\Ticagrelor\Simulation\Data\indirect_death
  0.1.dta", nogen
36 replace lnhr =_b[trt] if _n==1
37 replace selnhr =_se[trt] if _n==1
38 replace hr =exp(_b[trt]) if _n==1
39 gen lnhrind365 = lnhr[1] - lnhr[2]
40 gen selnhrind365 = sqrt(selnhr[1]^2+selnhr[2]^2)
41 gen hrind365 = exp(lnhrind365)
42 return scalar lnhrind365=lnhrind365[1]
43 return scalar selnhrind365=selnhrind365[1]
44 return scalar hrind365=hrind365[1]
45
46 *30 days analysis
47 stset stime1, f(died=1) exit(time 1*(30))
```

```

48 stpm2 trt, df(1) scale(h) noorthog
49 gen time30=30 in 1
50 predict surv30_trt0, surv timevar(time30) at(trt 0) ci
51 predict surv30_trt1, surv timevar(time30) at(trt 1) ci
52 return scalar trt30=_b[trt]
53 return scalar hr30=exp(_b[trt])
54 return scalar gamma30=_b[_rcs1]
55 return scalar lambda30=exp(_b[_cons])
56 return scalar trt30se=_se[trt]
57 return scalar gamma30se=_se[_rcs1]
58 return scalar lnlambda30se=_se[_cons]
59 return scalar surv30_trt0=surv30_trt0[1]
60 return scalar surv30_trt1=surv30_trt1[1]
61 return scalar surv30_trt0_lci=surv30_trt0_lci[1]
62 return scalar surv30_trt0_uci=surv30_trt0_uci[1]
63 return scalar surv30_trt1_lci=surv30_trt1_lci[1]
64 return scalar surv30_trt1_uci=surv30_trt1_uci[1]
65 return scalar death30_trt0=1-surv30_trt0[1]
66 return scalar death30_trt0_lci=1-surv30_trt0_uci[1]
67 return scalar death30_trt0_uci=1-surv30_trt0_lci[1]
68 *Indirect comparison with Prasugrel
69 merge 1:1 _n using "\\uol.le.ac.uk\root\staff\home\p\pd135\My
Documents\Projects\Ticagrelor\Simulation\Data\indirect_death
0.1.dta", nogen
70 replace lnhr =_b[trt] if _n==1
71 replace selnhr =_se[trt] if _n==1
72 replace hr =exp(_b[trt]) if _n==1
73 gen lnhrind30 = lnhr[1] - lnhr[2]
74 gen selnhrind30 = sqrt(selnhr[1]^2+selnhr[2]^2)
75 gen hrind30 = exp(lnhrind30)
76 return scalar lnhrind30=lnhrind30[1]
77 return scalar selnhrind30=selnhrind30[1]
78 return scalar hrind30=hrind30[1]
79
80 *180 days analysis (i.e. 6 months)
81 stset stime1, f(died=1) exit(time 1*(180))
82 stpm2 trt, df(1) scale(h) noorthog
83 gen time180=180 in 1
84 predict surv180_trt0, surv timevar(time180) at(trt 0) ci
85 predict surv180_trt1, surv timevar(time180) at(trt 1) ci
86 return scalar trt180=_b[trt]
87 return scalar hr180=exp(_b[trt])
88 return scalar gamma180=_b[_rcs1]
89 return scalar lambda180=exp(_b[_cons])
90 return scalar trt180se=_se[trt]
91 return scalar gamma180se=_se[_rcs1]
92 return scalar lnlambda180se=_se[_cons]
93 return scalar surv180_trt0=surv180_trt0[1]
94 return scalar surv180_trt1=surv180_trt1[1]
95 return scalar surv180_trt0_lci=surv180_trt0_lci[1]
96 return scalar surv180_trt0_uci=surv180_trt0_uci[1]
97 return scalar surv180_trt1_lci=surv180_trt1_lci[1]
98 return scalar surv180_trt1_uci=surv180_trt1_uci[1]
99 return scalar death180_trt0=1-surv180_trt0[1]
100 return scalar death180_trt0_lci=1-surv180_trt0_uci[1]
101 return scalar death180_trt0_uci=1-surv180_trt0_lci[1]
102 *Indirect comparison with Prasugrel
103 merge 1:1 _n using "\\uol.le.ac.uk\root\staff\home\p\pd135\My
Documents\Projects\Ticagrelor\Simulation\Data\indirect_death
0.1.dta", nogen

```

```

104 replace lnhr =_b[trt] if _n==1
105 replace selnhr =_se[trt] if _n==1
106 replace hr =exp(_b[trt]) if _n==1
107 gen lnhrind180 = lnhr[1] - lnhr[2]
108 gen selnhrind180 = sqrt(selnhr[1]^2+selnhr[2]^2)
109 gen hrind180 = exp(lnhrind180)
110 return scalar lnhrind180=lnhrind180[1]
111 return scalar selnhrind180=selnhrind180[1]
112 return scalar hrind180=hrind180[1]
113
114 *Recruitment time 1 year
115 stset stime1 if rec<365, f(died=1)
116 stpm2 trt, df(1) scale(h) noorthog
117 gen timerecl=365 in 1
118 predict survrec1_trt0, surv timevar(timerecl) at(trt 0) ci
119 predict survrec1_trt1, surv timevar(timerecl) at(trt 1) ci
120 return scalar trtrecl=_b[trt]
121 return scalar hrrec1=exp(_b[trt])
122 return scalar gammarecl=_b[_rcs1]
123 return scalar lambdarecl=exp(_b[_cons])
124 return scalar trtreclse=_se[trt]
125 return scalar gammareclse=_se[_rcs1]
126 return scalar lnλdareclse=_se[_cons]
127 return scalar survrec1_trt0=survrec1_trt0[1]
128 return scalar survrec1_trt1=survrec1_trt1[1]
129 return scalar survrec1_trt0_lci=survrec1_trt0_lci[1]
130 return scalar survrec1_trt0_uci=survrec1_trt0_uci[1]
131 return scalar survrec1_trt1_lci=survrec1_trt1_lci[1]
132 return scalar survrec1_trt1_uci=survrec1_trt1_uci[1]
133 return scalar deathrec1_trt0=1-survrec1_trt0[1]
134 return scalar deathrec1_trt0_lci=1-survrec1_trt0_uci[1]
135 return scalar deathrec1_trt0_uci=1-survrec1_trt0_lci[1]
136 *Indirect comparison with Prasugrel
137 merge 1:1 _n using "\\uol.le.ac.uk\root\staff\home\p\pd135\My
Documents\Projects\Ticagrelor\Simulation\Data\indirect_death
0.1.dta", nogen
138 replace lnhr =_b[trt] if _n==1
139 replace selnhr =_se[trt] if _n==1
140 replace hr =exp(_b[trt]) if _n==1
141 gen lnhrindrec1 = lnhr[1] - lnhr[2]
142 gen selnhrindrec1 = sqrt(selnhr[1]^2+selnhr[2]^2)
143 gen hrindrec1 = exp(lnhrindrec1)
144 return scalar lnhrindrec1=lnhrindrec1[1]
145 return scalar selnhrindrec1=selnhrindrec1[1]
146 return scalar hrindrec1=hrindrec1[1]
147 end
148
149 local simlist
150 local simlisttot
151 foreach sec in 30 180 365 rec1 {
152 local simlist trt`sec'=r(trt`sec') hr`sec'=r(hr`sec')
trt`sec'se=r(trt`sec'se) ///
153 lambda`sec'=r(lambda`sec') gamma`sec'=r(gamma`sec') ///
154 lnλambda`sec'se=r(lnλambda`sec'se)
gamma`sec'se=r(gamma`sec'se) ///
155 death`sec'_trt0=r(death`sec'_trt0)
death`sec'_trt0_lci=r(death`sec'_
_trt0_lci) ///

```



```

156 death`sec'_trt0_uci=r(death`sec'_trt0_uci) ///
157 lnhrind`sec'=r(lnhrind`sec') hrind`sec'=r(hrind`sec')
selnhrind`sec'=r(
selnhrind`sec')
158 local simlistttot `simlistttot' `simlist'
159 }
160
161 simulate `simlistttot', reps(10): simplatoinv
162 su *
163
164 foreach var in lambda365 lambda30 lambda180 lambdarec1 {
165 su `var'
166 local m`var'=r(mean)
167 local v`var'=r(Var)
168 local reallambda=0.0048822993962951
169 gen pb`var'=((`m`var'-'`reallambda')/`reallambda')*100
170 gen mse`var'=((`m`var'-'`reallambda')^2)+ `v`var''
171 gen in`var'=inrange(ln(`reallambda'), ln(`var')-
1.96*ln`var'se, ln(`var')+1.96*ln`var'
se)
172 }
173
174 foreach var in gamma365 gamma30 gamma180 gammarec1 {
175 su `var'
176 local m`var'=r(mean)
177 local v`var'=r(Var)
178 local realgamma=0.4093665719972181
179 gen pb`var'=((`m`var'-'`realgamma')/`realgamma')*100
180 gen mse`var'=((`m`var'-'`realgamma')^2)+ `v`var''
181 gen in`var'=inrange(`realgamma', (`var')-1.96*`var'se,
(`var')+1.96*`var'se)
182 }
183
184 foreach var in trt365 trt30 trt180 trtrecl {
185 su `var'
186 local m`var'=r(mean)
187 local v`var'=r(Var)
188 local realtrt=-0.210721031315653
189 gen pb`var'=((`m`var'-'`realtrt')/`realtrt')*100
190 gen mse`var'=((`m`var'-'`realtrt')^2)+ `v`var''
191 gen in`var'=inrange(`realtrt', `var'-1.96*`var'se,
`var'+1.96*`var'se)
192 }
193
194 su pb*
195 su mse*
196 tab inlambda365
197 tab ingamma365
198 tab intrt365
199 tab inlambda30
200 tab ingamma30
201 tab intrt30
202 tab inlambda180
203 tab ingamma180
204 tab intrt180
205 tab inlambdarec1
206 tab ingammarec1
207 tab intrtrecl
208
209 save simplatoinvdeath_1000reps, replace

```

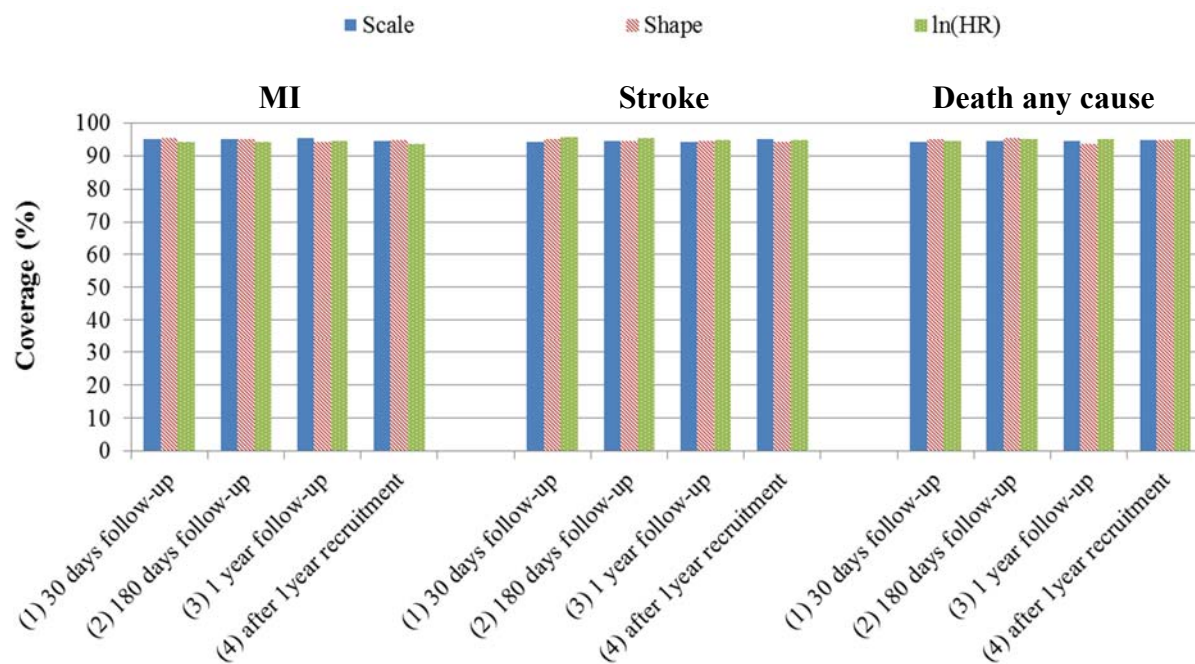
Table E1 Summary baseline transition probabilities for MI, stroke, and all-cause death for the base case simulation

	Baseline probabilities	Lower 95%CI	Upper 95% CI
MI			
(1) 30 days follow-up	0.021	0.019	0.025
(2) 180 days follow-up	0.044	0.040	0.048
(3) 1 year follow-up	0.058	0.053	0.063
(4) after 1yr recruitment	0.058	0.052	0.065
Stroke			
(1) 30 days follow-up	0.003	0.002	0.004
(2) 180 days follow-up	0.006	0.005	0.008
(3) 1 year follow-up	0.009	0.007	0.011
(4) after 1yr recruitment	0.009	0.006	0.012
Death any cause			
(1) 30 days follow-up	0.020	0.017	0.023
(2) 180 days follow-up	0.043	0.039	0.048
(3) 1 year follow-up	0.058	0.054	0.064
(4) after 1yr recruitment	0.058	0.052	0.065

Table E2 Summary percentage bias and MSE for the subgroup analysis simulation

Variable	MI		Stroke		Death any cause	
	% bias	MSE	% bias	MSE	% bias	MSE
Scale						
(1) 30 days follow-up	-0.25	0.00	-0.61	0.00	-1.45	0.00
(2) 180 days follow-up	0.04	0.00	-0.07	0.00	-0.72	0.00
(3) 1 year follow-up	-0.12	0.00	0.12	0.00	-0.68	0.00
(4) after 1year recruitment	-0.35	0.00	0.25	0.00	-0.44	0.00
Shape						
(1) 30 days follow-up	0.43	0.00	2.70	0.01	0.76	0.00
(2) 180 days follow-up	0.18	0.00	1.53	0.00	0.34	0.00
(3) 1 year follow-up	0.27	0.00	1.13	0.00	0.25	0.00
(4) after 1year recruitment	0.50	0.00	2.02	0.00	0.30	0.00
ln(HR)						
(1) 30 days follow-up	-0.93	0.01	-14.17	0.11	-5.04	0.02
(2) 180 days follow-up	0.28	0.01	-7.47	0.05	-1.15	0.01
(3) 1 year follow-up	-0.21	0.00	-5.70	0.03	-1.91	0.01
(4) after 1year recruitment	0.91	0.01	-6.77	0.07	-1.30	0.01

Figure E1 Coverage across simulation scenarios for shape, scale, and treatment effect parameters for the subgroup analysis simulation

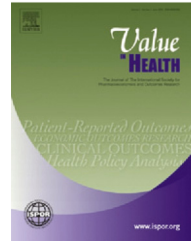


Appendix F

Dequen, P., A. J. Sutton, D. A. Scott and K. R. Abrams (2014). "Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery." Value Health **17**(4): 416-423

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Comparative Effectiveness Research/HTA

Searching for Indirect Evidence and Extending the Network of Studies for Network Meta-Analysis: Case Study in Venous Thromboembolic Events Prevention Following Elective Total Knee Replacement Surgery

Pascale Dequen, MSc^{1,*}, Alex J. Sutton, PhD¹, David A. Scott, MA, MSc², Keith R. Abrams, PhD¹¹Department of Health Sciences, University of Leicester, Leicester, UK; ²Oxford Outcomes Ltd., Oxford, UK

ABSTRACT

Objective: To evaluate the effect of study identification methods and network size on the relative effectiveness and cost-effectiveness of recommended pharmacological venous thromboembolic events (VTEs) prophylaxis for adult patients undergoing elective total knee replacement surgery in the United Kingdom. **Methods:** A stepwise literature search specifically designed to identify indirect evidence was conducted to extend the original clinical review from the latest National Institute for Health and Care Excellence (NICE) VTE technology appraisal. Different network sizes or network orders, based on the successive searches, informed three network meta-analyses (NMAs), which were compared with a replicated base case. The resulting comparative estimates were inputted in an economic model to investigate the effect of network size on cost-effectiveness probabilities. **Results:** Searches increased the number of indirect comparisons between VTE interventions, progressively widening the relevant network of studies for NMA. Precision around mean relative treatment

effects was increased as the network was extended from the base case to first-order NMA, but further extensions had limited effect. Cost-effectiveness analysis results were largely insensitive to variation in clinical inputs from the different NMA orders. **Conclusions:** No standard methodology is currently recommended by NICE to identify the most relevant network of studies for NMA. Our study showed that optimizing the identification of studies for NMA can extend the evidence base for analysis and reduce the uncertainty in relative effectiveness estimates. Although in our example network extensions did not affect the acceptability of available treatments in VTE prevention based on cost-effectiveness results, it may in other applications. **Keywords:** evidence synthesis, indirect treatment comparison, network meta-analysis, relative effectiveness, venous thromboembolism.

Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

The quantitative synthesis of clinical data is a key and often necessary step to the relative effectiveness assessment of medical interventions both premarket and postmarket launch. Meta-analysis is widely used to combine results from multiple clinical studies and considered best practice by many regulatory and health technology assessment bodies in Europe and worldwide [1]. The potential advantages, as well as standard methodology for conducting meta-analysis, are well established in the scientific community with acknowledged guidelines by the Cochrane Collaboration and the Centre for Reviews and Dissemination [2,3]. Recent statistical developments are extending this analytical approach to networks of studies, synthesizing evidence from both direct and indirect treatment comparisons [4–6].

When no head-to-head trial is available, studies evaluating A versus B and B versus C can be used to compare A and C indirectly using network meta-analysis (NMA). Indirect comparisons must be connected by at least one common comparator, that is, treatment B. Additional intermediate links may be required to connect two treatments of interest, thereby increasing the degree of “removal” or “separation” between comparisons and decreasing the degree of influence on the analysis [7]. A number of methodological concerns have been raised when extending an evidence base to include indirect comparisons within a network of studies such as how to best identify indirect evidence. The ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices published guidance on how to conduct NMA and recommended Hawkins et al.’s iterative search strategy to identify indirect evidence [7,8]. Although this search methodology can maximize the NMA

* Address correspondence to: Pascale Dequen, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK.

E-mail: pascale.dequen@le.ac.uk.

1098-3015/\$36.00 – see front matter Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2014.02.013>

network by efficiently identifying indirect evidence, authors warn that if more than a few links separate treatments (e.g., A and C), results may be unreliable. Additional links can provide useful information but may also increase between-study heterogeneity, uncertainty around estimates, and inconsistency between direct and indirect comparisons [7–9]. We carried out a case study to evaluate the effect of study identification methods and network size on indirect treatment comparisons for the prevention of venous thromboembolic events (VTEs) after total knee replacement (TKR) surgery.

The use of pharmacological, as well as mechanical, prophylaxis for VTE—deep vein thrombosis (DVT) and/or pulmonary embolism—after elective orthopaedic surgery is common practice in the United Kingdom. In 2010, the National Institute for Health and Care Excellence (NICE) published a clinical guideline on reducing the risk of VTE in patients admitted to hospital; at that time, five drugs were recommended: dabigatran etexilate, fondaparinux sodium, low molecular weight heparins, rivaroxaban, and unfractionated heparin for patients with renal failure [10]. Based on relative effectiveness estimates compared with these existing medicines, apixaban was also recommended in 2012 by NICE for use in adult patients scheduled for elective total hip or knee replacement [11]. These drugs were evaluated over time in single technology appraisals and all shown to be cost-effective for their given indication [11–13].

Objectives

We built on the latest NICE VTE technology appraisal TA245 for apixaban [11] to reanalyze the relative effectiveness and cost-effectiveness of recommended pharmacological VTE prophylaxis for adult patients undergoing elective TKR surgery in the United Kingdom using NMA. We sought to evaluate the effect of different network sizes on decision making for VTE prevention.

Methods

Literature Review

A stepwise systematic literature review was conducted in MEDLINE, Medline-in-Process, OLD Medline, EMBASE, and the Cochrane Library in October 2012 to identify relevant studies. The searches were replicated using the reported search strategies for the apixaban appraisal clinical review and adapted using

Hawkins et al.'s [7] breadth-first search methodology presented in Table 1 [11,14,15].

Breadth-first searching is based on graph theory; it is an uninformed or “naive” search process that aims to exhaustively search a sequence or a combination of sequences from a “root” node on a graph to all “neighboring” nodes without considering a final limit until it is reached. A parallel can be drawn between nodes on a graph to interventions on a network map and the need to identify all common comparators within a network without knowing the final size or shape of the network. Hawkins et al. [7] refer to search “orders” and associated search comparators to describe each sequential step in the breadth-first search. Treatments directly compared with first-order comparators following first-order searches become second-order comparators, and so on. The sequence of searches in Table 1 progressively include first-, second-, and third-order comparators, allowing us to identify all trials contributing to a network of evidence, until no further comparators are identified. From the set of identifiable trials, all relevant indirect comparisons are also identified at any given order.

In accordance with Hawkins et al. [7], searches were divided further for each order. In Table 1, search orders are numbered 1 to 3 and searches within each order i to vi. For example, in the first-order searches, all but one first-order comparator are included in the search terms (cf. search (1i) in Table 1). The omitted comparator is searched separately in a subsequent search iteration to ensure that all trials including one or more first-order comparators are captured and all possible second-order comparators identified (cf. search (1ii) in Table 1). Search (1i) will identify all trials comparing more than one of the first-order treatments, thus identifying any direct head-to-head evidence, albeit one of the treatments is not included in the search syntax. If the objective is to capture only first-order (i.e. direct) comparisons, the subsequent search (1ii) of the omitted comparator is not required. In this instance, dividing the search into two steps has the potential to reduce the search burden if a particular comparator is associated with a large number of hits. Hawkins et al. [7] thus recommend omitting a widely used comparator such as placebo or best supportive care; however, this is arbitrary. If further search orders are conducted and abstracts reviewed, search (1ii) is redundant and each order comparators could be searched at once. First-order comparators can be arbitrarily selected within or outside the original scope of searches and include treatments not of interest for appraisal. Moreover, study selection is intentionally broadened to include all clinical trials evaluating a first-order comparator without a restriction on comparator criteria, allowing for treatments that may not fall within the scope for appraisal, such as unlicensed drugs, nonrelevant treatments for decision making, or nonpharmacological interventions, to contribute to the network of evidence.

Studies were selected at the abstract and publication level on the basis of the indicated population for TKR and restricted to prospective, phases II to IV randomized controlled trials. To replicate the search conditions and provide comparable model results to the original technology appraisal, abstracts were further restricted by date to studies published before September 2011 and to English language. Date restrictions were included in the search strategy and exclusion of non-English abstracts and publications took place during the screening phase.

Network Meta-Analysis

Network sizes were based on the studies selected following each search order, thereafter referred to as first-, second-, and third-network orders. The base case was defined a priori in the apixaban appraisal from three pivotal phase III clinical trials

Table 1 – Breadth-first search strategy.

Search order	Search iteration	Search comparators
1	i	All first-order comparators except one
	ii	First-order comparator previously omitted
2	iii	All second-order comparators except one
	iv	Second-order comparator previously omitted
3	v	All third-order comparators except one
	vi	Third-order comparator previously omitted

Note. Adapted from Table 1 of Hawkins et al. [7].

comparing apixaban 2.5 mg/bd, dabigatran etexilate 220 mg/qd, and rivaroxaban 10 mg/qd to enoxaparin 40 mg/qd, respectively [16–18]. In accordance with the submitted apixaban economic model [14], these interventions form the decision space for VTE prevention after TKR and are routinely used in clinical practice in the United Kingdom. A comparison with fondaparinux was not considered relevant by manufacturers or the evidence review group because of its low market share in the United Kingdom and was therefore excluded from the analysis. The evidence network used in the original technology appraisal is referred to as the base case and shown in Figure 2A.

A Bayesian NMA was conducted for each network order for the composite outcome of total VTE and all-cause death, as well as for total DVT, and any bleeds. Multiple outcomes were analyzed for economic modeling purposes and to curb potential outcome reporting bias for the composite measure of all VTE/all-cause death used in more recent trials as primary outcome measure but not frequently calculated in older studies [16–18]. Fixed- and random-effects NMA models adjusted for multiarm trials were used in WinBUGS version 1.4.3 to estimate odds ratios (ORs), using ADES et al.'s codes available online [19,20].

The first 20,000 simulations were discarded as a burn-in and achieved reasonable convergence according to visual inspection of trace and history plots. Main analyses were based on a further 50,000 iterations to ensure robustness of results. Model fit was evaluated using the total residual deviance and the deviance information criterion (DIC) for each network size [21]. Between-study heterogeneity was compared using the standard deviation (SD) across random-effects models [22]. Inconsistency was assessed by plotting the residual deviances against the number of intervention arms in each included study, and looking at the proportion of mixed *P* values under 5% and 10% significance [23,24]. We expect that if there was no inconsistency, the residual deviance would equal the number of arms in each trial because it should be equal to 1 for each data point. Mixed *P* values provide an approximation to cross-validation *P* values, which can be calculated in a single model run. According to Welton et al. [25], mixed *P* values calculated from the same data set should follow a uniform distribution on the interval (0,1). We plotted the ordered *P* values for each study and each network order against uniform order statistics to evaluate inconsistency looking at unusually small or large *P* values [25].

Economic Model

A combined decision tree and Markov chain was built in Excel to model the initial prophylaxis/90-day postsurgery phase and the following 35-year time horizon, respectively. The economic model was rebuilt using the input data provided in the apixaban manufacturer submission and evidence review group report. The modeling approach and assumptions were externally validated against the original model [14,15]. Figure A in the Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013> illustrates the two-phase model diagram.

Treatment effect was demonstrated only during the first 90 days of the clinical pathway. We applied the ORs for all VTE/all-cause death and any bleeds from the NMA to adjust the baseline risk and inform transition probabilities in the decision tree. Baseline risks were taken from the Apixaban Dose Orally vs. Anticoagulation with Enoxaparin-2 trial for enoxaparin 40 mg/qd as in the original technology appraisal [16]. The parameterization of the Markov model was identical for all treatments compared. Uncertainty around parameters was expressed in distributions; a probabilistic sensitivity analysis was performed using 1000 model runs sampling from these distributions. ORs for all VTE/all-cause

death and any bleeds were sampled from 10,000 Markov chain Monte-Carlo simulations extracted from WinBUGS. Quality-adjusted life-years were used to estimate incremental cost-effectiveness ratios (ICERs) compared with enoxaparin; the 2.5th and 97.5th percentiles were also extracted to demonstrate the variation in uncertainty around mean ICER estimates at each given order.

Results

Literature Review

We considered the list of comparators included in the original apixaban submission search strategy as first-order comparators. More than 25 product names and drug classes of interest for VTE prevention in both total hip and knee replacement were included as first-order comparators. Different dosages were considered as individual treatments in the analysis. A full search strategy and the complete list of comparators included in each search order are included in the Appendix (cf. Table A1-3 and Table B) in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013>.

Fifty-three clinical trials met the inclusion criteria over the three network orders. Figure 1 shows the study selection flow diagram broken down by search and network order. The numbers of studies included and excluded for each search iteration are also presented and totaled by network order. Figure B in the Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013> illustrates the network map representing all treatment comparisons identified by successive search orders. The number of randomized controlled trials included in the NMA was limited to focus solely on treatment comparisons that would inform the relative effectiveness estimates for apixaban 2.5 mg/bd versus relevant comparators for decision making (i.e., dabigatran etexilate 220 mg/qd, enoxaparin 40 mg/qd, rivaroxaban 10 mg/qd). Graphically, these comparisons are referred to as “closed loops” within the network of studies. Focusing on these loops allowed us to reduce the size of the evidence base and make data sets more manageable without biasing results, because excluded studies did not contribute to indirect comparisons relevant to the decision space. Figure 2 illustrates the network diagrams for each search order including only the closed loops with the interventions of interest shaded in gray, as well as the base-case Indirect Treatment Comparison (ITC) network for reference. Asterisks in Figure 2 indicate that multiple drug dosages were represented by one node; although different dosages were considered as individual treatments in the analyses, these were not illustrated in the networks for readability. Note that we included interventions from three-arm trials even if only one treatment comparison from the trial was of interest, such as in Wang et al. [26] comparing placebo, fraxiparine (nadroparin calcium) 0.2 to 0.4 ml/qd, and indomethacin 25 mg/bd. Lastly, not all studies reported the outcomes of interest and were *de facto* excluded from the NMA. The final numbers of studies in each NMA order for TKR are included in Figure 1 and presented in tabular format in the Appendix (cf. Table C) in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013>, including studies reporting separate results for total hip and knee replacement in the same publication.

