Author:
Pedder, Hugo

Title:
Bridging the gap

using evidence synthesis to incorporate information from drug development into Health Technology Appraisal and beyond

Title

Bridging the gap: Using evidence synthesis to incorporate information from drug development into Health Technology Appraisal and beyond

PhD Candidate

Hugo Pedder
Student number: 2250849

Supervisors

Professor Nicky J Welton
Dr David M Phillippo
Professor Sofia Dias (External Adviser)

Abstract

Within the drug development process, drugs are first tested in preclinical trials consisting of animal and/or in vitro studies, followed by several phases of clinical trials in humans. For a new drug to be licensed it must pass through regulatory approval and, for it to be reimbursed by a national healthcare provider such as the UK NHS, be assessed in a Health Technology Appraisal (HTA). Treatments may subsequently be evaluated and recommended as part of national or international guidelines.

Synthesis of evidence is performed at different points throughout this process, but often in relative isolation and using different methodologies. Furthermore, available evidence can be sparse, which poses challenges for robust synthesis, particularly if such analyses are intended to inform decision-making.

This thesis aims to characterise the use of evidence synthesis within drug development and reimbursement, and to explore which sources of data may be useful at different points for decision-making by addressing the following questions:

1. What are the barriers to using evidence synthesis to support translation of research findings from preclinical studies to human trials?
2. What robust methods can we use to incorporate dose-response and time-course information into evidence synthesis?
3. Can we use early phase evidence to connect evidence networks in reimbursement decision-making?
4. Which tools may be helpful for evidence synthesis of drug development trials?

The thesis includes published papers and software to explore these questions, including developing and evaluating a framework for explicitly modelling dose-response and time-course relationships using Model-Based Network Meta-Analysis (MBNMA). The integrative chapter places these developments in the context of other work, discusses strengths and limitations of proposed approaches, and suggests future research that could build upon the work in the thesis.

Acknowledgements

Nicky Welton, who has been a unending source of support, inspiration and mentorship throughout my time in Bristol and who has set aside time and resources to work on this.

Sofia Dias, who mentored me in my journey into the world of evidence synthesis, and who took the time to help me learn the skills I needed when there was no-one else around me who could do so.

Meg Bennetts and Martin Boucher, who made me feel so at home in my weeks spent in Sandwich whilst developing MBNMA methods, and who gave me an insight into the pressures and interests of working with industry.

Malcolm Macleod, who set me on this journey, sparked my interest in evidence synthesis, and supported me on my first steps in academia.

David Phillippo, for his kindness, friendship, and coding support.

Gergana Tatarova, who assisted in designing the figures.

Author Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....  ..... DATE: 13th January 2023

Contents

List of Abbreviations

BEST-ITP: Bayesian Evidence Synthesis Techniques Integrated Two-component Prediction

CRAN: Comprehensive R Archive Network

EMA: European Medicines Agency

FDA: Food and Drug Administration

HTA: Health Technology Assessment

IPD: Individual Participant Data

ITC: Indirect Treatment Comparison

JAGS: Just Another Gibbs Sampler

MAIC: Matched Adjusted Indirect Comparison

MBMA: Model-Based Meta-Analysis

MBNMA: Model-Based Network Meta-Analysis

NICE: National Institute for Health and Social Care Excellence

NMA: Network Meta-Analysis

RCS: Restricted Cubic Spline

STC: Simulated Treatment Comparison

UME: Unrelated Mean Effects

1. Introduction

New medicines undergo a rigorous development process generating evidence from preclinical animal and in vitro studies, first-in-human studies, to larger scale clinical trials. This is then followed by appraisal both for the medicine's safety and for its cost-effectiveness. Evidence synthesis is the methodology used to pool results from multiple evidence sources within this process.

This thesis aims to describe the use of evidence synthesis at different stages in drug development and explore opportunities for how this can be improved, particularly with regard to making the best use of available data for analysis at each stage. More specifically, it seeks to address the following questions:

1. What are the barriers to using evidence synthesis to support translation of research findings from preclinical studies to human trials?
2. What robust methods can we use to incorporate dose-response and time-course information into evidence synthesis?
3. Can we use early phase evidence to connect evidence networks in reimbursement decision-making?
4. Which tools may be helpful for evidence synthesis of drug development trials?

The thesis starts by describing the drug development pathway, the different types of studies that are conducted at each stage, and the utilisation of evidence synthesis within it (Section 2). Different statistical approaches are used for synthesising different types of study, and these approaches may have different parameters of interest depending on their intended purpose.

In Section 3 I discuss problems that can arise in evidence synthesis during drug development and reimbursement. The subsequent impacts of these issues on drug development are explored, and the challenges in synthesis of preclinical evidence and early phase clinical trials are highlighted.

Section 4 proposes various solutions to the problems described in Section 3, identifying ways to bridge the gap between evidence syntheses conducted at different stages of drug development. My published contributions are described, along with tools to help researchers use the methods (software packages that I have developed and maintain, and a tool for data extraction).

I end in Section 5 with a discussion of the publications in the thesis and the strength and limitations of the solutions proposed. Ongoing work to resolve methodological complexities highlighted in Section 4 is described as well as suggestions for future research.

The subsequent chapters provide the publications that comprise the thesis, with each publication listed as a separate chapter. Paper P1 describes a meta-analysis of study quality in animal models of lacunar stroke, highlighting validity concerns in preclinical research. Paper P2 discusses the assessment of consistency in dose-response Model-Based Network Meta-Analysis (MBNMA). Paper P3 proposes an approach to time-course MBNMA, the performance of which is evaluated in Paper P4. Paper P5 explores how dose-response MBNMA can be used to link disconnected networks of evidence, a common issue in evidence synthesis. Paper P6 develops a tool for data extraction which can be used to minimise extraction errors and improve efficiency for systematic reviewers when extracting data for complex meta-analyses, such as those with multiple doses and time-points.

Appendices A and B consist of vignettes for the software packages I have developed, MBNMAdose and MBNMAtime. These provide detailed documentation and examples for dose-response and time-course MBNMA, guiding users through use of the packages whilst introducing them to the statistical frameworks.

2. Background

The journey of a drug from identification and early testing through to market consists of several stages of preclinical and clinical research, followed by regulatory/reimbursement appraisal (Figure 1) [1]. Evidence synthesis is employed at various points in this process, though different statistical approaches tend to be used at each stage due to the different questions addressed and the parameters targeted (Table 1). To illustrate the different stages of a drug's journey, I use eletriptan, a treatment for acute migraine, as an example in the descriptions below.

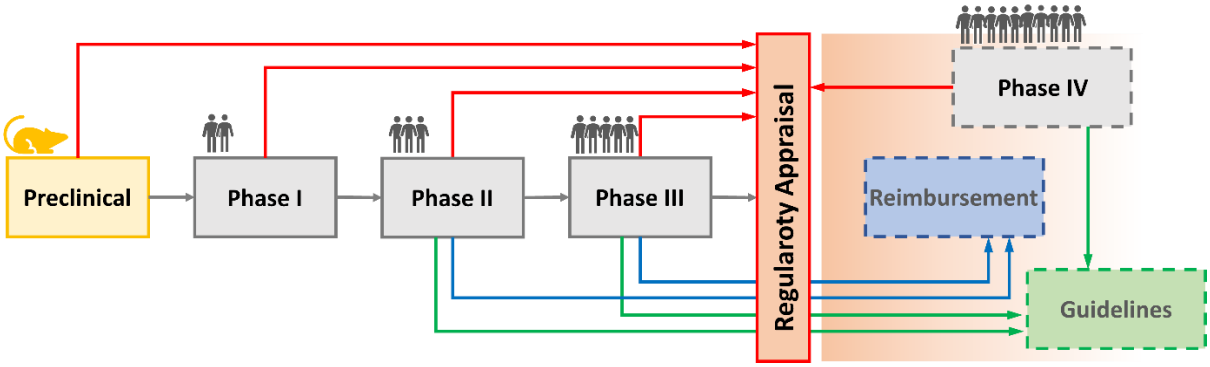

Figure 1: Illustrates the phases of drug development up to regulatory appraisal, followed by subsequent post-regulatory appraisal processes (Phase IV trials, Reimbursement, Guidelines) that are not usually required for a drug to be licensed and may be conducted once the drug is available on the market. Clinical research trial phases are shown by grey boxes, with each subsequent trial including larger numbers of participants for greater statistical power. Coloured arrows indicate which studies are used to inform the different appraisal processes, and which may therefore be used in evidence synthesis (either qualitative or quantitative). The pathway illustrated here represents a simplification of the process, as in practice a "failure" at one trial phase may result in a refinement of the target population (either for a different indication or a more specific subgroup), making the process less linear than the figure implies.

2.1 Evidence synthesis in drug development

Drug development typically consists of preclinical research, followed by "translation" of findings into clinical applications. A drug then proceeds through several stages of clinical trials. It is important to note that whilst drug development is often described as a "pathway", a drug may be tested in different phases simultaneously, because additional trials may be conducted to shed light on a specific aspect of the drug that may be of interest (e.g. different indications, specific populations).

Preclinical research consists of in vivo or in vitro studies that seek to identify and test novel promising compounds for different disease targets that could be tested in subsequent human trials[2]. In vivo studies will administer a compound to animal models of the disease in question to investigate toxicology, pharmacodynamics and whether it has the desired physiological response. For example in the case of eletriptan, a drug used to treat migraines, animal studies have been used to explore its potential as a vasoconstrictor[3] and its impact on neuroinflammation[4,5], both of which are factors thought to be implicated in migraines. In vitro studies seek to investigate more detailed receptor binding properties, as in the case of a study of eletriptan investigating its binding affinity for 5-HT receptors[6].

Evidence synthesis in preclinical research has mostly consisted of meta-analyses that seek to identify aspects of study design, methodology and reporting that impact treatment efficacy, known as "meta-epidemiological" studies[7–9]. This is in part because the substantial heterogeneity in animal models of diseases make meaningful estimation of treatment efficacy challenging, especially given the poor generalisability of such models to humans[10]. However, meta-epidemiological studies that seek to understand the ways in which animal research is carried out can be used to inform the design of

further preclinical studies that may provide additional value, or to help make a decision to proceed to running trials in humans[11]. A meta-analysis investigating treatments for migraine in preclinical rodent models of the disease aimed to optimise the design of further preclinical studies by estimating effect sizes that could be used to inform sample size calculations, whilst also estimating the impact of publication bias and risk of bias from poor methodological conduct[12].

If preclinical trials show promise and there is sufficient evidence that the safety profile of the drug is acceptable, clinical trials in humans may be initiated. A substantial number of drugs fail during clinical research, with only around 1 in every 10 drugs that are tested in humans achieving regulatory approval[13]. Phase I trials evaluate safety and tolerability, and are often conducted in a small number of healthy volunteers recruited from a homogeneous population. A phase I study for eletriptan included only 20 participants, all of whom were men aged 18-35 weighing 60-89kg[14]. Evidence syntheses of phase I trials are rare, in part because they are predominantly non-randomised but also because results for these (even aggregate data) are not made publicly available. However, synthesis of Phase I trials has previously been used to explore the risks of adverse events across a portfolio of potential therapies[15].

Phase II trials focus on exploring the pharmacokinetic properties of a drug in humans, and will typically include several doses, either in a randomised parallel or dose-escalating design. Results from Phase II trials help design dosing regimens that balance efficacy with adverse events to be used in subsequent trials. Three different doses (5mg, 20mg and 30mg) of eletriptan and placebo were compared in a randomised Phase II trial of 365 patients, which found that higher doses were likely to be required for it to be clinically and commercially attractive[16]. Subsequent studies therefore typically have investigated higher clinical doses of 40mg and 80mg[17].

Evidence from multiple Phase II studies can be synthesised to compare a drug to competitors and to help make go/no-go decisions about whether there is sufficient evidence of efficacy at a target dose[18]. Model-Based Meta-Analysis (MBMA) is the analytical method typically used for this, as it allows synthesis of both dose-response and time-course characteristics which are key for designing Phase III trials[19,20]. MBMA was used to compare the therapeutic benefit of eletriptan and sumatriptan across a range of doses and time-points, demonstrating the superiority of a lower dose of eletriptan at multiple follow-up times[21].

If the pharmacokinetic properties of the drug are promising, phase III trials are then used to demonstrate efficacy and to provide key evidence for regulatory approval and reimbursement. They are normally randomised, may include an active treatment as a comparator, and must be appropriately powered to estimate treatment efficacy for the primary outcome of interest. For eletriptan, several phase III studies have been conducted, two compared multiple doses of eletriptan with placebo[22,23] and two compared multiple doses of eletriptan with sumatriptan and placebo[24,25]. Synthesis of these trials will often be key to informing clinical inputs for models used to estimate cost-effectiveness in reimbursement submissions. Note that the phase III trial outcomes may be designed to pass regulatory approval, but that surrogate outcomes sufficient for approval can lead to considerable uncertainty for reimbursement decision-making. This will be discussed in more detail in Section 2.2.

Phase IV studies are typically carried out post-regulatory approval and are used to evaluate side effects and longer-term efficacy through what is known as “post-marketing surveillance”. These may be requested following regulatory approval and can be helpful in determining how effective a drug may be in routine clinical use, the rates of adverse events, and how the drug is used in practice (e.g. treatment duration, adherence). Such studies may be randomised or can involve observational data collection. Whilst they are not typically required for reimbursement, they may be helpful to inform economic model parameters. A phase IV non-randomised study of eletriptan was used to explore its efficacy in a specific population for whom non-steroidal anti-inflammatory drugs did not provide

satisfactory pain relief[26]. A randomised phase IV study was also conducted that aimed to investigate early migraine treatment using eletriptan[27].

Results from phase IV studies may be synthesised simultaneously with phase III trials if outcomes/populations are sufficiently similar and, given that they are often larger than phase III trials, they may be of particular interest when performing syntheses of adverse events. However, more complex methods are required if data are non-randomised[28,29]. Imbalances in variables that affect the outcome ("prognostic factors") or those that impact relative treatment effects ("effect modifiers") can cause confounding, and therefore models to synthesise non-randomised studies must ensure that all prognostic factors and effect modifiers have been accounted for.

2.2 Evidence synthesis in regulatory approval and reimbursement

For a drug to be made available it must obtain regulatory approval, which requires submission of detailed data from all trials in which the drug has been tested. This is typically sought after the successful completion of a phase III study, though increasingly manufacturers are seeking and obtaining regulatory approval based on phase II trials conducted on fewer patients that more frequently are single-arm or non-randomised[30]. This makes robust synthesis of evidence more challenging, and methods therefore need to make the best use of the data that is available.

Relevant questions raised during regulatory evaluations often relate to side effects or the potential for harm, and whether the benefits of the drug outweigh these[31]. Regulatory approval must be sought in each market separately (e.g. via the Food and Drug Administration (FDA) in the USA and via the European Medicines Agency (EMA) in Europe) and in each case regulators will examine detailed clinical study reports. Eletriptan was approved by the FDA in 2002 based on clinical study reports from eight trials, although subsequent evidence has led to updates regarding the warnings, precautions and interactions provided in the approval documentation[32].

Synthesis of all available evidence on the drug is a key part of regulatory approval, though statistical techniques for synthesis such as meta-analysis are rarely employed at this stage. Regulators may conduct evidence syntheses and meta-analyses of adverse events if concerns are raised once a drug has been available within the market, and approval may be rescinded, or warnings added, based on these[33].

Reimbursement is also necessary if a manufacturer is to make a return on their investment, and this therefore requires agreement of a price at which the drug can be purchased. Within centrally funded healthcare systems, such as in the UK, Health Technology Appraisal (HTA) provides a framework via which high value health technologies can be evaluated, not just for their efficacy, but also for their costs. Such cost-effectiveness analyses combine clinical inputs and costs to develop a Decision Analytic Model that can be used to determine whether a drug should be preferentially recommended for a specific indication[34]. In the UK, the provision of a technology that has been approved via HTA is mandatory[35]. Unlike regulators, who are more concerned with weighing the benefits versus the harms of a target drug, HTA bodies are more interested in assessing comparative evidence versus the current standard of care[31].

Given the comparative nature of the decision problem, clinical inputs typically consist of estimates of relative efficacy for which evidence synthesis and meta-analysis can be valuable. Where there are multiple studies that compare the target drug to the current standard of care, pairwise meta-analysis may be used to synthesise relative effects. However, there may be multiple comparators of interest, and/or trials evaluating the drug may have used a different comparator treatment, meaning there is no head-to-head evidence for the comparison of interest.

One common approach for estimating relative efficacy in the absence of head-to-head evidence for a comparison of interest is to perform an Indirect Treatment Comparison (ITC). For example, if we were interested in comparing eletriptan to almotriptan, there are no randomised trials that compare these treatments (Figure 2). However, in panel A both eletripan and almotriptan have been compared to placebo in RCTs. Using the consistency assumption, that direct and indirect relative effects are equal, we can estimate the indirect relative efficacy of eletriptan versus almotriptan ($\delta_{A,E}$) from the direct relative efficacies of eletriptan versus placebo ($\delta_{P,E}$) and almotriptan versus placebo ($\delta_{P,A}$):

$$\delta_{A,E} = \delta_{P,E} - \delta_{P,A}$$

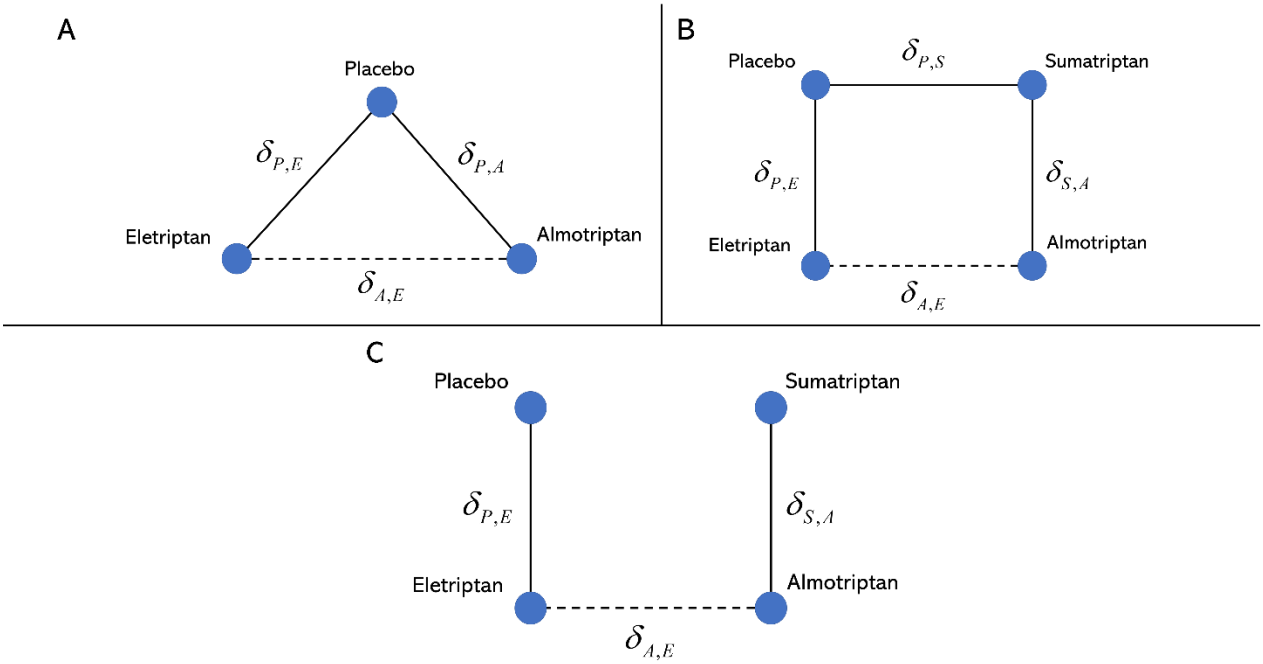

Figure 2: Network plots illustrating several ITCs. Nodes represent treatments and solid lines represent RCT evidence comparing two treatments. The dashed line represents an ITC between eletriptan and almotriptan. Panel A shows a simple triangle ITC, Panel B shows an ITC with a longer pathway of head-to-head evidence. Panel C shows an unanchored (disconnected) ITC. $\delta_{P,E}$ *represents the relative efficacy between eletriptan and placebo,* $\delta_{P,A}$ *the relative efficacy between almotriptan and placebo,* $\delta_{P,S}$ *the relative efficacy between sumatriptan and placebo, and* $\delta_{S,A}$ *the relative efficacy between almotriptan and sumatriptan.* $\delta_{A,E}$ *represents the relative efficacy between eletriptan and almotriptan, for which there is no direct RCT evidence.*

Whilst panel A shows a simple triangle ITC, consistency relationships can be extended to form indirect comparisons over longer paths (Figure 2B). The variances for relative effects are summed together when calculating an ITC, meaning that relative effects from ITCs with longer pathways of head-to-head comparisons (as in panel B) will be estimated with greater variance:

$$\text{var}(\delta_{A,E}) = \text{var}(\delta_{P,E}) + \text{var}(\delta_{P,S}) + \text{var}(\delta_{S,A})$$

To make such a comparison using the consistency relationships requires treatments to be "anchored", meaning that they are connected by a pathway of evidence[36]. To estimate an unanchored ITC between these treatments in the absence of such a pathway of evidence, as in panel C, necessitates making additional, often untestable, assumptions that are likely to raise concerns for decision-makers[37].

It is also possible to pool evidence when there is both direct and indirect evidence, so that there are loops of evidence, the generalisation of which is known as Network Meta-Analysis (NMA)[38].Such networks of evidence are often displayed in network plots, as shown in Figure 3 for the dataset of triptans for migraine relief reported by Thorlund et al.[17]. This network includes 70 studies investigating 7 triptans, evaluated at different doses, and placebo. Most studies include a direct comparison with placebo, there is both direct and indirect evidence informing many of the active (triptan vs triptan) treatment comparisons in the network, and there are several active comparisons (e.g. Eletriptan 1 vs Almotriptan 1) for which only indirect evidence is available. All these types of comparisons can be analysed simultaneously using NMA.

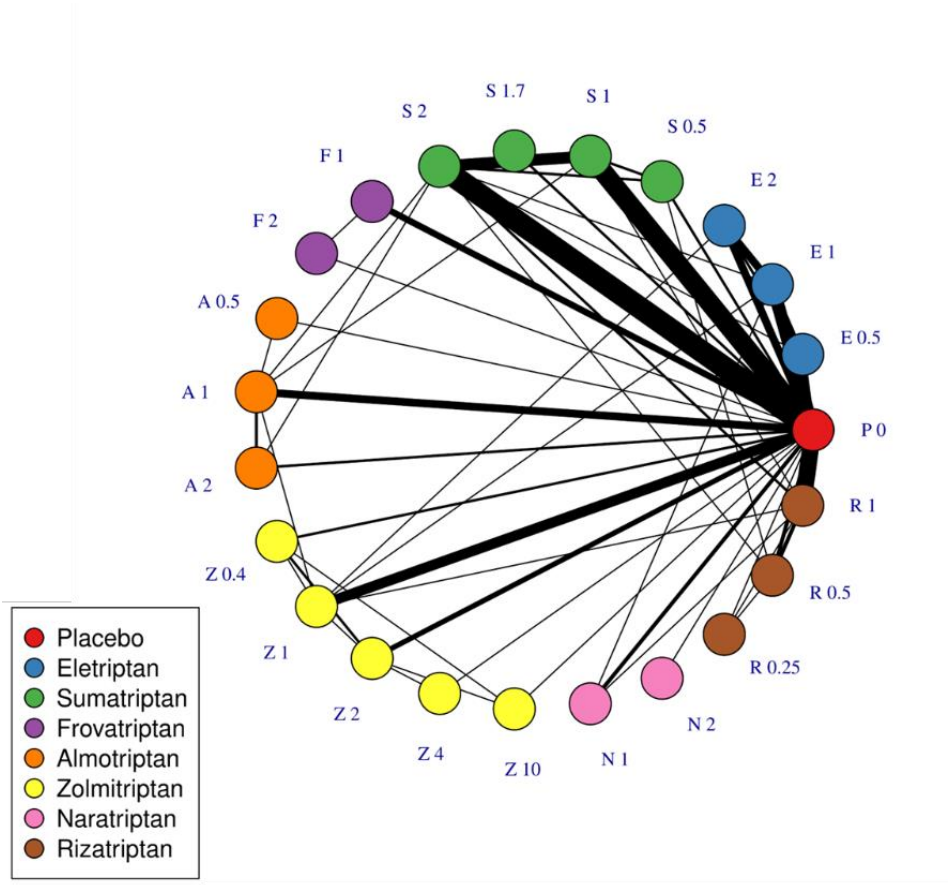

Figure 3: Network plot of studies investigating triptans for migraine relief. Nodes represent different treatments (defined as a specific dose of a specific drug) and are coloured by the drug. Connecting lines represent direct comparisons and their thickness is proportional to the number of RCTs that make a particular comparison. Treatment labels are given as the first letter of the specific triptan and the dose, standardised to the common dose of the specific triptan[17].

Various analyses assessing the cost-effectiveness of eletriptan have been conducted in different healthcare systems[39], in addition to several NMAs comparing multiple triptans (including eletriptan) to estimate their relative efficacies[40–42]. In fact, the number of treatments for migraine relief for which manufacturers have sought reimbursement is so high that this has led to the International Headache Society developing specific guidelines for their evaluation in HTA[43].

Beyond drug development and HTA, clinical guidelines may be developed that focus on the entire clinical pathway for a specific disease, rather than a single treatment or population. Nationally targeted programmes such as the National Institute for Health and Care Excellence (NICE) Centre for Guidelines[44], or international organisations such Cochrane[45] may seek to identify the most clinically and/or cost-effective of several potential treatments for different conditions, taking into account their

downstream effects and costs. For comparing multiple treatments across a wider evidence base, NMA is a useful approach for synthesis. However, heterogeneity can often be more of an issue due to the range of different populations in which different treatments are evaluated. Recommendations arising from guidelines in the UK are not mandatory, but adherence to them is encouraged and implementation is generally good[46].

Triptans, such as eletriptan, have been recommended for use in acute migraine relief in both British Association for the Study of Headaches guidelines[47] and NICE guidelines[48], which include a review of the relative strengths and limitations of each triptan based on multiple evidence syntheses and meta-analyses.

Phase of drug development	Questions asked	Parameters of interest typically targeted by evidence syntheses	Data used by[a]			
			Pharm	**Reg**	**Reimb**	**Guide**
Preclinical (animal / in vitro studies)	• How does the drug work in a biological system? • Are the effects generalisable to different species/conditions? • Should we proceed from preclinical trials to trials in humans?	• Relative treatment effects in animals • Pharmacokinetic parameters • Covariates impacting treatment efficacy • Heterogeneity	✓	✓		
Phase I	• What dose of the treatment is safe? • Are there adverse events and how frequent are they? • Should we proceed to phase II trials?	• Pharmacodynamic parameters • Proportions of adverse events	✓	✓		
Phase II	• What doses/time-points are likely to be effective? • Which subgroups of patients might benefit the most? • Should we proceed to phase III trials?	• Absolute/relative treatment effects • Covariates impacting treatment efficacy • Pharmacodynamic parameters • Proportions of adverse events	✓	✓	✓	✓
Phase III	• How effective is the drug compared to a clinically relevant comparator? • What is the frequency of adverse events? • Should the drug be recommended for clinical use?	• Relative treatment effects • Covariates impacting treatment efficacy • Proportions of adverse events	✓	✓	✓	✓
Phase IV	• What is the long-term safety? • What is the long-term efficacy?	• Absolute/relative treatment effects • Covariates impacting treatment efficacy • Proportions of adverse events	✓	✓		✓

Table 1: Use of evidence synthesis techniques across different phases of drug development

[a] ***Guide*** *= Guideline developers,* ***Pharm*** *= Pharmaceutical companies,* ***Reg*** *= Regulators,* ***Reimb*** *= Reimbursement agencies,*

3. The gap between drug development and reimbursement

As described in Section 2, although evidence synthesis is already performed at multiple stages within drug development using a variety of different data sources and methods, analyses at each stage are typically conducted in isolation of those performed` at other stages. Furthermore, data from earlier stages are rarely utilised in reimbursement decision-making or guideline development, even though they may provide valuable information to estimate clinical parameters of interest and reduce decision uncertainty.

Uncertainty in clinical inputs often arises where comparative trial evidence is limited, as is increasingly the case for HTA submissions[49], and this can lead to considerable decision uncertainty[50,51]. A survey of members of the International Network of Agencies for Health Technology Assessment indicated that data limitations were one of the top 10 challenges in appraising new technologies[52]. Data limitations can have significant consequences when conducting meta-analyses of clinical effectiveness. Direct head-to-head comparisons may be unavailable, and key parameters of interest may be difficult to reliably estimate, leading to imprecise or ungeneralisable estimates of relative treatment efficacy.