Network Meta-Analysis

Goodness-of-fit statistics for the fixed- and random-effects NMA models are presented in Table 2. Fixed-effects models for all network orders were used because they provided the best fit to the data according to the DIC. Forest plots in Figure 3 summarize

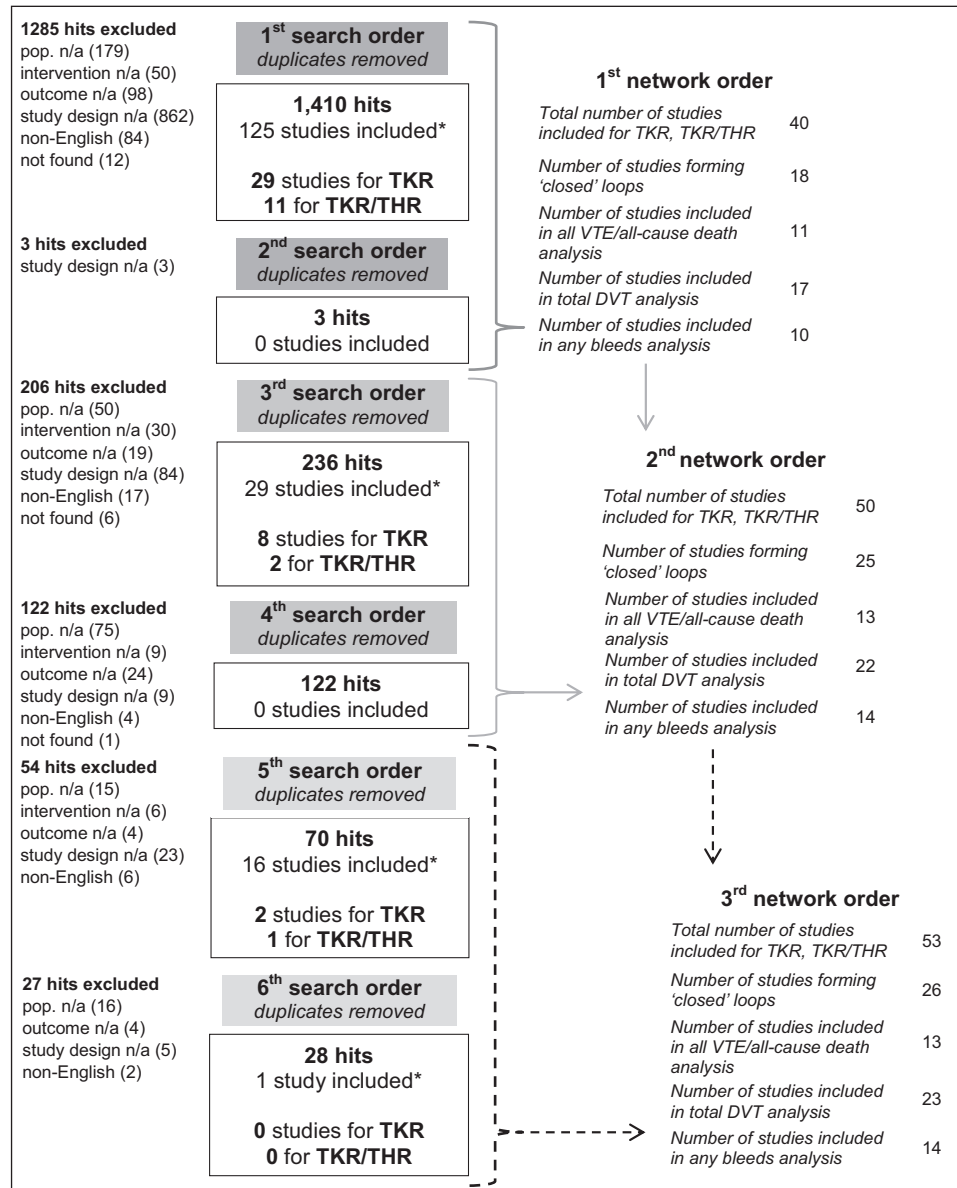


Fig. 1 – Study selection flow diagram. Asterisk indicates that the remainder of the included studies were THR only. DVT, deep vein thrombosis; n/a, not applicable; pop., population; THR, total hip replacement; TKR, total knee replacement; VTE, venous thromboembolic event.

the mean ORs and 95% credible intervals (CrI) for all VTE/all-cause death, total DVT, and any bleeds obtained for the base case and three network sizes. Given the number of studies included (cf. Fig. 1), second- and third-order NMAs for all VTE/all-cause death and any bleeds were the same and results in Figure 3 are presented only for completeness. The growing evidence base from the base case to first-network order marginally increased precision around the mean ORs for all outcomes. For example, the all VTE/all-cause death mean OR for dabigatran versus enoxaparin decreased from 0.95 (95% CrI 0.74–1.22) to 0.90 (0.73–1.10) between the base-case and first-order analysis; similarly, the uncertainty in any bleeds mean OR for apixaban versus enoxaparin was reduced from 0.78 (0.51–1.26) to 0.72 (0.55–0.97). Apixaban and rivaroxaban were superior to enoxaparin for both efficacy outcomes; however, ORs for dabigatran versus

enoxaparin were inconclusive. Results favored apixaban over dabigatran for all VTE/all-cause death for all network orders, with a mean OR of 0.65 (0.51–0.85) for first- and second-order analyses. The NMA also estimated that patients are less likely to experience a VTE event/death with rivaroxaban than with apixaban at higher network orders, although the base-case ITC did not support the statistical superiority of rivaroxaban and this was not demonstrated for total DVT. Apixaban showed the most favorable safety profile versus enoxaparin and versus rivaroxaban for first- and second-order NMA.

Although the fixed effects provided the best model fit for all outcomes and all network orders, we considered the random-effects models to assess between-study heterogeneity and the consistency of the evidence. Results for the random-effects models are included in the Appendix (cf. Figure C) in

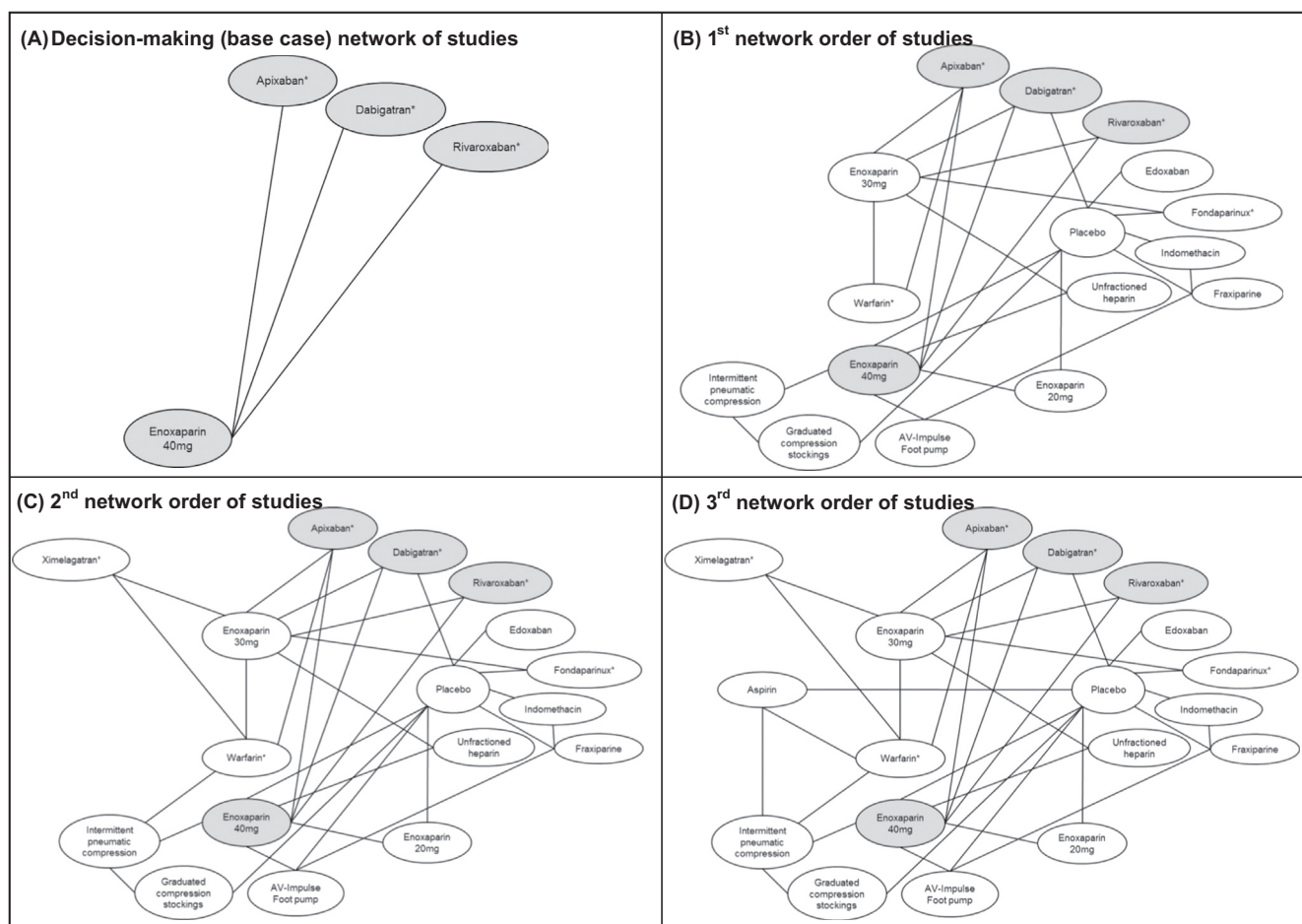


Fig. 2 – Network of studies including only “closed loops” based on search orders. Asterisk indicates multiple dosages included.

Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013> Overall, results were consistent across all network orders for both fixed- and random-effects models with little variation between respective point estimates and CrI. The between-study heterogeneity estimates and CrI were reduced

for all VTE/all-cause death from 0.156 (0.005–0.588) to 0.108 (0.004–0.379) and from 0.115 (0.003–0.569) to 0.108 (0.004–0.350) for any bleeds from first- to second-order NMA. The SDs increased, but not considerably, from 0.092 (0.002–0.307) to 0.112 (0.006–0.341) and

Table 2 – Goodness-of-fit statistics for fixed- and random-effects NMA models for all network orders.

Network order	Fixed effects		Random effects		
	DIC	Total residual deviance	DIC	Total residual deviance	SDs (95% CrI)
Total VTE/all-cause death					
First-order	260.97	39.92	262.27	39.33	0.156 (0.005– 0.588)
Second-order	303.14	44.23	304.45	43.95	0.108 (0.004– 0.379)
Third-order	NA	NA	NA	NA	NA
All DVT					
First order	366.15	52.48	369.14	52.96	0.092 (0.002– 0.307)
Second order	468.65	70.59	471.05	69.45	0.112 (0.006–0.341)
Third order	490.00	80.1	492.11	77.98	0.138 (0.015–0.391)
Any bleeds					
First order	237.87	33.46	239.46	34.12	0.115 (0.003–0.569)
Second order	303.89	42.29	305.36	42.86	0.108 (0.004–0.350)
Third order	NA	NA	NA	NA	NA

CrI, credible interval; DIC, deviance information criterion; DVT, deep vein thrombosis; NA, not available; NMA, network meta-analysis; VTE, venous thromboembolic event.

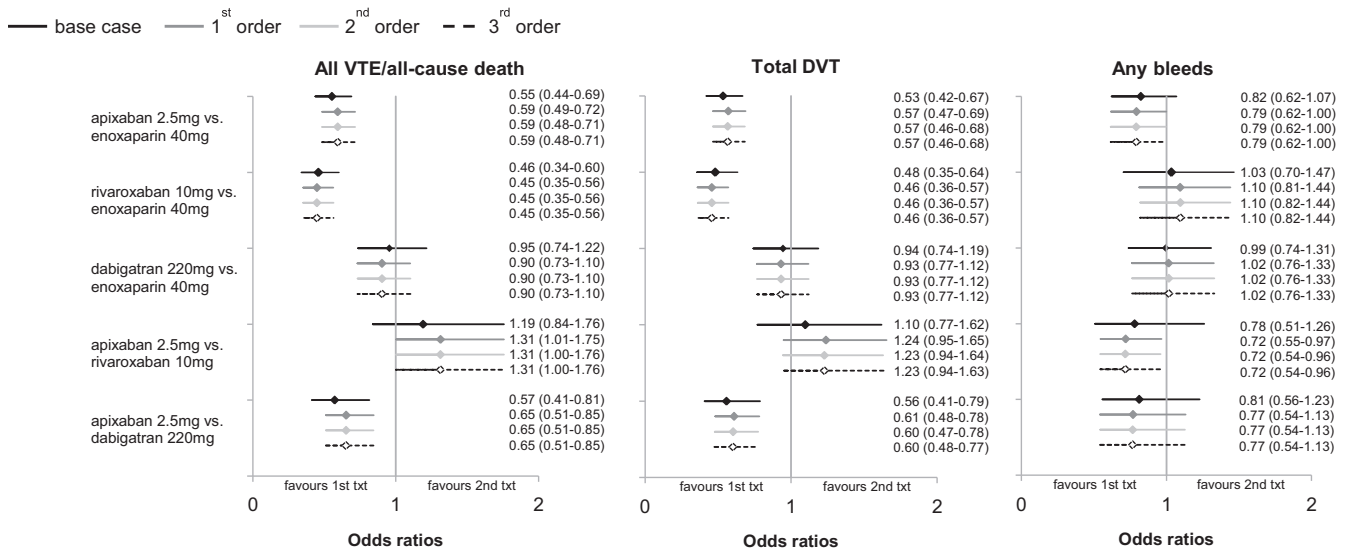


Fig. 3 – Odds ratios for all VTE/all-cause death, total DVT, and any bleeds from fixed-effects NMA models. DVT, deep vein thrombosis; NMA, network meta-analysis; VTE, venous thromboembolic event.

0.138 (0.015–0.391) as the network of studies grew across all three total DVT networks. Spiegelhalter et al. [22] provide a possible interpretation of the random-effects SD by describing a “range” of ORs. This range is in fact the ratio of the 97.5% to the 2.5% point of the distribution of ORs for any given relative treatment effect. They state that SDs on the OR scale of 0.1 or 0.2 will only ever correspond to a range of ORs of 1.48 or 2.19, respectively [22]. Therefore, the SDs reported in Table 2, all smaller than 0.2, showed little evidence of between-study heterogeneity.

Investigatory plots of residual deviances against the number of intervention arms for each trial, outcome, and network order, as shown in Figure D in the Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013>, do not suggest any inconsistencies between direct and indirect evidence across all models. We also plotted the ordered mixed predicted P values against uniform order statistics and found the evidence to be consistent across the three outcomes and network orders (cf. Figure E in the Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013>). Although the plotted mixed P values appear to deviate from a uniform distribution, no individual P value was significant at 5% or, more appropriately, at 10% due to the estimates being conservative by nature [25].

Lastly, analysis of both efficacy outcomes—that is, all VTE/all-cause death and total DVT—showed little variation largely due to the relatively low risks of pulmonary embolism (fatal and non-fatal) and death among surgical patients, suggesting no outcome reporting bias for composite measures in the VTE literature.

Economic Model

Apixaban, dabigatran etexilate, and rivaroxaban were found to be cost-effective versus enoxaparin for all network orders. These results were in line with findings from NICE appraisals that recommended these treatments on the basis of their dominance over enoxaparin. Table 3 presents the probabilistic sensitivity analysis means for total costs, total quality-adjusted life-years, and ICERs for the base case and first- and second-network orders. As previously stated, second- and third-order NMA results for all VTE/all-cause death and any bleeds were the same, because these were the clinical inputs to our model, and comparative effectiveness analysis results for the third-network order were redundant and not included in Table 3.

The mean ICERs for rivaroxaban, apixaban, and dabigatran etexilate were negative across all models, suggesting that treatments were on average both more effective and less costly than enoxaparin. The cost-utility analysis results showed little variation in outcomes despite the growing evidence base for the NMA parameterizing the economic model. Figure F in the Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.02.013> shows the cost-effectiveness planes based on the probabilistic sensitivity analysis results. At face value, these plots appear uninformative with regard to the effect of network size on the economic evaluation of compared pharmacological treatments for VTE. The percentages in Table 3, however, indicate a reduction in the uncertainty for which treatment is most cost-effective at a £20,000 willingness-to-pay threshold from the base case to first-network order, with rivaroxaban's predicted percentages increasing from 83.2% to 97.1% cost-effective. In addition, although the dominance of dabigatran versus enoxaparin is asserted by all network orders and the mean outcomes do not reflect any significant change, the 2.5th and 97.5th percentiles presented show the widest uncertainty in the ICERs.