Furthermore, indirect evidence in an anchored ITC may be imprecise, or the treatment comparison may only be possible via an unanchored ITC, which makes strong untestable assumptions. This could either be by providing additional evidence to improve precision for an anchored ITC or by providing additional connectivity that may anchor an otherwise unanchored ITC. For evidence syntheses for reimbursement decision-making, incorporating evidence from earlier stages of the drug development process may therefore be of benefit. However, synthesising this evidence is challenging because of various methodological and practical issues, several of which are described in the following sections.

3.1 The validity of preclinical evidence

Preclinical studies can provide estimates of relative treatment effects versus comparators, but for them to add value in evidence syntheses of clinical trials their results must be generalisable or meaningful to humans. Outcomes must be comparable and there must be a significant degree of mechanistic similarity between diseases across different species. For animal trials, concerns regarding this are significant, with some researchers questioning whether efficacy in animals can ever be meaningfully "translated" to efficacy in humans[10,53].

In addition to these mechanistic concerns there are also severe shortcomings in the methodological design and reporting of preclinical trials. Systematic reviews and meta-analyses in several disease areas have identified that the majority of animal studies fail to correctly take steps known to reduce bias (e.g. blinding of investigators, allocation concealment)[54–57], and there is clear evidence of publication bias[7,12,58,59]. These methodological shortcomings have been shown to substantially overstate efficacy, further undermining their translational value.

3.2 Evidence from early phase trials

Early phase clinical trials provide data that are largely ignored when making reimbursement decisions. This is due to differences in study characteristics that make their synthesis with later phase trials challenging.

The value of phase I studies for evaluation of efficacy is limited mainly by their focus being on safety and evaluation of serious adverse events. This means they have a small sample size, are typically non-randomised and include a narrowly defined population (who may often even be healthy and thus not generalisable to a patient population). Without a randomised comparison, synthesising them with

other studies requires careful adjustment of baseline prognostic factors and effect modifiers. This can be particularly challenging as complete information on all important prognostic factors are unlikely to be reported in trials. Furthermore, the results of these trials are only very rarely made publicly available.

Phase II studies are larger and often make randomised comparisons between at least two different treatments and therefore relative effect estimates can be robustly synthesised without requiring adjustment of baseline prognostic factors. For phase II trials, the challenge is that they often evaluate drugs at different doses or follow-up times than those used in phase III trials, which in turn can be different from doses/follow-up times relevant for reimbursement decision-making.

Model-Based Meta-Analysis (MBMA) is a framework used in pharmacometrics for synthesis of early phase trials that allows modelling of dose-response and time-course information simultaneously. Although the framework is used to describe a range of different modelling approaches, it typically involves pooling study arms ("arm-based") rather than relative effects ("contrast-based"), thus breaking within-study randomisation[21,60–68].

An arm-based analysis pools the study-specific absolute effects on each arm:

$$g(\theta_{i,k}) = \mu_{i,k}$$

where g is a link function that transforms the outcome onto an appropriate scale (e.g. the logistic function to convert a probability to the log-odds scale, or the identity function for a continuous outcome). $\theta_{i,k}$ is the model fitted value in arm k of study i, $\mu_{i,k}$ is the absolute effect on arm k of study i. The synthesis model is then applied to $\mu_{i,k}$.

An arm-based analysis has the potential to introduce bias if prognostic variables are not accounted for, and prevents "portability" of the relative effects to populations that differ in prognostic characteristics from those in the included studies[69]. The potential for mis-specified prognostic variables and lack of ability to easily make valid inferences in a specific population of interest makes MBMA a less appealing method for HTA and guideline development.

Alternatively, a contrast-based analysis uses a Generalised Linear Model Framework to express studies in terms of their relative effects:

$$g(\theta_{i,k}) = \begin{cases} \mu_i & \text{when } k = 1 \\ \mu_i + \delta_{i,k} & \text{when } k \geq 2 \end{cases} \quad [1]$$

where μ_i is the absolute effect on arm 1 of study i (which can be treated as a nuisance parameter), and $\delta_{i,k}$ is the relative effect for arm k of study i. For a contrast-based analysis, the synthesis model is applied to $\delta_{i,k}$.

Whilst there can be justification for using arm-based analyses in instances of disconnected networks or sparse data, HTA agencies have shown a preference for contrast-based analyses to avoid any potential bias from unbalanced prognostic variables, which (at least in theory) can be substantial[70]. There is therefore a need for robust contrast-based synthesis methods in order to account for multiple doses and time-points that are common in phase II trials.

3.2.1 Studies reporting different doses

For clarity, I refer hereon to a drug as an "agent", and a specific dose and agent combination as a "treatment". For meta-analysis of multiple doses of an agent, common approaches fall into two extremes, lumping or splitting, neither of which properly accounts for dose.

Lumping involves assuming that different doses of each agent have a similar or common effect. These models are identical to class effect models used in the literature, in which different treatments are nested within a class[36]. The common effect model for lumping assumes different treatments have an identical effect, as in the NICE Health Technology Appraisal on TNF-alpha inhibitors for ankylosing spondylitis (TA383)[71], whereas the more relaxed assumption of exchangeability around a mean class effect allows for variability between different treatments, as in the NICE Guideline on Depression in adults (CG222)[72]. Further details and explanation of common or exchangeable lumping of doses is given in Section 4.2.1. However, given that different doses typically do have different effects, lumping them together can introduce heterogeneity and/or inconsistency[36,73].

By contrast, splitting estimates an independent effect for each different dose of each agent. Whilst it reduces heterogeneity, it leads to a loss of precision because it ignores the relationship between doses and it may even lead to networks being disconnected, which poses a problem for synthesis and decision-making[37].

A more satisfactory approach than either lumping or splitting would be one that incorporates information from the relationship between doses, but that does not introduce heterogeneity or bias.

3.2.2 Studies reporting different follow-up times

The concept of splitting and lumping can similarly be applied to studies that report different follow-up times from each other, yet in this case they are performed on the *outcome* rather than on the treatment[74]. For example, if results from studies that report outcomes at different follow-up times are synthesised this may introduce heterogeneity and/or inconsistency, particularly if treatment effects are expected to vary over time. Conversely, restricting the analysis to studies reporting results at a specific follow-up time and analysing each follow-up time separately may limit the studies that can be included, resulting in sparse or disconnected networks that can lead to unanchored comparisons between treatments of interest. A further complication arises when studies report multiple time-points. Ideally we would like to use all the information reported in the studies, but note that results from multiple follow-up times from the same study are likely to be correlated[75].

However, it may be expected that there is a functional relationship between relative effects over time. An approach that robustly incorporates a time-course relationship could therefore allow for comparison between treatments evaluated at different follow-up times without introducing heterogeneity. If an outcome is absorbing (i.e. once experienced it cannot be experienced again, such as death) then a survival-based analysis using hazards is appropriate, and methods for evidence synthesis of this data are well described[36,38]. However, if data are continuous or events can be experienced multiple times by participants, then correlation between data points needs to be accounted for within the likelihood, which creates additional modelling complexities[76].

3.3 Data extraction

Finally, there are frequent and persistent issues with data extraction in evidence syntheses, particularly when data are complex, and when synthesis is used to answer a wide range of questions (as described in Section 2) that may extend beyond contrast-based treatment effects. Extraction of data at multiple follow-up times in particular has been identified as a common challenge[77]. Clear guidance on how best to extract complex data is needed to improve this.

4. Bridging the gap

The papers included in this thesis seek to address the issues identified in Section 3. They are described in the following sections, along with a review and discussion of alternative solutions. The focus of the work is the development of general Model-Based NMA (MBNMA) frameworks for incorporating dose-response and time-course information into NMA, the use of these approaches to link disconnected networks of evidence, and the development of accessible software tools to facilitate their implementation.

4.1 Improving the value of preclinical evidence

In paper P1 we explored the value of animal studies for testing treatments for lacunar stroke[78]. Lacunar strokes are small blood vessel occlusions that occur in subcortical regions of the brain, and they account for a fifth of all strokes[79]. Whilst the effects of lacunar strokes are less pronounced than cortical ischaemic strokes, they contribute to dementia and are strongly prognostic of further strokes. Despite an urgent need, to date there are no treatments for lacunar stroke, and drugs found to be effective in animal studies have consistently failed to translate to benefit in humans.

To explore reasons for this translational failure we performed a systematic review and meta-analysis of animal studies investigating treatments for lacunar stroke (paper P1). The efficacy of 43 treatments were described in 57 publications. As has been found in other systematic reviews of animal research (Section 3.1), we found that most studies failed to report key aspects of study design intended to limit bias such as randomisation, investigator blinding, and allocation concealment. Failure to report these was associated with higher estimates of efficacy. However, to explore the association between treatment efficacy and reporting, different treatment effects were pooled together and assumed to have similar effects. This is unlikely to be a realistic assumption in practice and may have confounded the association identified between study design reporting and treatment efficacy if this differed for different treatments.

We also identified a higher prevalence of low precision studies reporting higher estimates of neurobehavioural score efficacy which may have been caused by publication bias. This is a well-recognised result of the “file-drawer problem” in which both journals and researchers are less incentivised to publish less favourable findings[80].

Our study indicated that, for neurobehavioural score in particular, efficacy was more likely to arise from low internal validity (relating to bias arising from poor study design) and external validity (relating to publication bias), than from the efficacy of the treatments themselves. This suggests that proper conduct and reporting of preclinical research in this area is critical if it is intended to inform decisions to proceed to human clinical trials.

Our study contributed to the portfolio of the CAMARADES (Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies) research group[81], and has been used to highlight issues in preclinical research across different diseases. Such concerns have since led to the development of guidelines such as ARRIVE (Animal Research: Reporting of In Vivo Experiments), though to date these have failed to significantly improve internal validity in animal studies[82]. Whilst there is continued work in this area to try to improve the reproducibility and value of preclinical research, continued concerns regarding the external validity and generalisability persist[10,53].

4.2 Incorporating evidence from early phase trials

Although MBMA is used for synthesis of early phase clinical trials, the limitations of this arm-based approach (Section 3.2) make it unsuitable for HTA and guideline development. I therefore focus here on contrast-based approaches that pool relative effects. These could be used to incorporate early phase trials in a manner that respects randomisation. However, existing methods have so far focussed on synthesising *either* multiple doses *or* multiple time-points. Key mathematical notation for the following sections is given in Table 2, with further notation relevant for each method described in the text. separately. Note that because the thesis is a collection of papers developed over time there is some variation in notation across papers P2-P5 that comprise the subsequent chapters of the thesis.

Table 2: Glossary of key mathematic notation. Some parameters may vary by study, arm, time-point or time-course parameter, and may be specified with subscripts (i, k, m and p respectively) differently in different models to reflect this. Subscripts for these parameters in the table below are shown as …

Parameter	**Interpretation**
i	Study index
k	Arm index
m	Time-point index
$t_{i,k}$	Treatment index in arm k of study i
a	Agent index
$x_{i,k}$	Dose in arm k of study i
$\theta_{...}$	Model fitted values
$\mu_{...}$	Absolute effects on arm 1
$\delta_{...}$	Relative effects
τ	Between-study standard deviation
$d_{...,1,c}$ (also written as $d_{...,c}$)	Pooled treatment effect of treatment c versus the network reference treatment

Models for synthesis of relative effects can typically be illustrated as extensions of the NMA model introduced in Dias et al.[38], which expresses studies in terms of their relative effects, as shown in equation [1] (Section 3.2). These study-specific relative effects, $\delta_{i,k}$, are assumed to be exchangeable around a pooled treatment effect:

$$\delta_{i,k} \sim N(d_{t_{i,1},t_{i,k}}, \tau^2) \quad [2]$$

with between-study standard deviation (SD), τ. $t_{i,k}$ and $t_{i,1}$ are the treatments in arm k and arm 1 of study i, respectively. Correlation between relative effects in multi-arm (≥3) trials can be accounted for following an approach described in NICE Technical Support Document 2[38].

Pooled treatment effects can be expressed in terms of basic parameters for each treatment versus a network reference treatment:

$$d_{c,k} = d_{1,k} - d_{1,c} \quad [3]$$

Where $d_{1,k}$ is the pooled effect of treatment k versus the network reference treatment, and $d_{1,c}$ is the pooled effect of treatment c versus the network reference treatment. These are often termed

consistency equations. Since all models can be expressed via consistency relationships in terms of the basic parameters, I will subsequently simplify the notation of the basic parameters and give them a single index (e.g. $d_k = d_{1,k}$).

In equation [2] a common treatment effect can be specified by setting τ equal to zero hence:

$$\delta_{i,k} = d_{t_{i,k}} - d_{t_{i,1}}$$

4.2.1 Incorporating dose-response information

The simplest approach to model multiple doses is using the "lumping" approach described in Section 3.2.1. This fits a class effect model to the pooled treatment effects within an agent. By defining a treatment as a specific dose of a specific agent, an effect at the agent-level can be specified:

$$d_k \sim N(D_a, \tau_a^2) \quad \text{for } k \in \{a\} \qquad [4]$$

where each treatment k is a specific dose of agent a. The pooled treatment effects, d_k are assumed to be exchangeable around D_a, the mean pooled effect of agent a, with within-agent SD of τ_a. τ_a can be estimated separately for each agent, can be assumed to be equal across all agents in the network, or can be set to zero for a common agent effect.

As described in Section 3.2.1 this approach for modelling multiple doses of agents is likely to introduce heterogeneity/inconsistency, and it ignores the possibility of a dose-response relationship. A more robust method for incorporating multiple doses of different agents would be one that incorporates dose-response information in a manner that does not introduce heterogeneity or bias. Approaches can be split into those that formally model a dose-response function (parametric) and those that do not (non-parametric). Below I include a separate section to specifically describe the dose-response Model-Based NMA (MBNMA) framework and my contributions to its development.

Non-parametric models

Owen et al.[83] proposed a Bayesian hierarchical model in which doses were assumed to be exchangeable within agents, yet with an additional ordering constraint that imposed a monotonic dose-response such that the effect of each dose, d_k, of an agent has a greater or equal effect than the next lowest dose (e.g. $d_1 \leq d_2 \leq \cdots \leq d_k$). This was achieved via an indicator function set to be equal to 1 within each agent:

$$\prod_{c=1}^{k-1} I(d_{c+1} - d_c)$$

$$\text{where } I(d_{c+1} - d_c) = \begin{cases} 1 & \text{when } (d_{c+1} - d_c) \geq 0 \\ 0 & \text{when } (d_{c+1} - d_c) < 0 \end{cases}$$

Whilst this is a far more biologically plausible model than the classic lumped model [4], assuming monotonicity provides only limited gains in precision compared to the split standard NMA model. Furthermore, it is not possible to make predictions at doses not evaluated in included RCTs.

Del Giovane et al.[84] developed a series of Bayesian models for analysing different doses of multiple agents. One approach was a monotonic model similar to that of Owen et al., yet in which

monotonicity was achieved through a truncated prior distribution. An alternative model used a random walk process to make the assumption that doses of an agent that were close together had treatment effects that were more similar:

$$d_k \sim N(d_{k-1}, \tau_a^2) \quad \text{for } k \in \{a\}$$

The within-agent variance can be assigned a half-normal prior distribution whose variance hyperparameter is a function of the dose of treatment k, x_k, ensuring that the larger the difference between neighbouring doses, the less similar their treatment effects will be:

$$\tau_a^2 \sim N(0, f(x_k, x_{k-1}))I(0,) \quad \text{for } k \in \{a\}$$

where $f(x_k, x_{k-1})$ is the difference between doses or the log of the difference for more skewed distributions of doses.

Whilst this relaxes the assumption of monotonicity, gains in precision compared to the split NMA model are again limited. As with the Owen et al. model it is not possible to make predictions at doses not included in the dataset. Del Giovane et al.[84] also applied a regression coefficient to three different transformations of doses which allowed for fitting of parametric models more similar to those described below, though these can only be fitted to a single agent network.

Parametric models

Parametric models that apply a dose-response function to the relative effects have been proposed by several authors. Whilst they broadly follow a similar framework, different dose-response relationships have been fitted to different datasets. The assumption of all these methods is that the selected dose-response function correctly captures the true underlying dose-response relationship. However, an advantage of parametric models over non-parametric models is that they can be used to make predictions at doses not included in the dataset.

Thorlund et al.[17] proposed a Bayesian meta-regression model that adjusted treatment effects specified in [2] by the dose:

$$\delta_{i,k} \sim N(d_{t_{i,k}} - d_{t_{i,1}} + \beta x_{i,k}, \tau^2)$$

where β is a covariate for the impact of dose on the treatment effects (therefore assuming a linear dose-response) and $x_{i,k}$ is the dose in arm k of study i.

Del Giovane et al.[84] reported a similar model, in addition to their non-parametric models, but also proposed fitting the dose-response relationship to the log of the doses, thus fitting a more biologically plausible log-linear dose-response. Langford et al.[85] described several frequentist models for dose-response meta-analysis. Their contrast-based 2-stage models fitted an Emax function, commonly used in pharmacometrics[86], to a dataset of alogliptin for patients with type 2 diabetes. However, none of these models allow for dose-response parameters to vary by agent, and thus their use in networks with multiple agents is limited.

As an extension of work by Crippa et al.[87] on pairwise dose-response meta-analysis, Hamza et al.[88] developed a flexible and generalisable Bayesian approach for dose-response meta-analysis that first fits a study-specific dose-response function to relative effects, and then synthesises dose-response parameters across all studies (whilst allowing for between-study heterogeneity):

$$\delta_{i,k} = f(x_{i,k}, x_{i,1}, \mathbf{b}_{a_{i,k},i})$$

where $\mathbf{b}_{a_{i,k},i}$ is a vector of dose-response parameters for agent $a_{i,k}$ that need to be estimated within each study i.

A flexible function used in their study is one composed of a restricted cubic spline (RCS)[89]. For example, for a RCS with 3 knots there are 2 dose-response parameters ($p = 2$):

$$f(x_{i,k}, x_{i,1}, \mathbf{b}_{a_{i,k},i}) = \mathrm{b}_{1,i}\{f_1(x_{i,k}) - f_1(x_{i,1})\} + \mathrm{b}_{2,i}\{f_2(x_{i,k}) - f_2(x_{i,1})\}$$

where $\mathrm{b}_{1,i}$ and $\mathrm{b}_{2,i}$ are spline coefficients in study i, and f_1 and f_2 are the identity and RCS transformation functions respectively[89].

Study-specific dose-response parameters, $\mathbf{b}_{a_{i,k},i}$, can then by synthesised using a multivariate normal distribution around a vector of mean dose-response parameters, $\boldsymbol{\beta}_{a_{i,k}}$:

$$\mathbf{b}_{a_{i,k},i} \sim MVN(\boldsymbol{\beta}_{a_{i,k}}, \boldsymbol{\Sigma})$$

$\boldsymbol{\Sigma}$ is a $p \times p$ variance covariance matrix with diagonal elements τ^2 and off-diagonal elements equal to $\rho\tau^2$ allowing for correlation, ρ, between dose-response parameters, assumed to be equal across studies. A further hierarchy can also be introduced to assume agent-level dose-responses are exchangeable around an overall dose-response, $\mathbf{B}$:

$$\boldsymbol{\beta}_{a_{i,k}} \sim MVN(\mathbf{B}, \boldsymbol{\Sigma}^{class})$$

This additional hierarchy may improve identifiability but is only likely to be a valid assumption for agents that share the same class or mechanism of action. An extension to the method incorporates adjustment for effect modifiers[90].

A limitation of this approach is that, whilst a spline function allows for considerable flexibility in the dose-response relationship, the parameters are less interpretable than in the physiologically-derived functions used more commonly in pharmacometrics, such as the Emax model. This can create additional challenges if the use of informative priors is of interest, as might be the case for agents where limited doses are available to fully estimate the dose-response relationship. Another limitation is that RCS knots must be assumed to have the same location across agents, which requires harmonisation of doses (as described by Wu et al.[67] and Thorlund et al.[17]) that may introduce bias.

Dose-response Model-Based Network Meta-Analysis

Mawdsley et al.[91], whose work I have built upon (papers P2 and P5), fitted Emax and linear dose-response functions and allowed for dose-response parameters to vary by agent. This provides a general framework for Bayesian parametric dose-response NMA, which the authors termed "Model-Based" NMA (MBNMA). Study-specific relative effects in equation [2] can be specified in terms of a functional dose-response relationship:

$$\delta_{i,k} \sim N(f(x_{i,k}, a_{i,k}, \boldsymbol{\beta}_{a_{i,k}}) - f(x_{i,1}, a_{i,1}, \boldsymbol{\beta}_{a_{i,k}}), \tau^2) \quad [5]$$

where $x_{i,k}$ is the dose and $a_{i,k}$ is the agent in arm k of study i. $\boldsymbol{\beta}_{a_{i,k}}$ is a vector of dose-response parameters for agent $a_{i,k}$, whose length is equal to the number of dose-response parameters, p.

$f(x_{i,k}, a_{i,k}, \boldsymbol{\beta}_{a_{i,k}})$ can be any plausible function for the dose-response, providing enough information is available from the data or priors to estimate it.

A drawback of this approach is that sufficient dose-response information is required for reliable estimation of dose-response parameters for all agents in a network when complex functions are fitted. For example, the Emax function fitted by Mawdsley et al. has two dose-response parameters:

$$f(x_{i,k}, a_{i,k}, Emax_{a_{i,k}}, ED50_{a_{i,k}}) = \frac{Emax_{a_{i,k}} \times x_{i,k}}{ED50_{a_{i,k}} + x_{i,k}}$$

where the Emax parameter, $Emax_{a_{i,k}}$, is the maximum efficacy that can be achieved for agent $a_{i,k}$, and the ED50 parameter, $ED50_{a_{i,k}}$, is the dose at which 50% of the maximum efficacy is achieved for agent $a_{i,k}$. Estimation of both parameters requires at least three doses of an agent (this can include placebo which is equivalent to a dose of 0)[92].

Informative prior distributions can be used where RCT data are limited to help with estimation, and identification of these are aided by fitting a model with easily interpretable parameters, such as the Emax model. In the MBNMAdose[93] R package I developed described in Section 4.3.1, I have made improvements to the modelling approach by Mawdsley et al. to allow correlation between these parameters to inform agents with more limited dose-response information, which can also aid estimation. Other novel modelling assumptions that extend the work of Mawdsley et al. are further described in the vignette for MBNMAdose.

A significant challenge in dose-response MBNMA has been the question of how to assess the validity of the consistency assumption. In standard NMA several approaches exist to investigate loop inconsistency. The Bucher method can test for inconsistency in a single loop of treatments, though this becomes more challenging in larger networks[94]. Alternatively, an unrelated mean effects (UME) model can be used to investigate whether a model fitted only to the direct evidence in the network is a better fit to the data than the consistency NMA model[95]. Finally, node-splitting allows for direct and indirect evidence contributions to be estimated and compared[96]. Other methods for assessing consistency, such as testing for design inconsistency, are also possible[97].

However, for dose-response MBNMA consistency at both the level of treatment (dose) *and* agent is required. In paper P2, we have developed methods for assessing consistency at both levels in dose-response MBNMA models. Whilst the Bucher method is unlikely to be of value in the complex networks for which dose-response MBNMA is typically used, we describe how UME models and node-splitting can be adapted for use in MBNMA. For node-splitting we show that there are more opportunities to test for inconsistency as indirect evidence can arise from the consistency assumption and/or the dose-response relationship. A poorly fitted dose-response function can therefore introduce inconsistency at the treatment level. Finally, we propose a systematic approach for evaluating consistency in these models which improves the robustness and applicability of the dose-response MBNMA framework.

There is a degree of conceptual overlap between consistency in dose-response MBNMA and in class effect models (similar to the "lumped" approach shown in equation [4]), as both have an additional hierarchy (agent and class respectively) compared to standard NMA. The methods in paper P2 could be used to inform a similar systematic approach for assessing consistency in class models which would help improve their validity and promote their use.

4.2.2 Incorporating time-course information on continuous outcomes

As with dose-response modelling, models use either non-parametric or parametric approaches to incorporate time-course information. Whilst non-parametric approaches do not suffer from the risk of mis-specifying a time-course function, they will typically have less precision when estimating treatment effects, and effects cannot be predicted at follow-up times not included in the original studies. I focus here on models for outcomes that can be summarised as continuous summary outcomes, such as means or log-odds of response, as those for time-to-event data are well described elsewhere[36,38].

To incorporate time-course the Generalised Linear Model Framework in equation [1] (Section 3.2) has to be amended to include multiple time-points from each study and the likelihood and indices of parameters must be changed to reflect this.

Non-parametric models

Dakin et al.[98] proposed a Bayesian non-parametric approach that separated the range of follow-up times into time bins. An unconstrained study baseline was estimated separately for each bin, along with corresponding time-bin-specific relative effects.

$$g(\theta_{i,k,m}) = \begin{cases} \mu_{i,m} & \text{when } k = 1 \\ \mu_{i,m} + \delta_{i,k,m} & \text{when } k \geq 2 \end{cases} \quad [6]$$

where g is a link function that transforms the outcome onto an appropriate scale. $\theta_{i,k,m}$ is the model fitted value and $\delta_{i,k,m}$ is the relative effect at time-point m in arm k of study i, and $\mu_{i,m}$ is the absolute effect on arm 1 at time-point m of study i.

Study-specific relative effects at each time-point were normally distributed around pooled relative effects that were assumed to be piecewise constant within each time-bin, b:

$$\delta_{i,k,m} \sim N(d_{t_{i,k},b_{i,m}} - d_{t_{i,1},b_{i,m}}, \tau^2)$$

where $d_{t_{i,k},b_{i,m}}$ is the time-bin specific pooled treatment effect of treatment $t_{i,k}$ in time-bin $b_{i,m}$ versus the network reference treatment, and τ is the between-study SD.

Due to the large number of parameters (both constrained and unconstrained) in the model, estimates of effect size have been shown to be conservative in simulation[99].

Lu et al.[100] described a random walk approach that could be applied to either the baseline or the treatment effect for log-hazard models to assume that those from time-points adjacent to each other should be more similar than non-adjacent ones.

For the first time-bin, the model follows the standard Lu and Ades model[101]. Study baseline effects are assigned a vague normal prior and study relative-effects are assumed to be normally distributed around a mean pooled treatment effect:

$$\mu_{i,1} \sim N(0,1000) \text{ and } \delta_{i,k,1} \sim N(d_{t_{i,k}} - d_{t_{i,1}}, \tau^2)$$

Then for each subsequent time-bin, b, a random walk model is fitted:

$$\mu_{i,b} \sim N(\mu_{i,b-1}, \tau^2_{RW}) \text{ and } \delta_{i,k,b} \sim N(\delta_{i,k,b-1}, \tau^2_{RW})$$

where τ^2_{RW} controls the similarity of effects in neighbouring time-bins. As only a single pooled treatment effect is estimated, this model assumes proportionality of relative effects across the entire follow-up.

Multivariate meta-analysis has also been proposed as a way to model longitudinal data by dividing follow-up times into different time-bins (similarly to the above approaches) and analysing the responses within each time-bin as correlated[102]. These methods can be extended to networks of treatments[103].

Parametric models

Ding and Fu[104] describe several models, of which their "BEST-ITP" model incorporates multiple study time-points and multiple treatments, following a time-course function defined by Fu and Manner[105], which is a reparameterization of the Emax model that is commonly used in pharmacometrics:

$$g(\theta_{i,k,m}) = (\mu_i + \delta_{i,k})\left(\frac{1-e^{\lambda_{t_{i,k}} s_{i,m}}}{1-e^{\lambda_{t_{i,k}} s'_i}}\right)$$

where $s_{i,m}$ is the time at time-point m of study i, s'_i is the maximum time in study i, and $\lambda_{t_{i,k}}$ is the shape of the time-course for treatment $t_{i,k}$ in arm k of study i. Note that this model is for change from baseline summary measures and would require an extra parameter for the baseline outcome (at $s_{i,m} = 0$) if absolute summary measures are used.

$\delta_{i,k}$ thus determines the maximum study-specific relative effect across the time-course, which follow consistency equations given in [2] so that basic parameters $d_{t_{i,k}}$ represent the pooled maximum effect across the time-course for treatment $t_{i,k}$ versus the network reference treatment.

Although the model has shown to typically perform well, a simulation study has shown that it struggles to reliably estimate a constant time-course, and is likely to struggle with non-monotonic patterns[99]. Furthermore, the model does not explicitly account for correlation between study-level means at different follow-up times, although the time-course function is also fitted to the study-level variances, in addition to the means. This allows for increasing variance over time, which may help account for this correlation to some degree.