Discussion

Using a breadth-first search strategy specifically designed to optimize the identification of indirect evidence allowed us to extend the network of relevant studies for analysis. Extensions of the network maximized the number of indirect comparisons between existing VTE interventions, and precision was increased from the base case to first-network order because additional studies reduced the uncertainty around mean ORs for all VTE/all-cause death, total DVT, and any bleeds. Estimates, however, became more stable as fewer studies were included in the evidence networks with each subsequent search order. Authors believe that additional information provided by trials comparing existing treatments to a lower dose of enoxaparin (30 mg/bd) identified in first-order searches contributed in large part to the increased precision across all outcome estimates. Overall, results from the NMA were consistent across network orders and extending the networks did not increase heterogeneity or inconsistency between studies. The cost-utility analysis was insensitive to NMA results; variation in the clinical input data according

Table 3 – Cost-effectiveness results of apixaban, dabigatranetexilate, and rivaroxaban vs. enoxaparin for the base case, first-order, and second-order networks.

Interventions	Total costs (£)	Total QALYs	ICERs (£)	2.5th percentile uncertainty (£)	97.5th percentile uncertainty (£)	% cost-effective at 20K	% cost-effective at 30K
Base case (ITC)							
Rivaroxaban	703	9.32	– 3,412	–4,171	–2,957	83.2	83.3
Apixaban	810	9.27	–3,703	–4,627	–3,109	16.8	16.7
Dabigatran etexilate	1,377	9.04	– 17,920	–76,636	75,111	0	0
Enoxaparin 40 mg	1,746	9.02				0	0
First-network order							
Rivaroxaban	688	9.34	–3,387	–4,044	–2,956	97.1	97.1
Apixaban	860	9.26	–3,851	–4,807	–3,225	2.9	2.9
Dabigatran etexilate	1,275	9.09	–7,907	–48,454	25,412	0	0
Enoxaparin 40 mg	1,748	9.03				0	0
Second-network order							
Rivaroxaban	695	9.33	–3,380	–4,043	–2,920	96.3	96.3
Apixaban	868	9.25	–3,841	–4,771	–3,181	3.7	3.7
Dabigatran etexilate	1,293	9.08	–8,197	–55,296	47,541	0	0
Enoxaparin 40 mg	1,754	9.02				0	0

Note. Third-network orders for included model inputs are the same as second-network orders so model results not presented above. ICER, incremental cost-effectiveness ratio; ITC, indirect treatment comparison; QALY, quality-adjusted life-year.

to network order did not affect mean ICERs but reduced the uncertainty in outcomes without influencing the acceptability of interventions.

A number of limitations and methodological challenges should be addressed. Authors did not find an in-depth exploration of heterogeneity and inconsistency (e.g., node-splitting) across order NMAs was warranted given the results and therefore it was not performed. The selection of first-order comparators was arbitrary because no clear definition of how to optimally choose these search terms currently exists. Hawkins et al. start the iterative searches in their practical example looking at all currently licensed treatments for non-small cell lung cancer across regulatory jurisdictions [9]; our first-order search considered indicated pharmaceutical interventions for VTE prophylaxis in the United Kingdom. Although the NICE scoping process can provide some grounds for defining first-order comparators, depending on the therapeutic area, these can include four interventions, that is, for second-line stage III/IV non-small cell lung cancer, or 30 in our case study. This should not make a difference but could affect how many search iterations are needed in the breadth-first strategy. In our case study, no particular gains were achieved from further dividing search orders because the additional burden of including all comparators, even placebo, rather than all but one comparator was marginal. Ultimately, all relevant comparators will be identified in the sequence of searches; however, the incremental value of higher search and network orders for NMA should be weighed against the associated additional search and computational burden. For example, the authors found that initially splitting each search order as recommended by Hawkins et al. to minimize the search burden, that is, searching for “all except one” comparators and subsequently searching the omitted comparator separately, proved inefficient. We agree with Hawkins et al. [7] that such omission is redundant if the next search order is conducted and abstracts reviewed, as was the case in our

example. In practice, searches conducted as part of a clinical evidence review could inform first-network order searches, even if distinct study selection criteria may be required, and this could help alleviate the search burden.

Efforts to widen an evidence base for analysis are highly dependent not only on the literature available but also what outcomes are reported in trial publications. Across all networks, between 3 and 13 studies were excluded from our analyses because they did not report outcomes of interest. Recent work in multiple outcomes analysis could help maximize the evidence base and improve NMA methods [27–29]. Moreover, König et al. [30] propose a new method to characterize the flow of evidence in an NMA using linear coefficients to interpret the “parallelism” and “indirectness” of networks to gauge the risk of bias, heterogeneity, and inconsistency within an indirect treatment comparison. Such methodological extensions to understand an evidence base, including how searching and identifying indirect evidence could be examined quantitatively to optimize network shape and size, are desirable.

Our application of Hawkins et al. [7] search methods to evaluate the cost-effectiveness of VTE prophylaxis suggests that more exhaustive searches to identify indirect evidence can provide valuable additional information for NMA. As we extend the breadth of searches, we can draw on more treatment comparisons to inform the network of studies for analysis. However, we are also more likely to include small sample size and older studies, which may contribute to greater between-study heterogeneity and increase the potential for time bias. Given the contradictory results found by Hawkins et al. in their similar study evaluating relative effectiveness estimates for non-small cell lung cancer treatments across multiple network sizes, the effect of extending the network size on uncertainty remains case-specific [9]. Taken together with our findings, however, this highlights the case for examining a wider network of evidence and in the absence of guidelines, we tentatively recommend Hawkins et al.’s search strategy to both future researchers and

reviewers. This awareness should prevent, or at least discourage, “gaming” when undertaking and reporting NMAs. To ensure transparency, health technology assessment bodies should consider wider networks for clinical review and evidence synthesis, as well as to justify the use of narrower networks for economic modeling and decision making. A simulation study to evaluate the effect of network sizes and shapes for NMA would provide generalizable findings and help formalize guidance on the added value of indirect searching and network extensions.

Source of financial support: This study was funded by Pfizer Ltd. via an unrestricted educational grant to the University of Leicester. K. Abrams is partially supported by the National Institute for Health Research (NIHR) (grant no. NF-SI-0508-10061) as a senior investigator. K. Abrams is a member of the National Institute for Health and Care Excellence (NICE) Technology Appraisal Committee and is also a member of the NICE Decision Support Unit (DSU). K. Abrams, A.J. Sutton, and D.A. Scott have provided methodological consultancy services to Pfizer Ltd. The publication of study results was not contingent on the sponsor’s approval or censorship of the manuscript.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at <http://dx.doi.org/10.1016/j.jval.2014.02.013> or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

- Kleijnen S, George E, Gouden S, et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 Jurisdictions. *Value Health* 2012;15:954–60.
- Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions* Version 5.1.0 [updated March 2011]. 2011. Available from: www.cochrane-handbook.org. [Accessed October 15, 2012].
- Centre for Reviews and Dissemination. *Systematic Reviews CRD’s Guidance for Undertaking Reviews in Health Care*. York: Centre for Reviews and Dissemination, University of York. Available from: www.york.ac.uk. [Accessed August 28, 2012].
- Lumley T. *Network meta-analysis for indirect treatment comparisons*. *Stat Med* 2002;21:2313–24.
- Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23:3105–24.
- Cooper NJ, Peters J, Lai MCW, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value Health* 2011;14:371–80.
- Hawkins N, Scott DA, Woods B. How far do you go? Efficient searching for indirect evidence. *Med Decis Mak* 2009;29:273–81.
- Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health* 2011;14:429–37.
- Hawkins N, Scott DA, Woods BS, Thatcher N. No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data. *Value Health* 2009;12:996–1003.
- National Institute for Health and Care Excellence. *Venous Thromboembolism—Reducing the Risk*. Care guidelines CG92. London, UK: National Institute for Health and Care Excellence, 2010. Available from: guidance.nice.org.uk/CG92. [Accessed September 1, 2012].
- National Institute for Health and Care Excellence. 2012. Apixaban for the Prevention of Venous Thromboembolism After Total Hip or Knee Replacement in Adults. Technology Appraisal TA245. London, UK: National Institute for Health and Care Excellence. Available from: www.nice.org.uk. [Accessed September 1, 2012].
- National Institute for Health and Care Excellence. Dabigatran Eteixilate for the Prevention of Venous Thromboembolism After Hip or Knee Replacement Surgery in Adults. Technology Appraisal TA157. London, UK: National Institute for Health and Care Excellence, 2008. Available from: www.nice.org.uk. [Accessed September 1, 2012].
- National Institute for Health and Care Excellence. Rivaroxaban for the Prevention of Venous Thromboembolism After Total Hip or Total Knee Replacement in Adults. Technology Appraisal TA170. London, UK: National Institute for Health and Care Excellence, 2009. Available from: www.nice.org.uk. [Accessed September 1, 2012].
- Bristol-Myers Squibb Pharmaceuticals Ltd. and Pfizer Ltd. *Apixaban (Eliquis®) for the Prevention of Venous Thromboembolic Events in Adult Patients Who Have Undergone Elective Hip or Knee Replacement Surgery*. Venous Thromboembolism—Apixaban (Hip and Knee Surgery): Manufacturer Submission. London, UK: National Institute for Health and Care Excellence, 2011.
- Kleijnen Systematic Reviews Ltd. *Apixaban for the Prevention of Venous Thromboembolism in People Undergoing Elective Knee and Hip Replacement Surgery*. Venous Thromboembolism—Apixaban (Hip and Knee Surgery): Evidence Review Group Report. London, UK: National Institute for Health and Care Excellence, 2011.
- Lassen MR, Raskob GE, Gallus A, et al. Apixaban versus enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): a randomised double-blind trial. *Lancet* 2012;375:807–15.
- Eriksson BI, Dahl OE, Rosencor N, et al. Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial. *J Thrombosis Haemostasis* 2007;5:2178–85.
- Lassen MR, Agno W, Borris LC, et al. (RECORD3 Investigators). Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. *N Engl J Med* 2008;358:2776–86.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;10:325–37.
- Ades AE, Welton N, Lu G. *Introduction to Mixed Treatment Comparisons*. Bristol, UK: University of Bristol. Available from: www.bristol.ac.uk. [Accessed October 23, 2012].
- Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. *J Royal Stat Soc Ser B (Stat Methodol)* 2002;64:583–616.
- Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, UK: John Wiley & Sons Ltd., 2003.
- Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J Clin Epidemiol* 2010;63:875–82.
- Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. *Stat Med* 2003;22:1649–60.
- Welton NJ, Sutton AJ, Copper NJ. *Evidence Synthesis for Decision Making in Healthcare*. Chichester, UK: John Wiley & Sons Ltd., 2012.
- Wang CJ, Wang JW, Weng LH, et al. Prevention of deep-vein thrombosis after total knee arthroplasty in Asian patients: comparison of low-molecular-weight heparin and indomethacin. *J Bone Joint Surg* 2004;86-A:136–40.
- Ades AE, Mavranzeouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. *Value Health* 2004;13:976–83.
- Welton NJ, Cooper NJ, Ades AE, et al. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Stat Med* 2008;27:5620–39.
- Lu G, Ades AE, Sutton AJ, et al. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med* 2007;26:3681–99.
- König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Stat Med* 2013;32:5414–29.

References

- Academy of Managed Care Pharmacy (01/2010) "AMCP Format for Formulary Submissions Version 3.0". *J Manag Care Pharm*. **16**(1 Suppl A):1-30.
- Achana, F. A., N. J. Cooper, S. Bujkiewicz, S, *et al.* (2014). "Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes." *BMC Med Res Methodol* **14**: 92.
- Achana, F. A., N. J. Cooper, S. Dias, *et al.* (2013). "Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis." *Stat Med* **32**(5): 752-771.
- Ades, A. E., D. M. Caldwell, S. Reken, *et al.* (2013). "Evidence synthesis for decision making 7: a reviewer's checklist." *Med Decis Making* **33**(5): 679-691.
- Ades, A. E., I. Mavranetzouli, S. Dias, *et al.* (2010). "Network meta-analysis with competing risk outcomes." *Value Health* **13**(8): 976-983.
- Ades, A. E., M. Sculpher, A. Sutton, *et al.* (2006). "Bayesian methods for evidence synthesis in cost-effectiveness analysis." *Pharmacoeconomics* **24**(1): 1-19.
- Agency for Health Technology Assessment (AOTM/AHTAPol) (2009) Guidelines for conducting Health Technology Assessment (HTA) Version 2.1. Warsaw, Poland. www.aotm.gov.pl
- Allen, L. A., C. J. O'Donnell, C. A. Camargo, *et al.* (2006). "Comparison of long-term mortality across the spectrum of acute coronary syndromes." *Am Heart J* **151**(5): 1065-1071.
- AOK Niedersachsen (2016) Health Insurance in Germany. Accessed 11/2015 <http://en.zuwanderer.aok.de/home/staying-healthy-in-germany/health-care-for-you-in-germany/health-insurance-in-germany/>
- Ara, R. and J. E. Brazier (2010). "Populating an economic model with health state utility values: moving toward better practice." *Value Health* **13**(5): 509-518.
- Arnardottir, A. H., F. M. Haaijer-Ruskamp, S. M. Straus, *et al.* (2011). "Additional safety risk to exceptionally approved drugs in Europe?" *Br J Clin Pharmacol* **72**(3): 490-499.
- Assmann, S. F., S. J. Pocock, L. E. Enos and L. E. Kasten (2000). "Subgroup analysis and other (mis)uses of baseline data in clinical trials." *Lancet* **355**(9209): 1064-1069.

AstraZeneca UK Ltd. (12/11/2010a) Acute coronary syndromes - ticagrelor: manufacturer submission

<https://www.nice.org.uk/guidance/TA236/documents/acute-coronary-syndromes-ticagrelor-astrazeneca4>

AstraZeneca UK Ltd. (28/07/2010b) Core Presentation, Brilinta (Ticagrelor), for the July 28, 2010 Meeting of the Cardiovascular and Renal Drugs Advisory Committee. Accessed 08/2015.

<http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm221382.htm>

Avksentieva M. V, N. V. Antonova, G. P. Arutyunov , *et al.* (2010) Protocol on the procedure for clinical and economic evaluation of drug lists that are submitted for reimbursement coverage from public health care budget, decision-making criteria. Research Centre for Clinical and Economic Evaluation and Pharmacoeconomics of the Russian State Medical University named after N.I. Pirogov (Educational Institution of Higher Professional Training RSMU of Federal Agency for Public Health and Social Development), Moscow, Russia.

Bagust A, A. Boland, M. Blundell, *et al.* (4/01/2011) Ticagrelor for the treatment of acute coronary syndromes. Liverpool Reviews and Implementation Group (LRiG), Ticagrelor for the treatment of ACS, ERG Report.

<https://www.nice.org.uk/guidance/TA236/documents/acute-coronary-syndromes-ticagrelor-evidence-review-group-report2>

Baird, L. G., M. R. Trusheim, H.-G. Eichler, *et al.* (2013). "Comparison of Stakeholder Metrics for Traditional and Adaptive Development and Licensing Approaches to Drug Development." *Therapeutic Innovation & Regulatory Science* **47**(4): 474-483.

Baird, L. G., R. Banken, H. G. Eichler, *et al.* (2014). "Accelerated access to innovative medicines for patients in need." *Clin Pharmacol Ther* **96**(5): 559-571.

Bartolucci, A. A. and W. B. Hillegass (2010). Overview, strengths, and limitations of systematic reviews and meta-analyses. *Evidence-Based Practice: Toward Optimizing Clinical Outcomes*, Springer: 17-33.