A more flexible time-course model than that used by Fu and Manner[105] is the fractional polynomial model proposed by Jansen et al.[106], which can incorporate a wider range of time-course shapes. This approach uses a variance adjustment to assume constant within-study correlation by multiplying observed standard errors at time-point m, $se_{i,k,m}$, by $1-\rho^2$, where ρ is the correlation between neighbouring time-points. Although the strength of fractional polynomials is their flexibility, the interpretability of time-course parameters is a limitation, particularly if the use of informative priors is of interest as their definition on the correct parameter scale can be a challenge.

Another approach suggested by Heinecke et al.[107] uses B-splines to model the time-course function rather than fractional polynomials. The authors demonstrate in simulation that using splines provides a yet more flexible time-course function without the computational burden of fitting fractional polynomials. Parameter interpretation is also more straightforward, since it is directly linked to the particular interval (between two knots) of the corresponding coefficient.

Time-course Model-Based Network Meta-Analysis

In paper P3, we developed a general framework for NMA of studies reporting multiple follow-up times that overcomes the issues with other models in Section 4.2.2, termed time-course Model-Based NMA (MBNMA). It describes modelling of time-course parameters using relative or arm-based effects depending on the availability of information and provides an approach for testing consistency. B-spline models developed by Heinecke et al.[107] and fractional polynomial models developed by Jansen et al.[106] are special cases of the more general time-course MBNMA framework.

Data are modelled using a multivariate normal likelihood, which allows for the correlation between time-points to be accounted for:

$$\mathbf{y}_{i,k} \sim MVN(\boldsymbol{\theta}_{i,k}, \boldsymbol{\Sigma}_{i,k})$$

where $\mathbf{y}_{i,k}$ is a vector of observed summary continuous measures and $\boldsymbol{\theta}_{i,k}$ is a vector of model fitted values in arm k of study i, of length equal to the number of time-points in study i, M_i. $\boldsymbol{\Sigma}_{i,k}$ is a $M_i \times M_i$ covariance matrix:

$$\boldsymbol{\Sigma}_{i,k} = \begin{pmatrix} se_{i,k,1}^2 & \cdots & \rho_{i,1,M_i} se_{i,k,1} se_{i,k,M_i} \\ \vdots & \ddots & \vdots \\ \rho_{i,1,M_i} se_{i,k,1} se_{i,k,M_i} & \cdots & se_{i,k,M_i}^2 \end{pmatrix}$$

where $se_{i,k,m}$ is the standard error at time-point m in arm k of study i, and $\rho_{i,1,m}$ is the correlation between summary measures at time-point 1 and time-point m. Simplifying assumptions can be made to improve identifiability of $\rho_{i,1,m}$ by assuming constrained covariance structures (e.g. compound symmetry, autoregressive) and estimating a single parameter ρ to which a prior distribution (e.g. $U(0,1)$) can be assigned[74]. Alternatively, a yet simpler model can be fitted by assuming $\rho_{i,1,m} = 0$, in which case a univariate normal likelihood can be used. However, failing to account for within-study correlation if it is present can lead to estimates with biased standards errors, as we identified in a simulation study[108], and has been described in multivariate meta-analysis[75].

The time-course model is then applied to the model fitted values:

$$\boldsymbol{\theta}_{i,k} = f(\mathbf{s}_i, \boldsymbol{\lambda}_{i,k})$$

Where $\mathbf{s}_i$ is a vector of length M_i of follow-up times at each time-point in each study i, and $\boldsymbol{\lambda}_{i,k}$ is a vector of time-course parameters in arm k of study i, whose length corresponds to the number of time-course parameters, p. In paper P3, we fitted an Emax function to an illustrative dataset of treatments for osteoarthritis. However, the framework is general and can allow for modelling of any desired time-course function.

Each time-course parameter is split into study-specific baseline and relative effects:

$$\begin{aligned} g_1(\lambda_{1,i,k}) &= \mu_{1,i} + \delta_{1,i,k} \\ g_2(\lambda_{2,i,k}) &= \mu_{2,i} + \delta_{2,i,k} \\ &\vdots \\ g_p(\lambda_{p,i,k}) &= \mu_{p,i} + \delta_{p,i,k} \end{aligned}$$

An intercept parameter, $\lambda_{0,i}$, can also be included in $\boldsymbol{\lambda}_{i,k}$ if observed data are absolute continuous outcomes rather than change-from-baseline.

Consistency equations are applied to $\delta_{p,i,k}$ for each time-course parameter following equations [2] and [3], which results in a set of pooled treatment effects $d_{p,c}$ for each treatment c versus the reference treatment, for each time-course parameter p.

$\delta_{p,i,k}$ can be assumed to be common or exchangeable (random) around pooled treatment effects. If assumed random around multiple time-course parameters then they can be drawn from a multivariate normal distribution which allows for correlation between the effects on different time-course parameters. The full model is specified in paper P3, though in practice there is rarely likely to be sufficient data to fit such a model. Hence, various simplifying assumptions are described that can improve estimation, though they may introduce heterogeneity or bias.

Consistency testing follows standard methods[95], except that consistency is assumed (and can therefore be assessed) on all time-course parameters. Testing for consistency on multiple time-course parameters simultaneously is rarely likely to be possible due to data constraints. Simplifying assumptions that may be needed for estimation of these can in themselves introduce inconsistency.

To evaluate time-course MBNMA we designed and conducted a simulation study, reported in paper P4, which investigated the performance of these models in datasets with different constraints and time-course relationships. The study identified challenges in model convergence and bias for parameter estimates when more complex time-course functions were fitted when there is limited data. This may be because of difficulties in estimating the study-specific baseline time-course function when there are very few time-points within a study. Furthermore these convergence difficulties can persist in closed loops of treatments even when there is a well-estimated indirect time-course function, since estimation of a study-specific time-course function based on limited time-course data remains an issue.

However, the study showed that whilst parameter estimates may be biased when data are limited, predictions of absolute outcomes (e.g. mean response on each treatment) at specific time-points from the model were more reliably estimated. This gives guidance for when these models may be most useful, for example for informing economic models when estimation of absolute means is required.

Paper P4 highlighted the importance of correctly accounting for correlation between time-points, as has been previously described with regard to multivariate meta-analysis[75]. The study also proposed and illustrated a model selection strategy based on Deviance Information Criteria that was shown to be reliable for identifying an appropriate model, even when data were limited.

4.2.3 Linking disconnected networks of evidence

One of the benefits of incorporating evidence from early phase trials is that they may allow relative effects to be estimated between treatments that would otherwise be disconnected. These "unanchored" comparisons pose a significant problem for HTA, as analysis of them requires making strong, often untestable assumptions[109]. For example, in multiple myeloma an assumption of equivalence between two treatments was required to connect the network and estimate the comparison of interest[110]. Unanchored comparisons are becoming increasingly prevalent in HTA submissions in part due to accelerated regulatory approvals based on single-arm evidence[111].

Stevens et al.[37] categorised seven different methodological approaches that could be used to link disconnected networks. Of these, the most commonly employed in HTA is population adjustment, typically either via Simulated Treatment Comparison (STC) or Matched Adjusted Indirect

Comparison (MAIC). This is because Individual Participant Data (IPD), required for these methods, is typically available from a manufacturer's trial. However, making an unanchored comparison using population adjustment assumes conditional constancy of absolute effects (i.e. that all prognostic and effect modifying variables have been controlled for), a far stronger assumption than that required when making an anchored (connected) comparison[112]. This is a difficult assumption to meet, particularly given that data on important prognostic variables may not even have been collected in the relevant trials, and it cannot easily be tested. Reimbursement agencies such as NICE have issued guidance recommending against making such comparisons unless no alternatives exist, and this will typically lead to a higher threshold of cost-effectiveness in order to offset the additional uncertainty that an unanchored comparison will introduce[109].

Including earlier phase trials in HTAs may allow for anchored comparisons to be made without needing to make such strong assumptions. Challenges arising from incorporating data at different time-points or doses can be addressed using methods such as MBNMA, without the need to assume conditional constancy of absolute effects. Furthermore, the additional assumptions made in MBNMA (either that the dose-response or time-course has been correctly specified) can be assessed by comparing the model fit where data are available, and by clinical and biological expertise where interpolation or extrapolation outside the range of the data is required.

In paper P5 I describe how dose-response MBNMA can be used to link disconnected networks of evidence via the dose-response relationship. Two types of linking are described, the first between different doses of the same agent, and the second between different agents via an extrapolated placebo response. I illustrated this via artificially disconnecting a network of triptans for migraine relief and analysing different disconnected comparisons. MBNMA results for unanchored comparisons were consistent than those from connected networks analysed using standard NMA. The study also highlighted that even in a fully connected network, dose-response MBNMA provided greater precision over standard NMA.

4.3 Development of accessible tools

Whilst conducting this research, I have developed several comprehensive tools to facilitate the incorporation of data from drug development into evidence syntheses. Two of these are R packages, publicly available on the R package repository, CRAN (the Comprehensive R Archive Network)[113]. Listing on CRAN requires that uploaded packages pass an extensive battery of testing, that they include detailed documentation, and that they are regularly maintained to ensure continued functionality. Ongoing maintenance involves bug fixing and updates to ensure consistency with other R packages as well as introducing improvements following requests from users. Both packages use JAGS (Just Another Gibbs Sampler)[114] for fitting Bayesian models.

4.3.1 R packages for dose-response meta-analysis

I developed MBNMAdose, an R package for dose-response MBNMA[93]. It facilitates exploration of data, model fitting using a wide range of dose-response functions, consistency testing, prediction, and easy generation of informative graphical outputs. It can fit several of the different models described in Section 4.2.1, and extends existing methods by allowing for fitting of different dose-response functions to different agents in the network. Further details are given in the vignette I have written for MBNMAdose, listed on CRAN (https://cran.r-project.org/package=MBNMAdose) and also available in Appendix A.

Another package for conducting dose-response meta-analysis is available in R. MetaStan uses a more efficient Hamiltonian Monte Carlo sampler for analyses, and allows for inclusion of covariates to

control for effect modification[115]. However, it only allows for meta-analysis of dose-response within a single agent and so cannot be used in larger networks. It also only supports four different dose-response functions, all of which are monotonic, and whilst this is likely to allow modelling of a reasonably wide range of dose-response relationships found in practice it may limit the broader applicability of the package.

Agent-specific dose-response functions

An important feature that I have incorporated into MBNMAdose that has not been described in previous dose-response MBNMA methodological papers is the fitting of agent-specific dose-response functions. Networks of evidence may include agents with substantially different mechanisms of action that cannot be assumed to share the same dose-response function. For example, simultaneous analysis of agents that have a monotonic dose-response relationships and those that do not can be challenging in the MBNMA framework proposed by Mawdsley et al[73].

However, if placebo is the network reference treatment, use of a "reference treatment" parameterisation[116], an alternative to that proposed by Dias et al.[38], can allow for fitting of agent-specific dose-response functions. Study-specific effects can be specified as relative to placebo (even if the study does not include placebo):

$$g(\theta_{i,k}) = \mu_i + \delta_{i,k}$$

where g is a link function that transforms the outcome onto an appropriate scale, $\theta_{i,k}$ is the model fitted value in arm k of study i, μ_i is a study-specific intercept, and $\delta_{i,k}$ is the study-specific relative effect for arm k of study i *versus the network reference treatment* (placebo in the case of dose-response MBNMA).

Using this parameterization, the arm-level effect in each study can be specified *versus placebo*, and therefore as a single dose-response relationship:

$$\delta_{i,k} \sim N(f(x_{i,k}, a_{i,k}, \boldsymbol{\beta}_{a_{i,k}}), \tau^2)$$

This can then be extended to Z different agent-specific dose-response functions:

$$\delta_{i,k} \sim N(f_{z_{i,k}}(x_{i,k}, a_{i,k}, \boldsymbol{\beta}_{z_{i,k},a_{i,k}}), \tau^2)$$

Parameter interpretation is similar to that specified in equation [5], except that separate dose-response parameters are estimated for each dose-response function.

In practice, this fits parameters for all agents of all dose-response functions, but only a few (as indicated by the dose-response function index, $z_{i,k}$, in arm k of study i) are informed by the data. The others are treated as nuisance parameters and can be ignored in model results.

4.3.2 R packages for time-course meta-analysis

I developed MBNMAtime, an R package for time-course MBNMA[117]. Similarly to MBNMAdose it provides a simplified syntax in R to fit models using a wide range of functions, make predictions, and explore model results, though the focus is on time-course rather than dose-response models. Fractional polynomials, B-splines, the BEST-ITP function proposed by Fu and Manner[105], and others described in Section 4.2.2 can all be fitted within the package. Further details are given in the vignette

for MBNMAtime, listed on CRAN (https://cran.r-project.org/package=MBNMAtime) and also available in Appendix B.

Treatment-specific time-course functions

An extension to the time-course MBNMA framework described in paper P3 that I implemented in the MBNMAtime package allows for combining different time-course models fitted to different treatments in a network. This allows more flexibility in the choice of time-course functions specified in the network, and in particular it can help overcome issues identified in the simulation study reported in paper P4 in which limited time-course information on some treatments can lead to convergence issues.

For time-course MBNMA an imbalance in the richness of time-course information between treatments can be a significant modelling challenge (Paper P4). Fitting different functions for different treatments in the network, as illustrated for dose-response MBNMA in Section 4.3.1, is more challenging because the functional model is across the outcome, rather than across the treatment/arm. This means that all treatments in the network must share the same time-course function. Unless information on treatments with limited time-course data can be gleaned from other sources and incorporated using informative priors, the analysis will be restricted to fitting a simpler time-course function to all treatments in the network, which may be unrealistic and could lead to poor model fit for treatments with richer time-course data, or those with different mechanisms of action.

One solution to this is to use a two-stage approach. This involves splitting the dataset between treatments with richer time-course data and those with sparser time-course data into different "subnetworks". A separate time-course MBNMA can then be performed on each subnetwork, using a more complex time-course function for the richer dataset, and a simpler function with fewer time-course parameters for the sparser dataset (Figure 4). Relative comparisons between treatments in the two datasets at specific follow-up times can then be estimated from predicted effects versus a common comparator (e.g. placebo) using the Bucher method and assuming consistency[94]. Heterogeneity will be estimated separately in each network, though this can be incorporated into relative effects at specific follow-up times by using predictive intervals.

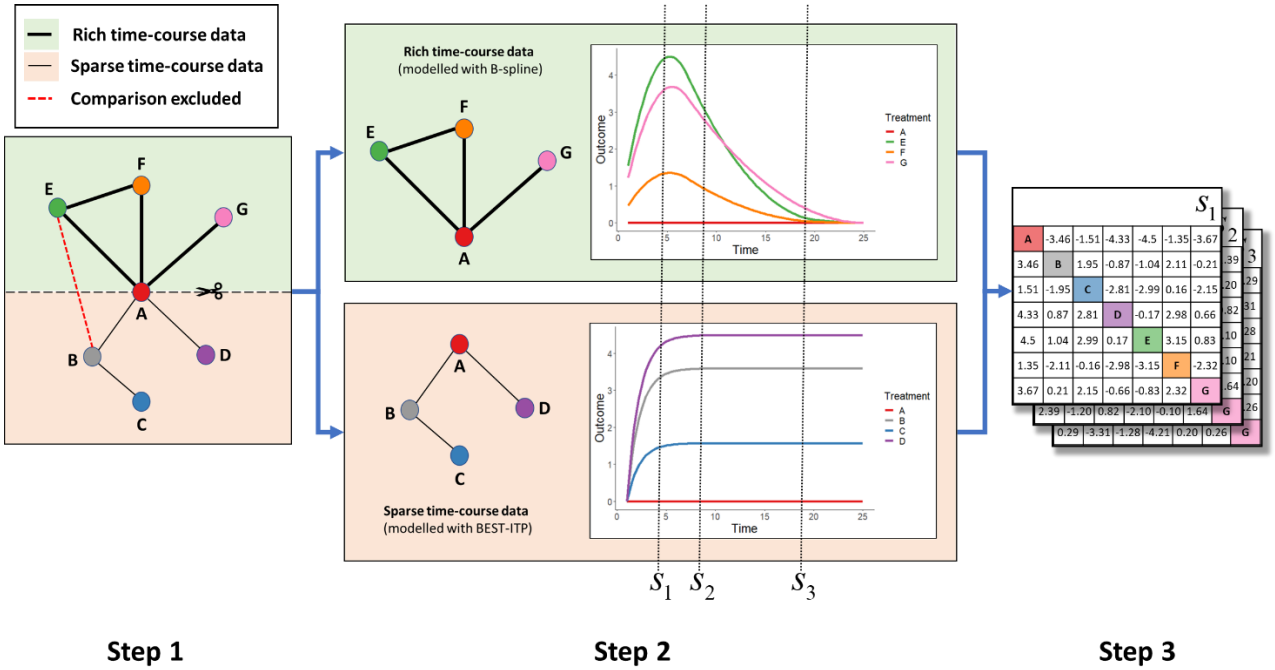

Figure 4: Two-stage time-course MBNMA fitted with different time-course functions. Step 1: Split the network at a chosen network reference treatment (A) into subnetworks with rich and sparse time-course data. Step 2: Fit separate time-course MBNMAs to each subnetwork using a different time-course function. Predict relative effects versus the network reference treatment over time. Step 3: Bucher method is used to calculate predicted relative effects between all treatments at specific time-points of interest (e.g. S_1, S_2 *and* S_3 *). For clarity, 95%CrIs are not shown in the plots or tables but can easily be calculated and computed.*
Thick connecting lines in network plots indicate comparisons with rich time-course data that can be modelled with a more complex function (e.g. B-spline), thin connecting lines in network plots indicate comparisons with sparse time-course data that can only be modelled with a less complex function (e.g. BEST-ITP). Comparisons between treatments in different subnetworks that are not the network reference must be excluded (red dashed line in network plot).

A drawback of this approach is that the dataset must be split at a common comparator, and studies comparing treatments in different subnetworks (indicated by the red dashed line in Figure 4) must be excluded, wasting potentially informative data. The decision of which treatment to split the network on is likely to be challenging in more densely connected networks. For Phase II trials many studies are placebo-controlled so splitting the network in this way may not be an issue. However, for later phase trials that compare active vs active treatments it could be more of a limitation, particularly as these trials are likely to be larger and contain more information.

This two-stage approach could be used to split the network multiple times and to fit a separate time-course function for each treatment, though in practice this is only likely to be possible in a star network with a single common comparator and few connections between other treatments to avoid excluding studies that link treatments in different subnetworks.

Although these excluded studies cannot contribute directly to the MBNMA models and results they can be used to assess consistency of the resulting relative effects since they provide a direct estimate to compare to the indirect estimate from the MBNMA analysis.

The two-stage time-course MBNMA approach is implemented in MBNMAtime and described in the vignette. It is currently being used in two ongoing projects I am collaborating with, of which a protocol is available for one[118].

4.3.3 Support for data extraction

A wide array of tools to support data extraction for meta-analysis exist, from standalone software such as Review Manager[119] and EPPI-Reviewer[120], to add-ons and packages such as amovi[121] and Meta-Essentials[122]. Automated data extraction tools, such as Robot Reviewer[123], are also providing novel ways to support systematic reviewers. However, these typically only support relatively "run-of-the-mill" meta-analyses, and for more complex analyses a standardised approach to which a software solution can be applied is challenging.

In paper P6 I developed the Data Extraction in Complex Meta-Analysis (DECiMAL) guide to support reviewers in a non-prescriptive way that allows for the flexibility required when conducting more complex data extractions. Whilst the issues raised in the DECiMAL guide can affect all types of evidence syntheses rather than specifically those used for HTA, they can be a barrier to extracting the types of data needed to perform more complex analyses that could be used to incorporate wider forms of evidence, such as those from earlier stages of drug development. The DECiMAL guide was created whilst developing the NICE guideline on Menopause (NG23)[124]. During the development of the guideline I conducted a NMA of treatments for vasomotor symptoms in menopause[125] in which several data extraction issues highlighted in the DECiMAL guide were encountered.

Issues such as these are common in guideline development due to the size and complexity of some networks of evidence. However, even for small networks, extracting and consistently coding data for multiple agents at multiple doses and/or time-points can be challenging. Given the relatively small number of data points that are often included in synthesis of aggregate data, even a single extraction error can have substantial impacts on a model's results if left unnoticed.

5. Discussion

In this thesis I have laid out some of the challenges in evidence syntheses during drug development and beyond. I have provided potential solutions to address these challenges by incorporating evidence from earlier in drug development, to bridge the gap between HTA and earlier phases of clinical and pre-clinical research. In addition to reviewing a range of different modelling approaches I have proposed methodological solutions, developed approaches to assess the validity of the data and models, evaluated the performance of the methods, developed computational tools, and a tool to enable appropriate data extraction. These contributions comprise the main body of papers included in the thesis.

5.1 Significance of publications

The publications included in this thesis have been well cited (Table 3) and their various results have been presented at 14 different conferences (12 of which were accepted for oral presentation), both nationally and internationally.

Table 3: Citations on Goggle Scholar (as of 14th November 2022)

Publication	**Citations on Google Scholar**
Paper P1[78]	44
Paper P2[126]	3
Paper P3[74]	17
Paper P4[108]	2
Paper P5[92]	6
Paper P6[127]	86

The MBNMA project was co-funded by Pfizer, who continue their interest in the work alongside developments in MBMA for application to pharmacometrics. With collaborators from Pfizer, we co-hosted a post-conference workshop on the methods at the American Conference on Pharmacometrics (AcoP) in San Diego 2018.

The collection of manuscripts detailing the MBNMA methodologies are accompanied by the two comprehensive R packages I wrote to facilitate their implementation, which include detailed vignettes and documentation. MBNMAdose and MBNMAtime have been downloaded from CRAN a total of 18,294 and 16,287 times respectively since their release (as of 14th November 2022). They are currently being used in two research projects led by international groups that I am collaborating with.

Paper P1 contributes to the body of evidence that is used by the CAMARADES group[81] to highlight widespread issues in preclinical research that can help persuade the research community of the need for robust study design.

5.2 Strengths and limitations

The MBNMA frameworks for dose-response and time-course NMA proposed and explored in this thesis are significant because they provide generalised approaches to modelling dose-response and time-course data respectively to which any functional forms can be applied. They allow inclusion of additional data that could not easily be analysed using a standard NMA approach. Given that there is often limited well-powered evidence to inform meta-analyses[128], and that networks of evidence can be disconnected, this additional information may improve precision and facilitate decision-making.

The methods make strong assumptions regarding the functional shape of the dose-response/time-course relationship, which will introduce bias if specified incorrectly[92,108]. Although the fit of these can be compared to the data where doses/time-points are available (following the model criticism and selection strategy proposed in Papers P2 and P4), extrapolation beyond the range of these is likely to be more sensitive to the choice of function and can be less easily evaluated. A good understanding of the clinical/biological plausibility of a particular function can support this, emphasising the need for collaboration between analysts and clinical experts when using these methods.

This raises a key limitation in MBNMA – more complex functional forms are data-hungry, and sufficient data may often not readily be available. Whilst a manufacturer may have access to the relevant data on their own compound (e.g. from Phase II studies), such data may not be accessible from other competitors with agents included in the network. Given that this information often exists, and is provided when seeking regulatory approval, it highlights the need for such evidence to be made publicly available for inclusion in systematic reviews. Many schemes and repositories now exist to improve data sharing in ways that avoid compromising anonymity or confidentiality, often following principles defined by the FAIR acronym (Findable, Accessible, Interoperable and Reusable)[129]. However, their adoption by industry has remained low – Tatsioni et al.[130] reported that as of April 2016, results from 67 of the 500 largest clinical trials initiated after June 2007 and completed before June 2012 remained unpublished. This proportion is likely to be far higher for smaller, earlier phase trials that can still provide valuable information for evidence synthesis.

A further issue can arise when analysing datasets with substantially different degrees of information on different comparisons within the network. For dose-response MBNMA this could be when there is RCT evidence at many doses for some agents, whilst only a single dose for others, and for time-course MBNMA when there are many time-points for some treatments but very few for others. Solutions to this are to fit agent-specific dose-response (Section 4.3.1) or treatment-specific time-course (Section 4.3.2) functions, allowing for more complexity where there is more evidence available to inform it. However, these methods have so far only been developed for R packages

MBNMAdose and MBNMAtime respectively, and they require further investigations to assess their statistical performance.

Finally, I acknowledge that paper P1 focuses on problems without providing concrete solutions. Although some general solutions are suggested in the paper, in the years since this paper's publication none have been widely implemented, largely due to inertia in research practices. The high prevalence of poor-quality research is not limited to preclinical research but permeates many other fields of academia including clinical research[131]. A detailed discussion of perverse incentives affecting research is outside the scope of this thesis, but it poses an intrinsic problem for evidence synthesis. Without careful thought and consideration, a biased evidence base is likely to lead to biased meta-analysis results and has consequences for research, licensing, reimbursement, and guideline recommendations that are based upon it.

5.3 Ongoing and future research

5.3.1 Model-Based Network Meta-Analysis

There are several opportunities for further developments in this area of research. A clear extension to the work would be to combine MBNMA frameworks so that both dose-response and time-course can be modelled simultaneously. The algebra for this would be relatively straightforward, by allowing dose-response relationships to be fitted to at least one time-course relative effect parameter. However, whilst these models may be less highly parameterised than a time-course MBNMA model, assumptions regarding correct specification of both time-course and dose-response functions are required, and model convergence is likely to be an issue without sufficient information across all treatments/agents (either from data or from informative prior distributions). Combined dose-response and time-course MBNMA has been implemented in an ongoing project I have been collaborating on to fit different doses of physical exercise (specified as metabolic equivalent per day) across multiple follow-up measurements.

What may be of further interest for some interventions that can be regarded as "agents", such as physical exercise, is fitting several dose-response relationships simultaneously. For example, both the intensity and the frequency of an agent could be explained by different dose-response relationships. These could be expressed as different "components" of the agent, following an approach developed by Welton et al.[132], and each component could be modelled with its own dose-response relationship.

As highlighted in Section 5.2, a limitation of time-course MBNMA is that in practice there appear to be significant between-study differences in time-course, even though the between-study variability at a single time-point may be reasonably low. This implies that study-specific effects over time may be highly variable, which could have important implications for both time-course MBNMA and for meta-analyses conducted at a single time-point. This may arise from fitting a time-course function to the baseline treatment, meaning that a model with an unconstrained baseline may resolve the variability.

Following difficulties identified in Paper P4 when analysing studies for which time-course data is sparse, an alternative time-course model that estimates an unconstrained study-specific baseline for each time-point could be fitted:

$$\theta_{i,k} = \mu_i + \delta_{i,k}$$

μ_i and $\delta_{i,k}$ are vectors of study-specific baseline effects and study-specific relative effects respectively, whose length is equal to the number of time-points in each study. Study-specific relative effects can be specified in terms of a functional time-course relationship that adheres to consistency relationships:

$$\boldsymbol{\delta}_{i,k} \sim N(f(\boldsymbol{s}_i, \boldsymbol{d}_{t_{i,k}}) - f(\boldsymbol{s}_i, \boldsymbol{d}_{t_{i,1}}), \tau^2)$$

where $\boldsymbol{d}_{t_{i,k}}$ is a vector of pooled time-course parameter treatment effects for treatment $t_{i,k}$ versus the reference treatment and τ is the between-study SD.

This model may have more parameters if studies include many time-points, but it makes no time-course assumption about the baseline effects and therefore may perform better when analysing studies that report measurements at only very few follow-up times. Furthermore, it may be easier to specify treatment-specific time-course functions, similarly to that described for agent-specific dose-response functions in Section 4.3.1. However, the performance of this unconstrained baseline model compared to the time-course baseline model described in Paper P3 needs to be explored further in simulation. This model has not been described or explored in published time-course MBNMA papers to date but will in the future be incorporated into the MBNMAtime R package and described in the vignette.

Whilst the work in this thesis describes and develops statistical frameworks for MBNMA, I have not explored the consequences for decision-making arising from these models. The selection of doses/time-points to include in an economic model will primarily be informed by clinical expertise but is likely to have an impact on the cost-effectiveness of a drug. I have been awarded a University of York Centre for Health Economics Visiting Fellowship to explore this further for a psoriasis Decision-Analytic Model in which we will investigate sharing of dose-response information from adult studies to inform a sparse network of studies in children and adolescents.

MBNMA models allow for information sharing using a biological and clinical understanding of the structural dose-response/time-course relationship. This means that assumptions regarding similarity of specific dose-response parameters or functional forms between different populations or even disease indications may be justifiable. Non-randomized evidence from earlier trials (e.g. phase 1) may also be used to inform this. The visiting fellowship will conclude in May 2022 and will be used to develop a future research fellowship into information sharing using structural relationships in evidence synthesis.

Finally, although I have conducted a simulation study of time-course MBNMA (paper P4), a detailed simulation study of dose-response MBNMA would be of great benefit. There now exist several methods for dose-response NMA, and a study that evaluates these across a range of different dataset characteristics would help analysts choose which is likely to be the most robust in any given scenario. Factors such as the number of doses available, the presence of agent-specific dose-response functions, unbalanced prognostic variables, disconnected treatments, and the impact of heterogeneity would be interesting to explore.