Barton, P., S. Bryan and S. Robinson (2004). "Modelling in the economic evaluation of health care: selecting the appropriate approach." *J Health Serv Res Policy* **9**.

Bassand, J. P., C. W. Hamm, D. Ardissino, *et al.* (2007). "Guidelines for the diagnosis and treatment of non-ST-segment elevation acute coronary syndromes." *Eur Heart J* **28**(13): 1598-1660.

- Bauer, K. A., B. I. Eriksson, M. R. Lassen and A. G. Turpie (2001). "Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after elective major knee surgery." N Engl J Med **345**(18): 1305-1310.
- Berntgen, M, A. Gourvil, M. Pavlovic, *et al.* (2014) "Improving the contribution of regulatory assessment reports to health technology assessments--a collaboration between the European Medicines Agency and the European network for Health Technology Assessment." Value Health **17**(5):634-41
- Biondi-Zoccai, G., M. Lotrionte, P. Agostoni, *et al.* (2011). "Adjusted indirect comparison meta-analysis of prasugrel versus ticagrelor for patients with acute coronary syndromes." Int J Cardiol **150**(3): 325-331.
- Blanchard, J., J. Y. Meuwly, P. F. Leyvraz, *et al.* (1999). "Prevention of deep-vein thrombosis after total knee replacement. Randomised comparison between a low-molecular-weight heparin (nadroparin) and mechanical prophylaxis with a foot-pump system." J Bone Joint Surg Br **81**(4): 654-659.
- Blasco, J., A. Martinez-Ferrer, J. Macho, *et al.* (2012). "Effect of vertebroplasty on pain relief, quality of life, and the incidence of new vertebral fractures: a 12-month randomized follow-up, controlled trial." J Bone Miner Res **27**(5): 1159-1166.
- Boonen, S., J. Van Meirhaeghe, L. Bastian, *et al.* (2011). "Balloon kyphoplasty for the treatment of acute vertebral compression fractures: 2-year results from a randomized trial." Journal of Bone and Mineral Research **26**(7): 1627-1637.
- Bormann I. (2012) DigitizeIt - Digitizer software - digitize a scanned graph or chart into (x,y)-data. <http://www.digitizeit.de/index.html>
- Briggs, A., K. Ritchie, E. Fenwick, *et al.* (2010). "Access with evidence development in the UK: past experience current initiatives and future potential." Pharmacoeconomics **28**(2):163–70.
- Briggs, A, K. Claxton, M. Sculpher. (2006) Decision modelling for health economic evaluation. Oxford University Press Inc., New York, USA.
- Briggs, A. H. (2000). "Handling uncertainty in cost-effectiveness models." Pharmacoeconomics **17**(5): 479-500.
- Briggs, A. and M. Sculpher (1998). "An introduction to Markov modelling for economic evaluation." Pharmacoeconomics **13**(4): 397-409.
- Bristol-Myers Squibb Pharmaceuticals Ltd. and Pfizer Ltd.(19/07/2011) Apixaban (Eliquis®) for the prevention of venous thromboembolic events in adult patients who have undergone elective hip or knee replacement surgery.

<https://www.nice.org.uk/guidance/TA245/documents/venous-thromboembolism-apixaban-hip-and-knee-surgery-bristolmyers-squibb-and-pfizer2>

- Buchbinder, R., R. H. Osborne, P. R. Ebeling, *et al.* (2009). "A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures." *N Engl J Med* **361**(6): 557-568.
- Bucher, H. C., G. H. Guyatt, L. E. Griffith and S. D. Walter (1997). "The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials." *J Clin Epidemiol* **50**(6): 683-691.
- Burge, R. T., D. Worley, A. Johansen, S. Bhattacharyya and U. Bose (2001). "The cost of osteoporotic fractures in the UK: projections for 2000–2020." *Journal of Medical Economics* **4**(1-4): 51-62.
- Burnham K. P and D. R. Anderson. (1998) Model selection and inference: A practical Information-Theoretical Approach. Springer-Verlag, New York, USA.
- Burton, A., D. G. Altman, P. Royston and R. L. Holder (2006). "The design of simulation studies in medical statistics." *Statistics in Medicine* **25**(24): 4279-4292.
- Caldwell, D. M., S. Dias and N. J. Welton (2015). "Extending Treatment Networks in Health Technology Assessment: How Far Should We Go?" *Value Health* **18**(5): 673-681.
- Caldwell, D. M. (2014). "An overview of conducting systematic reviews with network meta-analysis." *Systematic Reviews* **3**.
- Caldwell, D. M., N. J. Welton and A. E. Ades (2010). "Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency." *J Clin Epidemiol* **63**(8): 875-882.
- Canadian Agency for Drugs and Technologies in Health (16/12/2011) Common drug review, CDEC Final Recommendation, Ticagrelor (Brilinta – AstraZeneca) Indication: Prevention of Thrombotic Events in Acute Coronary Syndromes. https://www.cadth.ca/media/cdr/complete/cdr_complete_Brilinta_Dec-20-11.pdf
- Canadian Agency for Drugs and Technologies in Health (2006) HTA Guidelines for the economic evaluation of health technologies: Canada. 3rd edition. www.cadth.ca
- Canadian Agency for Drugs and Technologies in Health (2009) Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. www.cadth.ca
- Canadian Agency for Drugs and Technologies in Health (2009) USER GUIDE - Indirect Treatment Comparison. www.cadth.ca

- Canadian Agency for Drugs and Technologies in Health (2001, revised 2003). Guidelines for Authors of CADTH Health Technology Assessment Reports. www.cadth.ca
- Cannon, C. P., R. A. Harrington, S. James, *et al.* (2010). "Comparison of ticagrelor with clopidogrel in patients with a planned invasive strategy for acute coronary syndromes (PLATO): a randomised double-blind study." Lancet **375**(9711): 283-293.
- Cannon, C. P., S. Husted, R. A. Harrington, *et al.* (2007). "Safety, tolerability, and initial efficacy of AZD6140, the first reversible oral adenosine diphosphate receptor antagonist, compared with clopidogrel, in patients with non-ST-segment elevation acute coronary syndrome: primary results of the DISPERSE-2 trial." J Am Coll Cardiol **50**(19): 1844-1851.
- Carlson, J. J., S. D. Sullivan, L. P. Garrison, *et al.* (2009) "PHP64 Linking payment and health outcomes: a systematic review and taxonomy of performance-based health outcomes agreements between health care payers and manufacturers." Value Health **12**(3): A90.
- Caro, J. J. (2005). "Pharmacoeconomic analyses using discrete event simulation." Pharmacoeconomics **23**(4): 323-332.
- Carroll, K., H. C. Ross, D. Evans, *et al.* (2008). "Conditional approval: discussion points from the PSI conditional approval expert group." Pharm Stat **7**(4): 263-269; discussion 270-261.
- Cauley, J., D. Thompson, K. Ensrud, *et al.* (2000). "Risk of mortality following clinical fractures." Osteoporosis international **11**(7): 556-561.
- Centre for Reviews and Dissemination. (2009) Systematic Reviews CRD's guidance for undertaking reviews in health care. CRD, University of York, York, UK.
- Chaikledkaew, U. and K. Kittrongsiri (2014). "Guidelines for health technology assessment in Thailand (second edition)--the development process." J Med Assoc Thai **97 Suppl 5**: S4-9.
- Chalkidou, K., A. Hoy, P. Littlejohns. (2007) "Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research." J R Soc Med **100**(10):453–60.
- Chen, A. T., D. B. Cohen and R. L. Skolasky (2013). "Impact of nonoperative treatment, vertebroplasty, and kyphoplasty on survival and morbidity after vertebral compression fracture in the medicare population." J Bone Joint Surg Am **95**(19): 1729-1736.
- Chen, C. and L. Z. Sun (2011). "Quantification of PFS Effect for Accelerated Approval of Oncology Drugs." Statistics in Biopharmaceutical Research **3**(3): 434-444.

- Chen, M. H. and J. G. Ibrahim (2000). "Bayesian predictive inference for time series count data." *Biometrics* **56**(3): 678-685.
- Chin, P. L., M. S. Amin, K. Y. Yang, S. J. Yeo and N. N. Lo (2009). "Thromboembolic prophylaxis for total knee arthroplasty in Asian patients: a randomised controlled trial." *J Orthop Surg (Hong Kong)* **17**(1): 1-5.
- Claxton, K., S. Martin, M. Soares, *et al.* (10/06/2013) "Methods for the Estimation of the NICE Cost Effectiveness Threshold". Revised Report Following Referees Comments. Centre for Health Economic, University of York, York, UK.
https://www.york.ac.uk/media/che/documents/reports/resubmitted_report.pdf
- Claxton, K., S. Palmer, L. Longworth, *et al.* (2012) "Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development." *Health Technol Assess* **16**(46)
- Claxton, K., S. Palmer, L. Longworth, *et al.* (2011) Uncertainty, evidence and irrecoverable costs: informing approval, pricing and research decisions for health technologies. CHE research paper no. 69. Centre for Health Economics, University of York, York, UK.
- Cleemput, I., P. Van Wilder, R. Vrijens, *et al.* (2008) Guidelines for pharmacoeconomic evaluations in Belgium. Health Technology Assessment (HTA). Brussels: Health Care Knowledge Centre (KCE). KCE reports 78C.
https://kce.fgov.be/sites/default/files/page_documents/d20081027327.pdf
- Cochrane Comparing Multiple Interventions Group (2016) Accessed 10/2014.
<http://cmim.cochrane.org/>
- College voor zorgverzekeringen (23/05/2011) CFH-rapport 11/31: Ticagrelor (Brilique®). [https://www.zorginstituutnederland.nl/binaries/content/documents/zinl-www/documenten/publicaties/geneesmiddelbeoordelingen/2011/1105-ticagrelor-brilique/ticagrelor+\(Brilique\).pdf](https://www.zorginstituutnederland.nl/binaries/content/documents/zinl-www/documenten/publicaties/geneesmiddelbeoordelingen/2011/1105-ticagrelor-brilique/ticagrelor+(Brilique).pdf)
- College voor zorgverzekeringen (2006) Guidelines for pharmacoeconomic research, updated version.
<http://www.ispor.org/peguidelines/source/HTAGuidelinesNLUpdated2006.pdf>
- Colwell, C. W., Jr., S. D. Berkowitz, J. R. Lieberman, *et al.* (2005). "Oral direct thrombin inhibitor ximelagatran compared with warfarin for the prevention of venous thromboembolism after total knee arthroplasty." *J Bone Joint Surg Am* **87**(10): 2169-2177.

- Colwell, C. W., Jr., T. E. Spiro, A. A. Trowbridge, *et al.* (1995). "Efficacy and safety of enoxaparin versus unfractionated heparin for prevention of deep venous thrombosis after elective knee arthroplasty. Enoxaparin Clinical Trial Group." Clin Orthop Relat Res(321): 19-27.
- Cooper, N. J., A. J. Sutton, D. Morris, *et al.* (2009). "Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation." Stat Med **28**(14): 1861-1881.
- Cooper, N. J., J. Peters, M. C. Lai, *et al.* (2011). "How valuable are multiple treatment comparison methods in evidence-based health-care evaluation?" Value Health **14**(2): 371-380.
- Cortazar, P., R. Justice, J. Johnson, *et al.* (2012). "US Food and Drug Administration Approval Overview in Metastatic Breast Cancer." Journal of Clinical Oncology **30**(14): 1705-1711.
- Crombie, I. K. and H. TO. Davies. (4/2009) "What is meta-analysis?" Evidence-based medicine. What is...? series, second edition. Hayward Medical Communications, a division of Hayward Group Ltd, Newmarket, UK. www.whatisseries.co.uk
- Crowther, M. J. and P. C. Lambert (2012). "Simulating complex survival data." Stata Journal **12**(4): 674-687.
- Crowther, M. J. and P. C. Lambert (2013). "Simulating biologically plausible complex survival data." Statistics in Medicine **32**(23): 4118-4134.
- Davis, C. and J. Abraham (2011). "Desperately seeking cancer drugs: explaining the emergence and outcomes of accelerated pharmaceutical regulation." Sociol Health Illn **33**(5): 731-747.
- de Jong, J. P., D. E. Grobbee, B. Flamion, *et al.* (2013). "Appropriate evidence for adaptive marketing authorization." Nat Rev Drug Discov **12**(9): 647-648.
- Deeks, J. J., J. P. T. Higgins, D. G. Altman (editors). (2011) Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org
- Deeks, J. J., J. Dinnes, R. D'Amico, *et al.* (2003). "Evaluating non-randomised intervention studies." Health Technol Assess **7**(27): iii-x, 1-173.
- Denaro, V., U. G. Longo, N. Maffulli and L. Denaro (2009). "Vertebroplasty and kyphoplasty." Clinical Cases in Mineral and Bone Metabolism **6**(2): 125-130.

- Dennis, M. S., J. P. Burn, P. A. Sandercock, *et al.* (1993). "Long-term survival after first-ever stroke: the Oxfordshire Community Stroke Project." Stroke **24**(6): 796-800.
- Dequen, P., A. J. Sutton, D. A. Scott and K. R. Abrams (2014). "Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery." Value Health **17**(4): 416-423.
- Dias, S., N. J. Welton, A. J. Sutton, A. E. Ades. (04/2011) NICE DSU Technical Support Document 1: Introduction to Evidence Synthesis for Decision Making. Decision Support Unit, ScHARR, University of Sheffield, Sheffield, UK. <http://www.nicedsu.org.uk/TSD1%20Introduction.final.08.05.12.pdf>
- Dias, S., A. J. Sutton, N. J. Welton and A. E. Ades (2013). "Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis." Med Decis Making **33**(5): 671-678.
- Dias, S., N. J. Welton, D. M. Caldwell and A. E. Ades (2010). "Checking consistency in mixed treatment comparison meta-analysis." Stat Med **29**(7-8): 932-944.
- Dias, S., N. J. Welton, V. C. C. Marinho, *et al.* (2010). "Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis." Journal of the Royal Statistical Society: Series A (Statistics in Society) **173**(3): 613-629.
- Donegan, S., P. Williamson, U. D'Alessandro and C. Tudur Smith (2013). "Assessing key assumptions of network meta-analysis: a review of methods." Research Synthesis Methods **4**(4): 291-323.
- Douw, K., H. Vondeling, J. Sorensen, *et al.* (2004). ""The future should not take us by surprise": preparation of an early warning system in Denmark." Int J Technol Assess Health Care **20**(3): 342-350.
- Draborg, E., D. Gyrd-Hansen, P. B. Poulsen and M. Horder (2005). "International comparison of the definition and the practical application of health technology assessment." Int J Technol Assess Health Care **21**(1): 89-95.
- Drummond, M. F., M. J. Sculpher, G. W. Torrance, B. J. O'Brien, and G. L. Stoddar. (2005b) *Methods for the Economic Evaluation of Health Care Programmes*. Third edition. Oxford University Press Inc., New York, USA.
- Drummond, M. F. and T. O. Jefferson (1996). "Guidelines for authors and peer reviewers of economic submissions to the BMJ." BMJ **313**(7052): 275-283.
- Drummond, M. F., J. S. Schwartz, B. Jonsson, *et al.* (2008). "Key principles for the

- improved conduct of health technology assessments for resource allocation decisions." Int J Technol Assess Health Care **24**(3): 244-258; discussion 362-248.