5.3.2 In silico trials

Although preclinical studies have thus far generally failed to demonstrate robust translational evidence, there are findings from these studies that could in the future be used to inform syntheses of clinical evidence.

In silico trials provide a potential avenue for incorporating quantitative information from preclinical studies into clinical trials[133]. These are virtual trials in which the effects of different drugs are estimated via simulation using systems biology. Through an understanding of receptor binding activity in different biological compartments derived from preclinical trials, a drug's absorption, distribution, metabolism and excretion can be predicted.

In silico trials have been used to predict outcomes for diseases in which receptor binding properties are well characterised or where collecting data is challenging, such as for paediatric rare

diseases[134,135]. In 2009 a proof-of-concept in silico preclinical trial was deemed a reasonable alternative to an animal trial by the US FDA for type 1 diabetes[136]. If these continue to show promise in the future this this framework could be extended to clinical trials and evidence syntheses in which results from in silico trials could be used to develop prior distributions for Bayesian meta-analyses, and could be particularly informative for structural dose-response or time-course synthesis models.

5.4 Conclusions

This thesis explores opportunities for incorporating additional evidence into syntheses informing reimbursement and guideline development. Dose-response and time-course MBNMA frameworks provide robust randomised methods to achieve this by making use of data from earlier in drug development. In addition to methodological papers developing and evaluating the MBNMA frameworks, I have also developed tools and software packages to support their implementation, and these have helped contribute to wider use of the methods.

Key concerns raised in the thesis relate to the availability and robustness of preclinical and early phase clinical trial data. Increased alignment and communication between regulators and reimbursement bodies would help facilitate collection and interpretation of evidence that could be valuable for decision-making, as well as making the process more transparent to stakeholders.

References

1. Hill RG, Richards D. *Drug Discovery and Development: Technology in Transition*. 3rd ed. Elsevier Health Sciences; 2021.

2. Lo B, Field MJ, Institute of Medicine (US) Committee on Conflict of Interest in Medical Research E. *The Pathway from Idea to Regulatory Approval: Examples for Drug Development*. National Academies Press (US); 2009. Accessed January 24, 2022. https://www.ncbi.nlm.nih.gov/books/NBK22930/

3. Gupta P, Butler P, Shepperson NB, McHarg A. The in vivo pharmacological profile of eletriptan (UK-116,044): a potent and novel 5-HT1B/1D receptor agonist. *Eur J Pharmacol*. 2000;398(1):73-81. doi:10.1016/S0014-2999(00)00240-5

4. Knyihár-Csillik E, Tajti J, Csillik AE, Chadaide Z, Mihály A, Vécsei L. Effects of eletriptan on the peptidergic innervation of the cerebral dura mater and trigeminal ganglion, and on the expression of c-fos and c-jun in the trigeminal complex of the rat in an experimental migraine model. *Eur J Neurosci*. 2000;12(11):3991-4002. doi:10.1046/j.1460-9568.2000.00299.x

5. Goadsby PJ, Hoskin KL. Differential effects of low dose CP122,288 and eletriptan on Fos expression due to stimulation of the superior sagittal sinus in cat. *Pain*. 1999;82(1):15-22. doi:10.1016/S0304-3959(99)00025-1

6. Napier C, Stewart M, Melrose H, Hopkins B, McHarg A, Wallis R. Characterisation of the 5-HT receptor binding profile of eletriptan and kinetics of w3Hxeletriptan binding at human 5-HT1B and 5-HT1D receptors. Published online 1999:10.

7. Korevaar DA, Hooft L, Ter Riet G. Systematic reviews and meta-analyses of preclinical studies: Publication bias in laboratory animal experiments. *Lab Anim*. 2011;45(4):225-230. doi:10.1258/la.2011.010121

8. Sena ES, Currie GL, McCann SK, Macleod MR, Howells DW. Systematic Reviews and Meta-Analysis of Preclinical Studies: Why Perform Them and How to Appraise Them Critically. *J Cereb Blood Flow Metab*. 2014;34(5):737-742. doi:10.1038/jcbfm.2014.28

9. van der Worp HB, Howells DW, Sena ES, et al. Can Animal Models of Disease Reliably Inform Human Studies? *PLoS Med*. 2010;7(3):e1000245. doi:10.1371/journal.pmed.1000245

10. Pound P, Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. *J Transl Med*. 2018;16(1):304. doi:10.1186/s12967-018-1678-1

11. de Vries RBM, Wever KE, Avey MT, Stephens ML, Sena ES, Leenaars M. The Usefulness of Systematic Reviews of Animal Experiments for the Design of Preclinical and Clinical Studies. *ILAR J*. 2014;55(3):427-437. doi:10.1093/ilar/ilu043

12. Dolgorukova A, Protsenko E, Isaeva J, Gagloeva V, Verbitskaya E, Sokolov AY. *Meta-Analysis of the Effects of Clinically-Effective Therapeutics in the Preclinical Migraine Model as a Tool for Design Optimisation*. Neuroscience; 2022. doi:10.1101/2022.07.05.498780

13. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32(1):40-51. doi:10.1038/nbt.2786

14. Milton K, Allen M, Abel S, James G, Rance D. The safety, tolerability and pharmacokinetics of intravenuous eletriptan (UK-116,044), a new potent and selective “5HT1D-like” receptor agonist. *Cephalalgia*. 2016;16(5):76.

15. Emanuel EJ, Bedarida G, Macci K, Gabler NB, Rid A, Wendler D. Quantifying the risks of non-oncology phase I research in healthy volunteers: meta-analysis of phase I studies. *BMJ*. 2015;350(jun26 2):h3271-h3271. doi:10.1136/bmj.h3271

16. Farkkila M. A dose-finding study of eletriptan (UK-116,044)(5-30 mg) for the acute treatment of migraine. *Cephalalgia*. 1996;16(5):81.

17. Thorlund K, Mills EJ, Wu P, et al. Comparative efficacy of triptans for the abortive treatment of migraine: a multiple treatment comparison meta-analysis. *Cephalalgia*. 2014;34(4):258-267. doi:10.1177/0333102413508661

18. Mandema JW, Gibbs M, Boyd RA, Wada DR, Pfister M. Model-based meta-analysis for comparative efficacy and safety: application in drug development and beyond. *Clin Pharmacol Ther*. 2011;90(6):766-769. doi:10.1038/clpt.2011.242

19. Mould DR. Model-based meta-analysis: an important tool for making quantitative decisions during drug development. *Clin Pharmacol Ther*. 2012;92(3):283-286. doi:10.1038/clpt.2012.122

20. Mould DR. Models for disease progression: new approaches and uses. *Clin Pharmacol Ther*. 2012;92(1):125-131. doi:10.1038/clpt.2012.53

21. Mandema JW, Cox E, Alderman J. Therapeutic benefit of eletriptan compared to sumatriptan for the acute relief of migraine pain--results of a model-based meta-analysis that accounts for encapsulation. *Cephalalgia*. 2005;25(9):715-725. doi:10.1111/j.1468-2982.2004.00939.x

22. Sheftell F, Ryan R, Pitman V, Committee FTES. Efficacy, Safety, and Tolerability of Oral Eletriptan for Treatment of Acute Migraine: A Multicenter, Double-Blind, Placebo-Controlled Study Conducted in the United States. *Headache J Head Face Pain*. 2003;43(3):202-213. doi:10.1046/j.1526-4610.2003.03043.x

23. Stark R, Dahlöf C, Haughie S, Hettiarachchi J. Efficacy, Safety and Tolerability of Oral Eletriptan in the Acute Treatment of Migraine: Results of A Phase III, Multicentre, Placebo-Controlled Study Across Three Attacks. *Cephalalgia*. 2002;22(1):23-32. doi:10.1046/j.1468-2982.2002.00300.x

24. Sandrini G, Färkkilä M, Burgess G, Forster E, Haughie S. Eletriptan vs sumatriptan: A double-blind, placebo-controlled, multiple migraine attack study. *Neurology*. 2002;59(8):1210-1217. doi:10.1212/WNL.59.8.1210

25. Goadsby PJ, Ferrari MD, Olesen J, et al. Eletriptan in acute migraine: A double-blind, placebo-controlled comparison to sumatriptan. *Neurology*. 2000;54(1):156-156. doi:10.1212/WNL.54.1.156

26. Chia Y, Lim S, Wang S, Cheong YM, Denaro J, Hettiarachchi J. Efficacy of Eletriptan in Migraineurs With Persistent Poor Response to Nonsteroidal Anti-inflammatory Drugs. *Headache J Head Face Pain*. 2003;43(9):984-990. doi:10.1046/j.1526-4610.2003.03190.x

27. Brandes J, Kudrow D, Cady R, Tiseo P, Sun W, Sikes C. Eletriptan in the Early Treatment of Acute Migraine: Influence of Pain Intensity and Time of Dosing. *Cephalalgia*. 2005;25(9):735-742. doi:10.1111/j.1468-2982.2005.00981.x

28. Efthimiou O, Mavridis D, Debray TP, et al. Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med*. 2017;36(8):1210-1226. doi:10.1002/sim.7223

29. Metelli S, Chaimani A. Challenges in meta-analyses with observational studies. *Evid Based Ment Health*. 2020;23(2):83-87. doi:10.1136/ebmental-2019-300129

30. Booth CM, Del Paggio JC. Approvals in 2016: questioning the clinical benefit of anticancer therapies. *Nat Rev Clin Oncol*. 2017;14(3):135-136. doi:10.1038/nrclinonc.2017.18

31. Vreman RA, Naci H, Goettsch WG, et al. Decision Making Under Uncertainty: Comparing Regulatory and Health Technology Assessment Reviews of Medicines in the United States and Europe. *Clin Pharmacol Ther*. 2020;108(2):350-357. doi:10.1002/cpt.1835

32. U.S.A. Food and Drug Administration. *FDA Approval: RELAPAX (Eletriptan Hydrobromide) Tablets*. USA Food and Drug Administration; 2013:1-17.

33. *Evidence Synthesis and Meta-Analysis for Drug Safety: Report of CIOMS Working Group X*. CIOMS; 2016:1-156.

34. Edlin R, McCabe C, Hulme C, Hall P, Wright J. *Cost Effectiveness Modelling for Health Technology Assessment: A Practical Course*. Springer International Publishing; 2015.

35. Department of Health and Social Care. The NHS Constitution for England. www.gov.uk. Accessed June 9, 2022. https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england

36. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network Meta-Analysis for Decision Making*. Wiley; 2018.

37. Stevens JW, Fletcher C, Downey G, Sutton A. A review of methods for comparing treatments evaluated in studies that form disconnected networks of evidence. *Res Synth Methods*. 2018;9(2):148-162. doi:10.1002/jrsm.1278

38. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Mak*. 2013;33(5):607-617. doi:10.1177/0272989X12458724

39. Bhambri R, Mardekian J, Liu LZ, Schweizer E, Ramos E. A review of the pharmacoeconomics of eletriptan for the acute treatment of migraine. *Int J Gen Med*. 2015;8:27-36. doi:10.2147/IJGM.S73673

40. Xu H, Han W, Wang J, Li M. Network meta-analysis of migraine disorder treatment by NSAIDs and triptans. *J Headache Pain*. 2016;17(1):113. doi:10.1186/s10194-016-0703-0

41. VanderPluym JH, Halker Singh RB, Urtecho M, et al. Acute Treatments for Episodic Migraine in Adults: A Systematic Review and Meta-analysis. *JAMA*. 2021;325(23):2357-2369. doi:10.1001/jama.2021.7939

42. Yang CP, Liang CS, Chang CM, et al. Comparison of New Pharmacologic Agents With Triptans for Treatment of Migraine: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2021;4(10):e2128544. doi:10.1001/jamanetworkopen.2021.28544

43. Diener HC, Ashina M, Durand-Zaleski I, et al. Health technology assessment for the acute and preventive treatment of migraine: A position statement of the International Headache Society. *Cephalalgia*. 2021;41(3):279-293. doi:10.1177/0333102421989247

44. NICE guidelines | NICE guidance | Our programmes | What we do | About. NICE. Accessed February 7, 2022. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines

45. Cochrane. Accessed February 7, 2022. https://www.cochrane.org/welcome

46. Lowson K, Jenks M, Filby A, Carr L, Campbell B, Powell J. Examining the implementation of NICE guidance: cross-sectional survey of the use of NICE interventional procedures guidance by NHS Trusts. *Implement Sci*. 2015;10(1):93. doi:10.1186/s13012-015-0283-4

47. BASH Guideline Writing Committee. *National Headache Management System for Adults 2019*. British Association for the Study of Headache; 2019:1-75.