- Drummond, M. F., M. Sculpher, G. W. Torrance, B. J. O'Brien and G. L. Stoddart (2005). Methods for the Economic Evaluation of Health Care Programmes, Volume 3rd. Oxford, Oxford University Press.
- Eddy, D. M., V. Hasselblad, and R. Shachter (1992). Meta-Analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence. Har/Dskt edition. Academic Press, Boston, USA.
- Eddy, D. M., V. Hasselblad and R. Shachter (1990). "A Bayesian Method for Synthesizing Evidence: The Confidence Profile Method." International Journal of Technology Assessment in Health Care **6**(01): 31-55.
- Eddin, A. A., K. L. Ong, E. Lau and S. M. Kurtz (2011). "Mortality risk for operated and nonoperated vertebral fracture patients in the medicare population." J Bone Miner Res **26**(7): 1617-1626.
- Eddin, A. A., K. L. Ong, E. Lau and S. M. Kurtz (2015). "Morbidity and Mortality After Vertebral Fractures: Comparison of Vertebral Augmentation and Nonoperative Management in the Medicare Population." Spine (Phila Pa 1976) **40**(15): 1228-1241.
- Edlin, R., C. McCabe, C. Hulme, P. Hall, J. Wright. (2015) Cost Effectiveness Modelling for Health Technology Assessment, A Practical Course. ADIS, Springer International Publishing, Switzerland.
- Eichler, H. G., F. Pignatti, B. Flamion, *et al.* (2008). "Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma." Nat Rev Drug Discov **7**(10): 818-826.
- Eichler, H. G., K. Oye, L. G. Baird, *et al.* (2012). "Adaptive licensing: taking the next step in the evolution of drug approval." Clin Pharmacol Ther **91**(3): 426-437.
- Eichler, H. G., L. G. Baird, R. Barker, *et al.* (2015). "From adaptive licensing to adaptive pathways: Delivering a flexible life-span approach to bring new drugs to patients." Clinical Pharmacology & Therapeutics **97**(3): 234-246.
- El Uso de Promedios de Modelo Bayesiano para Mejorar la Representación de la Incertidumbre en Modelos Ecológicos." Conservation Biology **17**(6): 1579-1590.
- Eriksson, B. I., O. E. Dahl, N. Rosencher, *et al.* (2007). "Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial." J Thromb Haemost **5**(11): 2178-2185.

- European Commission (16/10/2015). “Adaptive licensing” or “adaptive pathways”: Deregulation under the guise of earlier access (Joint Briefing Paper). Accessed 02/2016.
http://ec.europa.eu/health/files/committee/stamp/2015-10_stamp3/3c_prescrire_position_paper.pdf
- European Medicines Agency (2016). Adaptive pathways. Accessed 02/2016.
http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp
- European Medicines Agency (15/12/2014). Adaptive pathways to patients: report on the initial experience of the pilot project. Accessed 02/2016.
http://www.ema.europa.eu/docs/en_GB/document_library/Report/2014/12/WC500179560.pdf
- European Medicines Agency (23/09/2010) Summary of opinion (initial authorisation) – Brilique (Ticagrelor). Committee for Medicinal Products for Human Use.
http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/001241/WC500097028.pdf
- European Medicines Agency (5/12/2006). Guideline on the Scientific Application and the Practical Arrangements Necessary to Implement Commission Regulation (EC) No 507/2006 on the Conditional Marketing Authorisation for Medicinal Products for Human Use Falling Within the Scope of Regulation (EC) No 726/2004. Accessed 02/2016.
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004908.pdf
- European Medicines Agency (15/07/1998) Annex 1 – Summary of product characteristics, Plavix – clopidogrel.
http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000174/WC500042189.pdf
- European network for health technology assessment (2016) EUnetHTA About us. Accessed 02/2016. <http://www.eunetha.eu/about-us>
- European network for health technology assessment (02/2013) COMPARATORS & COMPARISONS - Criteria for the choice of the most appropriate comparator(s) Summary of current policies and best practice recommendations.
https://5026.fedimbo.belgium.be/sites/5026.fedimbo.belgium.be/files/Choice_of_comparator.pdf

- Faria, R., M. Hernandez Alava, A. Manca, A. J. Wailoo. (05/2015) NICE DSU Technical Support Document 17: The Use of Observational Data to Inform Estimates of Treatment Effectiveness in Technology Appraisal: Methods for Comparative Individual Patient Data. Decision Support Unit. ScHARR, University of Sheffield, Sheffield, UK.
- <http://www.nicedsu.org.uk/TSD17%20-20DSU%20Observational%20data%20FINAL.pdf>
- Farrokhi, M. R., E. Alibai and Z. Maghami (2011). "Randomized controlled trial of percutaneous vertebroplasty versus optimal medical management for the relief of pain and disability in acute osteoporotic vertebral compression fractures." J Neurosurg Spine **14**(5): 561-569.
- Fauno, P., O. Suomalainen, V. Rehnberg, *et al.* (1994). "Prophylaxis for the prevention of venous thromboembolism after total knee arthroplasty. A comparison between unfractionated and low-molecular-weight heparin." J Bone Joint Surg Am **76**(12): 1814-1818.
- Fenwick, E., K. Claxton and M. Sculpher (2001). "Representing uncertainty: the role of cost-effectiveness acceptability curves." Health economics **10**.
- Ferrario, A. and P. Kanavos. (2013) Managed entry agreements for pharmaceuticals: the European experience. EMiNet, Brussels, Belgium.
- Fitzgerald, R. H., Jr., T. E. Spiro, A. A. Trowbridge, *et al.* (2001). "Prevention of venous thromboembolic disease following primary total knee arthroplasty. A randomized, multicenter, open-label, parallel-group comparison of enoxaparin and warfarin." J Bone Joint Surg Am **83-a**(6): 900-906.
- Francis, C. W. (2007). "Prophylaxis for Thromboembolism in Hospitalized Medical Patients." New England Journal of Medicine **356**(14): 1438-1444.
- Francis, C. W., B. L. Davidson, S. D. Berkowitz, *et al.* (2002). "Ximelagatran versus warfarin for the prevention of venous thromboembolism after total knee arthroplasty. A randomized, double-blind trial." Ann Intern Med **137**(8): 648-655.
- Francis, C. W., S. D. Berkowitz, P. C. Comp, *et al.* (2003). "Comparison of ximelagatran with warfarin for the prevention of venous thromboembolism after total knee replacement." N Engl J Med **349**(18): 1703-1712.
- Fuji, T., S. Fuijita, T. Ujihira and T. Sato (2010). "Dabigatran etexilate prevents venous thromboembolism after total knee arthroplasty in Japanese patients with a safety profile comparable to placebo." J Arthroplasty **25**(8): 1267-1274.

- Fuji, T., S. Fujita and T. Ochi (2008). "Fondaparinux prevents venous thromboembolism after joint replacement surgery in Japanese patients." Int Orthop **32**(4): 443-451.
- Fuji, T., T. Ochi, S. Niwa and S. Fujita (2008). "Prevention of postoperative venous thromboembolism in Japanese patients undergoing total hip or knee arthroplasty: two randomized, double-blind, placebo-controlled studies with three dosage regimens of enoxaparin." J Orthop Sci **13**(5): 442-451.
- Futterman, L. G. and L. Lemberg (2004). "A silent killer--often preventable." Am J Crit Care **13**(5): 431-436.
- Gangi, A., T. Sabharwal, F. G. Irani, X. Buy, J. P. Morales and A. Adam (2006). "Quality assurance guidelines for percutaneous vertebroplasty." Cardiovascular and interventional radiology **29**(2): 173-178.
- Garfin, S. R., H. A. Yuan and M. A. Reiley (2001). "New Technologies in Spine: Kyphoplasty and Vertebroplasty for the Treatment of Painful Osteoporotic Compression Fractures." Spine **26**(14): 1511-1515.
- Garg, P., D. J. Cohen, T. Gaziano and L. Mauri (2008). "Balancing the risks of restenosis and stent thrombosis in bare-metal versus drug-eluting stents: results of a decision analytic model." J Am Coll Cardiol **51**(19): 1844-1853.
- Geerts, W. H., D. Bergqvist, G. F. Pineo, *et al.* (2008). "Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition)." Chest **133**(6 Suppl): 381s-453s.
- Gelfand, A. E. and F. M. S. Adrian (1990). "Sampling-Based Approaches to Calculating Marginal Densities." Journal of the American Statistical Association **85**(410): 398-409.
- Gelman, A. and D. B. Rubin. (21/05/1996). Markov Chain Monte Carlo Methods in Biostatistics. Accessed 12/2015.
<http://www.stat.columbia.edu/~gelman/research/published/med4.pdf>
- Gelman, A. and D. B. Rubin (1992). "Inference from Iterative Simulation Using Multiple Sequences." Statistical Science **7**(4): 457-472.
- Ginsberg, J. S., B. L. Davidson, P. C. Comp, *et al.* (2009). "Oral thrombin inhibitor dabigatran etexilate vs North American enoxaparin regimen for prevention of venous thromboembolism after knee arthroplasty surgery." J Arthroplasty **24**(1): 1-9.
- Goodman, C. S. (2014) HTA 101: Introduction to Health Technology Assessment. Bethesda, MD: National Library of Medicine, USA.
- Goodman, C. S. and R. Ahn (1999). "Methodological approaches of health technology

- assessment." *Int J Med Inform* **56**(1-3): 97-105.
- Gray, A. M., P. M. Clarke, J. L. Wolstenholme, S. Wordsworth. (2001) *Applied Methods of Cost-effectiveness Analysis in Healthcare*. Oxford University Press, Oxford, UK.
- Griffin, S., L. Bojke, C. Main and S. Palmer (2006). "Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer." *Value Health* **9**(2): 123-131.
- Grimm, S., M. Strong, A. Brennan, and A. J. Wailoo. (2016) *Framework for analysing risk in Health Technology Assessments and its application to Managed Entry Agreements*. Decision Support Unit. ScHARR, University of Sheffield, Sheffield, UK. <http://www.nicedsu.org.uk/DSU%20Managed%20Access%20report%20FINAL.pdf>
- Guyatt, G. H., Oxman, A. D., and G. Vist *et al.*, for the GRADE Working Group (2008) "Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations." *BMJ* **336**(7650):924-6.
- Guyatt, G. H., Oxman, A. D., and E. A. Akl, *et al.* (2011). "GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables." *J Clin Epidemiol* **64**(4): 383-394.
- Haidich, A. B. (2010). "Meta-analysis in medical research." *Hippokratia* **14**(Suppl 1): 29-37.
- Hailey, D. (2003). "Toward transparency in health technology assessment: a checklist for HTA reports." *Int J Technol Assess Health Care* **19**(1): 1-7.
- Haute Autorité de santé (2010) *L'évaluation économique à la Haute Autorité de Santé Principes et méthodes*. 2010. www.has-sante.fr
- Haute Autorité de santé (2007) *General method for assessing health technologies*. Department of Medical and Surgical Procedures Assessment. www.has-sante.fr
- Haute Autorité de santé (2000) *Guide d'analyse de la littérature et gradation des recommandations*. www.has-sante.fr
- Hawkins, N., D. A. Scott and B. Woods (2009a). "How far do you go? Efficient searching for indirect evidence." *Med Decis Making* **29**(3): 273-281.
- Hawkins, N., D. A. Scott, B. S. Woods and N. Thatcher (2009b). "No study left behind: a network meta-analysis in non-small-cell lung cancer demonstrating the importance of considering all relevant data." *Value Health* **12**(6): 996-1003.
- Heit, J. A., C. W. Colwell, C. W. Francis, *et al.* (2001). "Comparison of the oral direct thrombin inhibitor ximelagatran with enoxaparin as prophylaxis against venous

- thromboembolism after total knee replacement: a phase 2 dose-finding study." Arch Intern Med **161**(18): 2215-2221.
- Henshall, C., L. Mardhani-Bayne, K. B. Frønsdal and M. Klemp (2011). "Interactions between health technology assessment, coverage, and regulatory processes: Emerging issues, goals, and opportunities." International Journal of Technology Assessment in Health Care **27**(03): 253-260.
- Higgins, J. P. T., S. Green (editors). (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. www.cochrane-handbook.org
- Higgins, J. P. T., S. G. Thompson and D. J. Spiegelhalter (2009). "A re-evaluation of random-effects meta-analysis." Journal of the Royal Statistical Society. Series A, (Statistics in Society) **172**(1): 137-159.
- Higgins, J. P., S. G. Thompson, J. J. Deeks and D. G. Altman (2003). "Measuring inconsistency in meta-analyses." Bmj **327**(7414): 557-560.
- Higgins, J., S. Thompson, J. Deeks and D. Altman (2002). "Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice." Journal of Health Services Research & Policy **7**(1): 51-61.
- High Level Pharmaceutical Forum (2008). Final Report 2005 – 2008. Accessed 01/2013. http://www.anm.ro/_/Final%20Conclusions%20and%20Recommendations%20of%20the%20High%20Level%20Pharmaceutical%20Forum.pdf
- Hoaglin, D. C., N. Hawkins, J. P. Jansen, *et al.* (2011). "Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2." Value Health **14**(4): 429-437.
- Hoekman, J., W. P. Boon, J. C. Bouvy, *et al.* (2015). "Use of the conditional marketing authorization pathway for oncology medicines in Europe." Clin Pharmacol Ther **98**(5): 534-541.
- Hofler, M., R. Lieb and H. U. Wittchen (2007). "Estimating causal effects from observational data with a model for multiple bias." Int J Methods Psychiatr Res **16**(2): 77-87.
- Holt, G., K. T. Khaw, D. M. Reid, *et al.* (2002). "Prevalence of osteoporotic bone mineral density at the hip in Britain differs substantially from the US over 50 years of age: implications for clinical densitometry." Br J Radiol **75**(897): 736-742.

- House of Commons Health Committee (8/03/2005) The Prevention of Venous Thromboembolism in Hospitalised Patients. Second Report of Session 2004-2005 (HC99). House of Commons, The Stationary Office Limited, London, UK. <http://www.publications.parliament.uk/pa/cm200405/cmselect/cmhealth/99/99.pdf>
- Husereau, D., C. Henshall and J. Jivraj (2014). "ADAPTIVE APPROACHES TO LICENSING, HEALTH TECHNOLOGY ASSESSMENT, AND INTRODUCTION OF DRUGS AND DEVICES." International Journal of Technology Assessment in Health Care **30**(03): 241-249.
- Hutton, J., P. Trueman and C. Henshall (2007). "Coverage with evidence development: an examination of conceptual and policy issues." Int J Technol Assess Health Care **23**(4): 425-432.
- Ibrahim, J., M. H. Chen, D. Sinha. (2003) On Optimality Properties of the Power Prior. Journal of the American Statistical Association. American Statistical Association. HighBeam Research. <https://www.highbeam.com/doc/1G1-100068883.html>
- Innovative Medicines Initiative (2016) IMI Home page. Accessed 02/2016. <http://www.imi.europa.eu/>
- Ioannidis, J. A., A. Haidich, M. Pappa, *et al.* (2001). "Comparison of evidence of treatment effects in randomized and nonrandomized studies." JAMA **286**(7): 821-830.
- Ioannidis, J. P. and J. Lau (1999). "Pooling research results: benefits and limitations of meta-analysis." Jt Comm J Qual Improv **25**(9): 462-469.