48. NICE. *Headaches in over 12s: Diagnosis and Management*. National Institute for Health and Care Excellence; 2021. Accessed June 9, 2022. https://www.nice.org.uk/guidance/cg150/chapter/Recommendations#management-2

49. Patel D, Grimson F, Mihaylova E, et al. Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials. *Value Health*. 2021;24(8):1118-1125. doi:10.1016/j.jval.2021.01.015

50. Dickson R, Boland A, Duarte R, et al. EMA and NICE Appraisal Processes for Cancer Drugs: Current Status and Uncertainties. *Appl Health Econ Health Policy*. 2018;16(4):429-432. doi:10.1007/s40258-018-0393-7

51. Walton MJ, O'Connor J, Carroll C, Claxton L, Hodgson R. A Review of Issues Affecting the Efficiency of Decision Making in the NICE Single Technology Appraisal Process. *PharmacoEconomics - Open*. 2019;3(3):403-410. doi:10.1007/s41669-018-0113-0

52. O'Rourke B, Werkö SS, Merlin T, Huang LY, Schuller T. The 'Top 10' Challenges for Health Technology Assessment: INAHTA Viewpoint. *Int J Technol Assess Health Care*. 2020;36(1):1-4. doi:10.1017/S0266462319000825

53. Sjoberg EA. Logical fallacies in animal model research. *Behav Brain Funct*. 2017;13(1):3. doi:10.1186/s12993-017-0121-8

54. Hirst JA, Howick J, Aronson JK, et al. The Need for Randomization in Animal Trials: An Overview of Systematic Reviews. *PLOS ONE*. 2014;9(6):e98856. doi:10.1371/journal.pone.0098856

55. Fabian-Jessing BK, Vallentin MF, Secher N, et al. Animal models of cardiac arrest: A systematic review of bias and reporting. *Resuscitation*. 2018;125:16-21. doi:10.1016/j.resuscitation.2018.01.047

56. Abdel-Sattar M, Krauth D, Anglemyer A, Bero L. The relationship between risk of bias criteria, research outcomes, and study sponsorship in a cohort of preclinical thiazolidinedione animal studies: a meta-analysis. *Evid-Based Preclin Med*. 2014;1(1):e00005. doi:10.1002/ebm2.5

57. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the Efficacy of NXY-059 in Experimental Focal Cerebral Ischaemia Is Confounded by Study Quality. *Stroke*. 2008;39(10):2824-2829. doi:10.1161/STROKEAHA.108.515957

58. Sena ES, Worp HB van der, Bath PMW, Howells DW, Macleod MR. Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. *PLOS Biol*. 2010;8(3):e1000344. doi:10.1371/journal.pbio.1000344

59. Friis Berntsen C, Rootwelt P, Dahm AEA. Bias in animal studies of estrogen effects on cardiovascular disease: A systematic review and meta-analysis. *Res Pract Thromb Haemost*. 2021;5(4):e12507. doi:10.1002/rth2.12507

60. Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J Pharmacokinet Pharmacodyn*. 2010;37(2):179-201. doi:10.1007/s10928-010-9152-6

61. Checchio T, Ahadieh S, Gupta P, et al. Quantitative Evaluations of Time-Course and Treatment Effects of Systemic Agents for Psoriasis: A Model-Based Meta-Analysis. *Clin Pharmacol Ther*. 2017;102(6):1006-1016. doi:10.1002/cpt.732

62. Demin I, Hamren B, Luttringer O, Pillai G, Jung T. Longitudinal model-based meta-analysis in rheumatoid arthritis: an application toward model-based drug development. *Clin Pharmacol Ther*. 2012;92(3):352-359. doi:10.1038/clpt.2012.69

63. Korell J, Martin SW, M.O. K, Ribbing J. A model-based longitudinal meta-analysis of FEV1 in randomized COPD trials. *Clin Pharmacol Ther*. 2016;99(3):314-324.

64. Luu KT, Raber SR, Nickens DJ, Vicini P. A model-based meta-analysis of the effect of latanoprost chronotherapy on the circadian intraocular pressure of patients with glaucoma or ocular hypertension. *Clin Pharmacol Ther*. 2010;87(4):421-425. doi:10.1038/clpt.2009.306

65. Mercier F, Claret L, Prins K, Bruno R. A Model-Based Meta-analysis to Compare Efficacy and Tolerability of Tramadol and Tapentadol for the Treatment of Chronic Non-Malignant Pain. *Pain Ther*. 2014;3(1):31-44. doi:10.1007/s40122-014-0023-5

66. Raber S, Mandema JW, Li H, Nickens DJ. A model-based dose-response meta-analysis of ocular hypotensive agents as a drug development tool to evaluate new therapies in glaucoma. *J Ocul Pharmacol Ther*. 2015;31(4):189-197. doi:10.1089/jop.2014.0106

67. Wu J, Banerjee A, Jin B, Menon SM, Martin SW, Heatherington AC. Clinical dose-response for a broad set of biological products: A model-based meta-analysis. *Stat Methods Med Res*. 2018;27(9):2694-2721. doi:10.1177/0962280216684528

68. Li H, Xu J, Fan X. Target-mediated pharmacokinetic/pharmacodynamic model based meta-analysis and dosing regimen optimization of a long-acting release formulation of exenatide in patients with type 2 diabetes mellitus. *J Pharmacol Sci*. 2015;127(2):170-180. doi:10.1016/j.jphs.2014.12.004

69. Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods*. 2016;7(1):23-28. doi:10.1002/jrsm.1184

70. White IR, Turner RM, Karahalios A, Salanti G. A comparison of arm-based and contrast-based models for network meta-analysis. *Stat Med*. 2019;38(27):5197-5213. doi:10.1002/sim.8360

71. Overview | TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis | Guidance | NICE. Accessed July 4, 2022. https://www.nice.org.uk/guidance/ta383

72. Overview | Depression in adults: treatment and management | Guidance | NICE. Accessed July 4, 2022. https://www.nice.org.uk/guidance/ng222

73. Mawdsley D, Bennetts M, Dias S, Boucher M, Welton N. Model-Based Network Meta-Analysis: A Framework for Evidence Synthesis of Clinical Trial Data. *CPT Pharmacomet Syst Pharmacol*. 2016;5(8):393-401. doi:10.1002/psp4.12091

74. Pedder H, Dias S, Bennetts M, Boucher M, Welton NJ. Modelling time-course relationships with multiple treatments: Model-based network meta-analysis for continuous summary outcomes. *Res Synth Methods*. 2019;10(2):267-286. doi:10.1002/jrsm.1351

75. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *J R Stat Soc A*. 2009;172:789-811.

76. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Stat Med*. 2008;27(5):670-686. doi:10.1002/sim.2913

77. Mathes T, Klaßen P, Pieper D. Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. *BMC Med Res Methodol*. 2017;17(1):152. doi:10.1186/s12874-017-0431-4

78. Pedder H, Vesterinen HM, Macleod MR, Wardlaw JM. Systematic Review and Meta-Analysis of Interventions Tested in Animal Models of Lacunar Stroke. doi:10.1161/STROKEAHA.113.003128

79. Wardlaw JM. What causes lacunar stroke? *J Neurol Neurosurg Psychiatry*. 2005;76(5):617-619. doi:10.1136/jnnp.2004.039982

80. Scargle JD. Publication Bias: The "File-Drawer" Problem in Scientific Inference. *J Sci Explor*. 14(1):91-106.

81. CAMARADES. The University of Edinburgh. Accessed January 25, 2022. https://www.ed.ac.uk/clinical-brain-sciences/research/camarades

82. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS ONE*. 2018;13(5):e0197882. doi:10.1371/journal.pone.0197882

83. Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value Health*. 2015;18(1):116-126. doi:10.1016/j.jval.2014.10.006

84. Giovane CD, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Stat Med*. 2013;32(1):25-39. doi:10.1002/sim.5512

85. Langford O, Aronson JK, van Valkenhoef G, Stevens RJ. Methods for meta-analysis of pharmacodynamic dose-response data with application to multi-arm studies of alogliptin. *Stat Methods Med Res*. Published online March 17, 2016. doi:10.1177/0962280216637093

86. Ette EI, Williams PJ. *Pharmacometrics: The Science of Quantitative Pharmacology*. John Wiley & Sons; 2007.

87. Crippa A, Discacciati A, Bottai M, Spiegelman D, Orsini N. One-stage dose–response meta-analysis for aggregated data. *Stat Methods Med Res*. 2019;28(5):1579-1596. doi:10.1177/0962280218773122

88. Hamza T, Cipriani A, Furukawa TA, Egger M, Orsini N, Salanti G. A Bayesian dose–response meta-analysis model: A simulations study and application. *Stat Methods Med Res*. 2021;30(5):1358-1372. doi:10.1177/0962280220982643

89. Harrell F. *Regression Modelling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer; 2015.

90. Hamza T, Furukawa TA, Orsini N, Cipriani A, Iglesias C, Salanti G. A dose-effect network meta-analysis model: an application in antidepressants. :33.

91. Mawdsley D, Bennetts M, Dias S, Boucher M, Welton NJ. Model-Based Network Meta-Analysis: A Framework for Evidence Synthesis of Clinical Trial Data. *CPT Pharmacomet Syst Pharmacol*. 2016;5(8):393-401. doi:10.1002/psp4.12091

92. Pedder H, Dias S, Bennetts M, Boucher M, Welton NJ. Joining the Dots: Linking Disconnected Networks of Evidence Using Dose-Response Model-Based Network Meta-Analysis. *Med Decis Making*. 2021;41(2):194-208. doi:10.1177/0272989X20983315

93. Pedder H. MBNMAdose R package: Dose-Response MBNMA Models. Accessed March 7, 2022. https://hugaped.github.io/MBNMAdose/index.html

94. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683-691.

95. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Mak*. 2013;33(5):641-656. doi:10.1177/0272989X12455847

96. van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods*. 2016;7(1):80-93. doi:10.1002/jrsm.1167

97. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Methods*. 2012;3(2):111-125. doi:10.1002/jrsm.1045

98. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat Med*. 2011;30(20):2511-2535. doi:10.1002/sim.4284

99. Tallarita M, De Iorio M, Baio G. A comparative review of network meta-analysis models in longitudinal randomized controlled trial. *Stat Med*. 2019;38(16):3053-3072. doi:10.1002/sim.8169

100. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med*. 2007;26(20):3681-3699. doi:10.1002/sim.2831

101. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105-3124. doi:10.1002/sim.1875

102. Ishak KJ, Platt RW, Joseph L, Hanley JA, Caro JJ. Meta-analysis of longitudinal studies. *Clin Trials*. 2007;4(5):525-539. doi:10.1177/1740774507083567

103. Achana FA, Cooper NC, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol*. 2014;14.

104. Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat Med*. 2013;32(15):2613-2628. doi:10.1002/sim.5688

105. Fu H, Manner D. Bayesian adaptive dose-finding studies with delayed responses. *J Biopharm Stat*. 2010;20(5):1055-1070. doi:10.1080/10543400903315740

106. Jansen JP, Vieira MC, Cope S. Network meta-analysis of longitudinal data using fractional polynomials. *Stat Med*. 2015;34(15):2294-2311. doi:10.1002/sim.6492

107. Heinecke A, Tallarita M, De Iorio M. Bayesian splines versus fractional polynomials in network meta-analysis. *BMC Med Res Methodol*. 2020;20(1):261. doi:10.1186/s12874-020-01113-9

108. Pedder H, Boucher M, Dias S, Bennetts M, Welton NJ. Performance of model-based network meta-analysis (MBNMA) of time-course relationships: A simulation study. *Res Synth Methods*. 2020;11(5):678-697. doi:10.1002/jrsm.1432

109. Phillippo DM, Dias S, Elsada A, Ades AE, Welton NJ. Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals. *Int J Technol Assess Health Care*. 2019;35(03):221-228. doi:10.1017/S0266462319000333

110. Overview | Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation | Guidance | NICE. Accessed July 18, 2022. https://www.nice.org.uk/guidance/ta311

111. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. *BMJ Open*. 2016;6(6):e011666. doi:10.1136/bmjopen-2016-011666

112. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Med Decis Mak*. 2018;38(2):200-211. doi:10.1177/0272989X17725740

113. The Comprehensive R Archive Network. Accessed March 14, 2022. https://cran.r-project.org/

114. Plummer M. JAGS. Published online 2017. http://mcmc-jags.sourceforge.net

115. Günhan BK, Röver C, Friede T. MetaStan: An R package for Bayesian (model-based) meta-analysis using Stan. *ArXiv220200502 Stat*. Published online February 1, 2022. Accessed March 7, 2022. http://arxiv.org/abs/2202.00502

116. Phillippo DM, Dias S, Ades AE, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc*. 2020;183(3):1189-1210. doi:10.1111/rssa.12579

117. Pedder H. MBNMAtime: Run Time-Course MBNMA Models. R package version 0.1.3. Published online 2020. https://CRAN.R-project.org/package=MBNMAtime

118. Belavy DL, Diwan AD, Ford J, et al. Network meta-analysis for comparative effectiveness of treatments for chronic low back pain disorders: systematic review protocol. *BMJ Open*. 2021;11(11):e057112. doi:10.1136/bmjopen-2021-057112

119. Review Manager (RevMan) [Computer program]. Version 5.4. The Cochrane Collaboration, 2020.

120. Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4.0: software for research synthesis. Published online 2010.

121. jamovi. Published online 2021. https://www.jamovi.org

122. Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Res Synth Methods*. 2017;8(4):537-553. doi:10.1002/jrsm.1260

123. RobotReviewer. RobotReviewer. Accessed July 18, 2022. https://www.robotreviewer.net

124. Overview | Menopause: diagnosis and management | Guidance | NICE. Accessed February 4, 2022. https://www.nice.org.uk/guidance/ng23

125. Sarri G, Pedder H, Dias S, Guo Y, Lumsden M. Vasomotor symptoms resulting from natural menopause: a systematic review and network meta-analysis of treatment effects from the National Institute for Health and Care Excellence guideline on menopause. *BJOG Int J Obstet Gynaecol*. 2017;124(10):1514-1523. doi:10.1111/1471-0528.14619

126. Pedder H, Dias S, Boucher M, Bennetts M, Mawdsley D, Welton NJ. Methods to assess evidence consistency in dose-response model based network meta-analysis. *Stat Med*. 41(4): 625-644. doi:10.1002/sim.9270

127. Pedder H, Sarri G, Keeney E, Nunes V, Dias S. Data extraction for complex meta-analysis (DECiMAL) guide. *Syst Rev*. 2016;5(1):212. doi:10.1186/s13643-016-0368-4

128. Turner RM, Bird SM, Higgins JPT. The Impact of Study Size on Meta-analyses: Examination of Underpowered Studies in Cochrane Reviews. *PLoS ONE*. 2013;8(3):e59202. doi:10.1371/journal.pone.0059202

129. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018. doi:10.1038/sdata.2016.18

130. Tatsioni A, Karassa FB, Goodman SN, Zarin DA, Fanelli D, Ioannidis JPA. Lost Evidence From Registered Large Long-Unpublished Randomized Controlled Trials: A Survey. *Ann Intern Med*. 2019;171(4):300-301. doi:10.7326/M19-0440

131. Vinkers CH, Lamberink HJ, Tijdink JK, et al. The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLOS Biol*. 2021;19(4):e3001162. doi:10.1371/journal.pbio.3001162

132. Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *Am J Epidemiol*. 2009;169(9):1158-1165. doi:10.1093/aje/kwp014

133. Viceconti M, Henney A, Morley-Fletcher E. In silico clinical trials: how computer simulation will transform the biomedical industry. *Int J Clin Trials*. 2016;3(2):37. doi:10.18203/2349-3259.ijct20161408

134. Clermont G, Bartels J, Kumar R, Constantine G, Vodovotz Y, Chow C. In silico design of clinical trials: A method coming of age. *Crit Care Med*. 2004;32(10):2061-2070. doi:10.1097/01.CCM.0000142394.28791.C3

135. Carlier A, Vasilevich A, Marechal M, de Boer J, Geris L. In silico clinical trials for pediatric orphan diseases. *Sci Rep*. 2018;8(1):2465. doi:10.1038/s41598-018-20737-y

136. Kovatchev BP, Breton M, Man CD, Cobelli C. In Silico Preclinical Trials: A Proof of Concept in Closed-Loop Control of Type 1 Diabetes. *J Diabetes Sci Technol Online*. 2009;3(1):44-55.