- Jalava, T., S. Sarna, L. Pylkkänen, *et al.* (2003). "Association between vertebral fracture and increased mortality in osteoporotic patients." Journal of Bone and Mineral Research **18**(7): 1254-1260.
- James, S., A. Akerblom, C. P. Cannon, *et al.* (2009). "Comparison of ticagrelor, the first reversible oral P2Y(12) receptor antagonist, with clopidogrel in patients with acute coronary syndromes: Rationale, design, and baseline characteristics of the PLATelet inhibition and patient Outcomes (PLATO) trial." Am Heart J **157**(4): 599-605.
- Jansen, J. P., R. Fleurence, B. Devine, *et al.* (2011). "Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1." Value Health **14**(1524-4733 (Electronic)): 417-428.
- Jansen, J. P., T. Trikalinos, J. C. Cappelleri, *et al.* (2014). "Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good

- Practice Task Force report." Value Health **17**(2): 157-173.
- Johnson and Johnson (8/02/2012) Vertebral fractures - vertebroplasty and kyphoplasty: manufacturer submissions (executive summary only)
<https://www.nice.org.uk/guidance/TA279/documents/vertebral-fractures-vertebroplasty-and-kyphoplasty-johnson-and-johnson4>
- Jönsson, B. (2015). "Bringing in health technology assessment and cost-effectiveness considerations at an early stage of drug development" Molecular Oncology **9**(5):1025-1033.
- Jönsson, B. (2013). "Technology assessment for new oncology drugs." Clin Cancer Res **19**(1): 6-11.
- Kaaniche, A., A. Troubat and A. Sherwood (2015). "Does Conditional Marketing Authorisation Influence Market Access in France, England, and Germany?" Value Health **18**(7): A559.
- Kado, D. M., W. S. Browner, L. Palermo, *et al.* (1999). "Vertebral fractures and mortality in older women: a prospective study." Archives of internal medicine **159**(11): 1215-1220.
- Kaempffe, F. A., R. M. Lifeso and C. Meinking (1991). "Intermittent pneumatic compression versus coumadin. Prevention of deep vein thrombosis in lower-extremity total joint arthroplasty." Clin Orthop Relat Res(269): 89-97.
- Kallmes, D. F., B. A. Comstock, P. J. Heagerty, *et al.* (2009). "A randomized trial of vertebroplasty for osteoporotic spinal fractures." N Engl J Med **361**(6): 569-579.
- Kesselheim, A. S., B. Wang, J. M. Franklin and J. J. Darrow (2015). "Trends in utilization of FDA expedited drug development and approval programs, 1987-2014: cohort study." BMJ **351**.
- Kind, P., P. Dolan, C. Gudex and A. Williams (1998). "Variations in population health status: results from a United Kingdom national questionnaire survey." BMJ : British Medical Journal **316**(7133): 736-741.
- Klazen, C. A., P. N. Lohle, J. de Vries, *et al.* (2010). "Vertebroplasty versus conservative treatment in acute osteoporotic vertebral compression fractures (Vertos II): an open-label randomised trial." Lancet **376**(9746): 1085-1092.
- Kleijnen, S., E. George, S. Goulden, *et al.* (2012). "Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions." Value Health **15**(6): 954-960.
- Klotzbuecher, C. M., P. D. Ross, P. B. Landsman, T. A. Abbott and M. Berger (2000).

- "Patients with Prior Fractures Have an Increased Risk of Future Fractures: A Summary of the Literature and Statistical Synthesis." Journal of Bone and Mineral Research **15**(4): 721-739.
- Konig, J., U. Krahn and H. Binder (2013). "Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons." Stat Med **32**(30): 5414-5429.
- Lacey, E. and S. Walters (2003). "Continuing inequality: gender and social class influences on self perceived health after a heart attack." Journal of Epidemiology and Community Health **57**(8): 622-627.
- Lam, P. K. (2008-2009). MCMC Methods: Gibbs Sampling and the Metropolis-Hastings Algorithm. Teaching Lecture, Bayesian Statistics in Political Science Research, Harvard University, Cambridge, Massachusetts, USA. Accessed 01/2016. <http://www.people.fas.harvard.edu/~plam/teaching/methods/mcmc/mcmc.pdf>
- Lambert, P. C. (15-17/01/2014) Computationally Intensive Methods (part 2) Lecture Slides for MSc Medical Statistics, University of Leicester, Leicester, UK.
- Lambert, P. C. and P. Royston (2009). "Further development of flexible parametric models for survival analysis." Stata Journal **9**(2): 265-290.
- Lange, A., C. Kasperk, L. Alvares, *et al.* (2014). "Survival and cost comparison of kyphoplasty and percutaneous vertebroplasty using German claims data." Spine (Phila Pa 1976) **39**(4): 318-326.
- Lassen, M. R., B. L. Davidson, A. Gallus, *et al.* (2007). "The efficacy and safety of apixaban, an oral, direct factor Xa inhibitor, as thromboprophylaxis in patients following total knee replacement." J Thromb Haemost **5**(12): 2368-2375.
- Lassen, M. R., G. E. Raskob, A. Gallus, *et al.* (2009). "Apixaban or enoxaparin for thromboprophylaxis after knee replacement." N Engl J Med **361**(6): 594-604.
- Lassen, M. R., G. E. Raskob, A. Gallus, *et al.* (2010). "Apixaban versus enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): a randomised double-blind trial." Lancet **375**(9717): 807-815.
- Lassen, M. R., W. Ageno, L. C. Borris, *et al.* (2008). "Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty." N Engl J Med **358**(26): 2776-2786.
- Latimer, N. R., K. R. Abrams, P. C. Lambert, *et al.* (2014). "Adjusting Survival Time Estimates to Account for Treatment Switching in Randomized Controlled Trials—an Economic Evaluation Context: Methods, Limitations, and Recommendations." Medical Decision Making **34**(3): 387-402.

- Latimer, N. R., K. R. Abrams, P. C. Lambert, *et al.* (03/2013) Adjusting survival time estimates to account for treatment switching in randomised controlled trials – a simulation study. Health Economics and Decision Science (HEDS) Discussion Paper No. 13.06. ScHARR, University of Sheffield, Sheffield, UK.
- Latimer, N. R. (06/2011) NICE DSU Technical Support Document 14 : Survival Analysis for Economic Evaluations Alongside Clinical Trials – Extrapolation with Patient-Level Data. Decision Support Unit, ScHARR, University of Sheffield, Sheffield, UK.
- Laws, A., R. Kendall and N. Hawkins (2014). "A Comparison of National Guidelines for Network Meta-Analysis." Value Health **17**(5): 642-654.
- Liberti, L., P. Stolk, N. McAuslane, *et al.* (2015). "Adaptive licensing and facilitated regulatory pathways: A survey of stakeholder perceptions." Clin Pharmacol Ther **98**(5): 477-479.
- Liu, J. T., W. J. Liao, W. C. Tan, *et al.* (2010). "Balloon kyphoplasty versus vertebroplasty for treatment of osteoporotic vertebral compression fracture: a prospective, comparative, and randomized clinical study." Osteoporos Int **21**(2): 359-364.
- Longo, U. G., M. Loppini, L. Denaro, *et al.* (2012). "Conservative management of patients with an osteoporotic vertebral fracture: A review of the literature." Journal of Bone & Joint Surgery, British Volume **94-B**(2): 152-157.
- Longworth, L., J. Youn, L. Bojke, *et al.* (2013) When does NICE recommend the use of health technologies within a programme of evidence development? : a systematic review of NICE guidance. Pharmacoeconomics **31**(2):137-49.
- Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." Stat Med **23**(20): 3105-3124.
- Lu, G., A. E. Ades, A. J. Sutton, *et al.* (2007). "Meta-analysis of mixed treatment comparisons at multiple follow-up times." Statistics in Medicine **26**(20): 3681-3699.
- Lumley, T. (2002). "Network meta-analysis for indirect treatment comparisons." Stat Med **21**(16): 2313-2324.
- Lunn D., C. Jackson, N. Best, *et al.* (2012) The BUGS Book - A Practical Introduction to Bayesian Analysis. CRC Press, Chapman and Hall, Boca Raton, Florida, USA.
- Lunn, D., A. Thomas, N. Best and D. Spiegelhalter (2000). "WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility." Statistics and Computing **10**(4): 325-337.
- MacLehose, R. R., B. C. Reeves, I. M. Harvey, *et al.* (2000). "A systematic review of

- comparisons of effect sizes derived from randomised and non-randomised studies." Health Technol Assess **4**(34): 1-154.
- Mahaffey, K. W., D. M. Wojdyla, K. Carroll, *et al.* (2011). "Ticagrelor compared with clopidogrel by geographic region in the Platelet Inhibition and Patient Outcomes (PLATO) trial." Circulation **124**(5): 544-554.
- Marshall, E. C. and D. J. Spiegelhalter (2003). "Approximate cross-validatory predictive checks in disease mapping models." Stat Med **22**(10): 1649-1660.
- McCullough, B. J., B. A. Comstock, R. A. Deyo, *et al.* (2013). "Major medical outcomes with spinal augmentation vs conservative therapy." JAMA Intern Med **173**(16): 1514-1521.
- McKenna, R., J. Galante, F. Bachmann, *et al.* (1980). "Prevention of venous thromboembolism after total knee replacement by high-dose aspirin or intermittent calf and thigh compression." Br Med J **280**(6213): 514-517.
- Medtronic (02/2012) Percutaneous Vertebroplasty (PVP) and Balloon Kyphoplasty (BKP) for the treatment of Osteoporotic Vertebral Compression Fractures (OVCFs) Executive Summary Only.
<https://www.nice.org.uk/guidance/TA279/documents/vertebral-fractures-vertebroplasty-and-kyphoplasty-medtronic6>
- Miyamoto, B. E. and E. D. Kakkis (2011). "The potential investment impact of improved access to accelerated approval on the development of treatments for low prevalence rare diseases." Orphanet Journal of Rare Diseases **6**(1): 1-13.
- National Institute for Health and Care Excellence (2016) What we do. Accessed 02/2016.
<https://www.nice.org.uk/about/what-we-do>
- National Institute for Health and Care Excellence (07/2014) Prasugrel with percutaneous coronary intervention for treating acute coronary syndromes. NICE technology appraisal guidance [TA317] <https://www.nice.org.uk/guidance/ta317>
- National Institute for Health and Care Excellence (4/04/2013a) Guide to the methods of technology appraisal 2013. Process and methods guides.
<http://publications.nice.org.uk/pmg9>
- National Institute for Health and Care Excellence (03/2013b) Venous thromboembolism in adults: diagnosis and management. NICE quality standard [QS29].
<https://www.nice.org.uk/guidance/qs29>

National Institute for Health and Care Excellence (04/2013c). Percutaneous vertebroplasty and percutaneous balloon kyphoplasty for treating osteoporotic vertebral compression fractures. NICE technology appraisal guidance [TA279]
<https://www.nice.org.uk/guidance/ta279>

National Institute for Health and Care Excellence (06/2012a) Venous thromboembolic diseases: diagnosis, management and thrombophilia testing. NICE guidelines [CG144]. <https://www.nice.org.uk/guidance/cg144>

National Institute for Health and Care Excellence (01/2012b) Apixaban for the prevention of venous thromboembolism after total hip or knee replacement in adults. NICE technology appraisal guidance [TA245]. <https://www.nice.org.uk/guidance/ta245>

National Institute for Health and Care Excellence (10/2011a) Ticagrelor for the treatment of acute coronary syndromes. NICE technology appraisal guidance [TA236].
<https://www.nice.org.uk/guidance/ta236>

National Institute for Health and Care Excellence (16/09/2011b) Final appraisal determination - Ticagrelor for the treatment of acute coronary syndromes.
<https://www.nice.org.uk/guidance/TA236/documents/acute-coronary-syndromes-ticagrelor-final-appraisal-determination-document2>

National Institute for Health and Care Excellence (01/2010a) Venous thromboembolism: reducing the risk for patients in hospital. NICE guidelines [CG92].
<https://www.nice.org.uk/guidance/cg92>

National Institute for Health and Care Excellence (03/2010b) Unstable angina and NSTEMI: early management. NICE guidelines [CG94]
<https://www.nice.org.uk/guidance/cg94>

National Institute for Health and Care Excellence (7/12/2010c) Acute coronary syndromes - ticagrelor: final scope.
<https://www.nice.org.uk/guidance/TA236/documents/acute-coronary-syndromes-ticagrelor-final-scope2>

National Institute for Health and Care Excellence (10/2009a) Prasugrel for the treatment of acute coronary syndromes with percutaneous coronary intervention. NICE technology appraisal guidance [TA182] <http://www.nice.org.uk/guidance/ta182>

National Institute for Health and Care Excellence (04/2009b) Rivaroxaban for the prevention of venous thromboembolism after total hip or total knee replacement in adults. NICE technology appraisal guidance [TA170].
<https://www.nice.org.uk/guidance/ta170>

- National Institute for Health and Care Excellence (2008a). Guide to the methods of technology appraisal. www.nice.org.uk
- National Institute for Health and Care Excellence (09/2008b) Dabigatran etexilate for the prevention of venous thromboembolism after hip or knee replacement surgery in adults. NICE technology appraisal guidance [TA157]
<https://www.nice.org.uk/guidance/ta157>
- National Institute for Health and Care Excellence Decision Support Unit (2016) Technical Support Documents. www.nicedsu.org.uk
- Neumann, P. J., D. Greenberg, N. V. Olchanski, P. W. Stone and A. B. Rosen (2005). "Growth and Quality of the Cost–Utility Literature, 1976–2001." *Value Health* **8**(1): 3-9.
- Office of National Statistics (21/03/2013) National Life Tables, 2009-2011. Interim Life Tables, England and Wales, 2009-2011.
<http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/2009-2011/index.html>
- Office of National Statistics (24/10/2013) National Life Tables, 2010-2012. Interim Life Tables, England and Wales, 2010-2012.
<http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/2010-2012/index.html>
- Owen, R. K., D. G. Tincello and R. A. Keith (2015). "Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints." *Value Health* **18**(1): 116-126.
- Pickin, M., C. L. Cooper, T. Chater, *et al.* (2009). "The Multiple Sclerosis Risk Sharing Scheme Monitoring Study--early results and lessons for the future." *BMC Neurol* **9**:1.
- Pharmaceutical Benefits Advisory Committee (07/2011) Public Summary Document - Ticagrelor, tablet, 90 mg, Brilinta®
http://www.pbs.gov.au/industry/listing/elements/pbac-meetings/psd/2011-07/Ticagrelor_BRILINTA_AstraZeneca_PSD_5-14_2011-07_FINAL.pdf
- Pharmaceutical Benefits Advisory Committee (2008). Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. www.health.gov.au
- Pharmaceutical Benefits Board (LFN) (2003). General guidelines for economic evaluations from the Pharmaceutical Benefits Board. version LFNAR 2003:2.
www.lfn.se

- Pharmaceutical Management Agency (PHARMAC) (2015) Guidelines for funding applications to PHARMAC, updated 2015.
<https://www.pharmac.govt.nz/assets/funding-applications-guidelines.pdf>
- Philips, Z., L. Ginnelly, M. Sculpher, *et al.* (2004). "Review of guidelines for good practice in decision-analytic modelling in health technology assessment."
- Poole, D., G. Bertolini and S. Garattini (2009). "Errors in the approval process and post-marketing evaluation of drotrecogin alfa (activated) for the treatment of severe sepsis." Lancet Infect Dis **9**(1): 67-72.
- Rawson, N. S., R. West and W. C. Appel (2000). "Could conditional release of new drugs provide the information required to study drug effectiveness? - A discussion paper." Can J Clin Pharmacol **7**(4): 185-190.
- Retel, V. P., M. J. Hummel and W. H. van Harten (2008). "Early phase Technology Assessment of nanotechnology in oncology." Tumori **94**(2): 284-290.
- Riemsma, R., M. Joore, J. Grutters, *et al.* (21/09/2011) Apixaban for the prevention of venous thromboembolism in people undergoing elective knee and hip replacement surgery. ERG Report. <https://www.nice.org.uk/guidance/TA245/documents/venous-thromboembolism-apixaban-hip-and-knee-surgery-evidence-review-group-report2>
- Robinson, M., L. Ginnelly, M. Sculpher, L. Jones and R. Riemsma (2002). "A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists." Health Technology Assessment **6**(25): 160.
- Robinson, R. (1993). "Cost-benefit analysis." BMJ **307**.
- Rothstein, H. R., A. J. Sutton and M. Borenstein (2006). Publication Bias in Meta-Analysis. Publication Bias in Meta-Analysis, John Wiley & Sons, Ltd: 1-7.
- Rousing, R., K. L. Hansen, M. O. Andersen, *et al.* (2010). "Twelve-months follow-up in forty-nine patients with acute/semiacute osteoporotic vertebral fractures treated conservatively or with percutaneous vertebroplasty: a clinical randomized study." Spine (Phila Pa 1976) **35**(5): 478-482.
- Rousing, R., M. O. Andersen, S. M. Jespersen, *et al.* (2009). "Percutaneous vertebroplasty compared to conservative treatment in patients with painful acute or subacute osteoporotic vertebral fractures: three-months follow-up in a clinical randomized study." Spine (Phila Pa 1976) **34**(13): 1349-1354.
- Sacks, H. S., D. Reitman, D. Pagano and B. Kupelnick (1996). "Meta-analysis: an update." Mt Sinai J Med **63**(3-4): 216-224.
- Salanti, G., C. Del Giovane, A. Chaimani, *et al.* (2014). "Evaluating the Quality of

- Evidence from a Network Meta-Analysis." PLoS ONE **9**(7).
- Salanti, G., F. K. Kavvoura and J. P. A. Ioannidis (2008). "Exploring the Geometry of Treatment Networks." Annals of Internal Medicine **148**(7): 544-553.
- Scannell, J. W., S. Hinds and R. Evans (2015). "Financial returns on R&D: looking back at history, looking forward to adaptive licensing." Rev Recent Clin Trials **10**(1): 28-43.
- Schmitz, S., R. Adams and C. Walsh (2013). "Incorporating data from various trial designs into a mixed treatment comparison model." Stat Med **32**(17): 2935-2949.
- Schneeweiss, S., J. J. Gagne, R. J. Glynn, *et al.* (2011). "Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development." Clin Pharmacol Ther **90**(6): 777-790.
- Schulz, K. F., I. Chalmers, R. J. Hayes and D. G. Altman (1995). "Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials." Jama **273**(5): 408-412.
- Scottish Medicines Consortium (9/05/2011) Brilique - Scottish Medicines Consortium. https://www.scottishmedicines.org.uk/files/advice/ticagrelor_Brilique_FINAL_APRI_L_2011_amended_030511_for_website.pdf
- Sculpher, M. J., F. S. Pang, A. Manca, *et al.* (2004). "Generalisability in economic evaluation studies in healthcare: a review and case studies." Health Technol Assess **8**.
- Sedgwick, P. (2015). "Meta-analyses: what is heterogeneity?" BMJ **350**.
- Shiroiwa, T., Y.-K. Sung, T. Fukuda, *et al.* (2010). "International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness?" Health Econ **19**.
- Siebert, U. and K. Saverno. (2011). Meta-Analysis and Indirect Treatment Comparisons from RCTs with Small Patient Numbers, Workshop presentation, ISPOR 2011, Madrid. http://www.ispor.org/congresses/spain1111/presentations/W22_siebertuwe.pdf
- Song, F., A. J. Eastwood, S. Gilbody, L. Duley and A. J. Sutton (2000). "Publication and related biases." Health Technol Assess **4**(10): 1-115.
- Sorenson, C., M. Drummond, P. Kanavos. (2008) Ensuring value for money in health care: the role of health technology assessment in the European Union. Observatory Studies Series No11. World Health Organization, on behalf of the European Observatory on Health Systems and Policies, MPG Books Ltd, Bodmin, Cornwall, UK.

- Spiegelhalter, D. J., K. R. Abrams, and J. P. Myles. (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Ltd, Chichester, UK.
- Spiegelhalter, D. J. and N. G. Best (2003). "Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling." *Stat Med* **22**(23): 3687-3709.
- Spiegelhalter, D. J., J. P. Myles, D. R. Jones and K. R. Abrams (1999). "Methods in health service research. An introduction to bayesian methods in health technology assessment." *Bmj* **319**(7208): 508-512.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. Van Der Linde (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4): 583-639.
- Spiegelhalter, D., J. Myles, D. Jones and K. Abrams (2000). "Bayesian methods in health technology assessment: a review." *Health Technology Assessment* **4**(38): 130.
- StataCorp. (2013) *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.
- Stephens, J. M., B. Handke, J. A. Doshi. (2012) "International survey of methods used in health technology assessment (HTA): Does practice meet the principles proposed for good research?" *Comparative Effectiveness Research*, **2**: 29–44.
- Sterne, J. A., P. Juni, K. F. Schulz, *et al.* (2002). "Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research." *Stat Med* **21**(11): 1513-1524.
- Stevenson, M., T. Gomersall, M. Lloyd Jones, *et al.* (8/08/2012) Percutaneous vertebroplasty and percutaneous balloon kyphoplasty for the treatment of osteoporotic vertebral fractures: a systematic review and cost effectiveness analysis. ERG report. SCHARR, University of Sheffield, Sheffield, UK.
<https://www.nice.org.uk/guidance/TA279/documents/vertebral-fractures-vertebroplasty-and-kyphoplasty-assessment-report4>
- Stevenson, M., M. Lloyd-Jones and D. Papaioannou (2009). "Vitamin K to prevent fractures in older women: systematic review and economic evaluation." *Health Technol Assess* **13**(45): iii-xi, 1-134.
- Stordeur, S., I. Vinck, M. Neyt, H. Van Brabant and F. Hulstaert (2013). "[Introduction of innovative high-risk medical devices in Europe: are clinical efficacy and safety guaranteed?]." *Rev Epidemiol Sante Publique* **61**(2): 105-110.
- Sturtz, S. and R. Bender (2012). "Unsolved issues of mixed treatment comparison meta-

- analysis: network size and inconsistency." Research Synthesis Methods **3**(4): 300-311.
- Sutton, A. J. and K. R. Abrams (2001). "Bayesian methods in meta-analysis and evidence synthesis." Statistical Methods in Medical Research **10**(4): 277-303.
- Sutton, A. J., S. J. Duval, R. L. Tweedie, K. R. Abrams and D. R. Jones (2000). "Empirical assessment of effect of publication bias on meta-analyses." BMJ : British Medical Journal **320**(7249): 1574-1577.
- Sutton, A. J., A. E. Ades, N. Cooper and K. Abrams (2008). "Use of indirect and mixed treatment comparisons for technology assessment." Pharmacoeconomics **26**(9): 753-767.
- Swedish Council on Health Technology Assessment (2012). Utvärdering av metoder i hälso- och sjukvården – En handbok. version 2010:1. www.sbu.se
- Tan, S. H., S. Bujkiewicz, A. Sutton, P. Dequen and N. Cooper (2013). "Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons." J.Health Serv.Res.Policy **18**(1758-1060 (Electronic)): 224-232.
- Tandvårds-och läkemedelsförmånsverket (08/06/2011) Brilique - application to the pharmaceutical benefits board.
http://www.tlv.se/Upload/Beslut_2011/bes110608-brilique.pdf
- Taylor, R. S. and J. Elston (2009). "The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports." Health Technol Assess **13**(8): iii, ix-xi, 1-50.
- The Institute for Quality and Efficiency in Health Care (29/09/11) IQWiG Reports - Commission No. A11-02; Ticagrelor –Benefit assessment according to § 35a Social Code Book V (Extract).
https://www.iqwig.de/download/A11-02_Extract_of_dossier_assessment_Ticagrelor.pdf
- Thompson, S. G. and J. P. T. Higgins (2002). "How should meta-regression analyses be undertaken and interpreted?" Statistics in Medicine **21**(11): 1559-1573.
- Thompson, S. G, U. Ekelund, S. Jebb, *et al.* (2011). "A proposed method of bias adjustment for meta-analyses of published observational studies." Int J Epidemiol **40**(3): 765-777.
- Thorlund, K., Z. Zafari, E. Druyts, *et al.* (2014). "The impact of incorporating Bayesian network meta-analysis in cost-effectiveness analysis - a case study of pharmacotherapies for moderate to severe COPD." Cost Effectiveness and Resource

Allocation: C/E 12(1).

- Thromboembolism in Hospitalised Patients, Second Report of Session 2004–05 (HC99). House of Commons, London: The Stationery Office Limited.
<http://www.publications.parliament.uk/pa/cm200405/cmselect/cmhealth/99/99.pdf>
- Tolley, K. (2010). "Pharmaceutical market access and the challenges of health technology assessment in the United Kingdom." *Drug Development Research* **71**(8): 478-484.
- Towse, A. (2010). "Value based pricing, research and development, and patient access schemes. Will the United Kingdom get it right or wrong?" *Br J Clin Pharmacol* **70**(3): 360-366.
- Trueman, P., D. L. Grainger and K. E. Downs (2010). "Coverage with Evidence Development: applications and issues." *Int J Technol Assess Health Care* **26**(1): 79-85.
- Tubach, F., V. Lamarque-Garnier and A. Castot (2011). "Role of the post-marketing authorisation studies in drug risk surveillance: specifications and methodologies." *Therapie* **66**(4): 355-362, 347-354.
- Turner, R. M., D. J. Spiegelhalter, G. C. Smith and S. G. Thompson (2009). "Bias modelling in evidence synthesis." *J R Stat Soc Ser A Stat Soc* **172**(1): 21-47.
- Turpie, A. G., M. R. Lassen, B. L. Davidson, *et al.* (2009). "Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty (RECORD4): a randomised trial." *Lancet* **373**(9676): 1673-1680.
- Turpie, A. G., W. D. Fisher, K. A. Bauer, *et al.* (2005). "BAY 59-7939: an oral, direct factor Xa inhibitor for the prevention of venous thromboembolism in patients after total knee replacement. A phase II dose-ranging study." *J Thromb Haemost* **3**(11): 2479-2486.
- U.S.A. Food and Drug Administration (07/20/2011) Drug Approval Package, Brilinta/Ticagrelor.
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022433Orig1s000TOC.cfm
- U.S.A. Food and Drug Administration (09/15/2014). Accelerated Approval. Accessed 02/2016. <http://www.fda.gov/ForPatients/Approvals/Fast/ucm405447.htm>
- U.S.A. Food and Drug Administration (17/11/1997) Drug Approval Package, Plavix/Clopidogrel bisulfate.
http://www.accessdata.fda.gov/drugsatfda_docs/nda/97/020839_plavix_toc.cfm

- U.S.A. Food and Drug Administration (31/12/2015) CDER Drug and Biologic Accelerated and Restricted Distribution Approvals. Accessed 02/2016. <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/DrugandBiologicApprovalReports/NDAandBLAApprovalReports/UCM404466.pdf>
- Voormolen, M. H., W. P. Mali, P. N. Lohle, *et al.* (2007). "Percutaneous vertebroplasty compared with optimal pain medication treatment: short-term clinical outcome of patients with subacute or chronic painful osteoporotic vertebral compression fractures. The VERTOS study." *AJNR Am J Neuroradiol* **28**(3): 555-560.
- Walker, S., M. Sculpher, K. Claxton and S. Palmer (2012). "Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions." *Value Health* **15**(3): 570-579.
- Wallentin, L., R. C. Becker, A. Budaj, *et al.* (2009). "Ticagrelor versus Clopidogrel in Patients with Acute Coronary Syndromes." *New England Journal of Medicine* **361**(11): 1045-1057.
- Wang, C. J., J. W. Wang, L. H. Weng, *et al.* (2004). "Prevention of deep-vein thrombosis after total knee arthroplasty in Asian patients. Comparison of low-molecular-weight heparin and indomethacin." *J Bone Joint Surg Am* **86-a**(1): 136-140.
- Wardlaw, D., S. R. Cummings, J. Van Meirhaeghe, *et al.* (2009). "Efficacy and safety of balloon kyphoplasty compared with non-surgical care for vertebral compression fracture (FREE): a randomised controlled trial." *Lancet* **373**(9668): 1016-1024.
- Warren, F. C., K. R. Abrams, S. Golder and A. J. Sutton (2012). "Systematic review of methods used in meta-analyses where a primary outcome is an adverse or unintended event." *BMC Med Res Methodol* **12**: 64.
- Warwick, D., J. Harrison, S. Whitehouse, *et al.* (2002). "A randomised comparison of a foot pump and low-molecular-weight heparin in the prevention of deep-vein thrombosis after total knee replacement." *J Bone Joint Surg Br* **84**(3): 344-350.
- Welton, N. J., A. J. Sutton, N. J. Cooper, *et al.* (2012) Front matter, in Evidence Synthesis for Decision Making in Healthcare Statistics in Practice. John Wiley & Sons, Ltd, Chichester, UK.
- Welton, N. J., A. E. Ades, J. B. Carlin, *et al.* (2009). "Models for potentially biased evidence in meta-analysis using empirically based priors." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(1): 119-136.
- Welton, N. J., N. J. Cooper, A. E. Ades, G. Lu and A. J. Sutton (2008). "Mixed treatment

- comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B." Stat Med **27**(27): 5620-5639.
- White, S., D. Ashby and P. Brown (2000). "An introduction to statistical methods for health technology assessment: a review." Health Technology Assessment **4**(8): 59.
- Wilson, N. V., S. K. Das, V. V. Kakkar, *et al.* (1992). "Thrombo-embolic prophylaxis in total knee replacement. Evaluation of the A-V Impulse System." J Bone Joint Surg Br **74**(1): 50-52.
- Wintle, B. A., M. A. McCarthy, C. T. Volinsky and R. P. Kavanagh (2003). "The Use of Bayesian Model Averaging to Better Represent Uncertainty in Ecological Models
- Wiviott, S. D., E. Braunwald, C. H. McCabe, *et al.* (2007). "Prasugrel versus Clopidogrel in Patients with Acute Coronary Syndromes." New England Journal of Medicine **357**(20): 2001-2015.
- Wonder, M., M. E. Backhouse and S. D. Sullivan (2012). "Australian managed entry scheme: a new manageable process for the reimbursement of new medicines?" Value Health **15**(3): 586-590.
- Yot Teerawattananon, M. (2008). "Thai health technology assessment guideline development." J Med Assoc Thai **91**(2): S11-15.
- Zambahari, R., O. H. Kwok, S. Javier, *et al.* (2007). "Clinical use of clopidogrel in acute coronary syndrome." International Journal of Clinical Practice **61**(3): 473-481.