Article:
Chang, J.-Y.A. orcid.org/0000-0001-6660-5246, Chilcott, J.B. orcid.org/0000-0003-1231-7817 and Latimer, N.R. orcid.org/0000-0001-5304-5585 (2024) Challenges and opportunities in interdisciplinary research and real-world data for treatment sequences in health technology assessments. PharmacoEconomics, 42 (5). pp. 487-506. ISSN 1170-7690

https://doi.org/10.1007/s40273-024-01363-1

Title:

Challenges and Opportunities in Interdisciplinary Research and Real-World Data for Treatment Sequences in Health Technology Assessments

Jen-Yu Amy Chang[1] (ORCID: 0000-0001-6660-5246); James B. Chilcott[1] (ORCID: 0000-0003-1231-7817); Nicholas R. Latimer[1,2] (ORCID: 0000-0001-5304-5585)

[1]Sheffield Centre for Health and Related Research (SCHARR), Division of Population Health, School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom

[2]Delta Hat Limited, United Kingdom

Contact information of corresponding author:

Jen-Yu Amy Chang, MSc, RPh

Sheffield Centre for Health and Related Research (SCHARR), Division of Population Health, School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom

Address: Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Tel: +44 (0) 114 222 5454

Fax: +44 (0) 114 222 0749

E-mail: jy.a.chang@sheffield.ac.uk

Running title: Treatment Sequences in HTA: Challenges and Opportunities

Keywords: treatment sequences, health economics, economic evaluation, interdisciplinary research, real-world data, causal inference

Abstract word count: 261/450

Word count (Systematic Review): 8,763/10,000

Number of pages: 34

Number of tables: 3

Number of figures: 2

See published version here: https://link.springer.com/article/10.1007/s40273-024-01363-1

Abstract

With an ever-increasing number of treatment options, the assessment of treatment sequences has become crucial in health technology assessment (HTA). This review systematically explores the multifaceted challenges inherent in evaluating sequences, delving into their interplay and nuances that go beyond economic model structures. We synthesised a "roadmap" of literature from key methodological studies, highlighting the evolution of recent advances and emerging research themes. These insights were compared against HTA guidelines to identify potential avenues for future research.

Our findings reveal a spectrum of challenges in sequence evaluation, encompassing selecting appropriate decision-analytic modelling approaches and comparators, deriving appropriate clinical effectiveness evidence in the face of data scarcity, scrutinising effectiveness assumptions and statistical adjustments, considering treatment displacement, and optimising model computations. Integrating methodologies from diverse disciplines—statistics, epidemiology, causal inference, operational research and computer science—has demonstrated promise in addressing these challenges. An updated review of application studies is warranted to provide detailed insights into the extent and manner in which these methodologies have been implemented.

Data scarcity on the effectiveness of treatment sequences emerged as a dominant concern, especially because treatment sequences are rarely compared in clinical trials. Real-world data (RWD) provides an alternative means for capturing evidence on effectiveness and future research should prioritise harnessing causal inference methods, particularly Target Trial Emulation, to evaluate treatment sequence effectiveness using RWD. This approach is also adaptable for analysing trials harbouring sequencing information and adjusting indirect comparisons when collating evidence from heterogeneous sources. Such investigative efforts could lend support to

reviews of HTA recommendations and contribute to synthesising external control arms involving treatment sequences.

Key Points for Decision Makers

- There has been a surge in health technology assessment (HTA) research into treatment sequence evaluation. Despite advancements in modelling frameworks, the field faces multifaceted challenges that go beyond economic model structures, including selecting appropriate sequencing comparators, scarcity of clinical effectiveness evidence necessitating simplifying assumptions and statistical adjustments, considering treatment displacement effects, and computational optimisation.
- Cross-disciplinary methodologies, such as statistics, epidemiology, causal inference, operational research and computer science, show promise in addressing these challenges. For instance, real-world data (RWD) has substantial potential for informing estimates of treatment sequence effectiveness but demands analysis with appropriate statistical methods—often adapted from epidemiological and statistical research—to effectively mitigate biases.
- Future research should emphasise harnessing causal inference methods in evaluating the effectiveness of sequences, especially leveraging the Target Trial Emulation approach for sequencing analysis using RWD or clinical trials, and adjusting indirect comparisons. These investigative efforts can better inform reviews of HTA recommendations and the synthesis of external control arms involving sequences.

Declarations

Funding: This study was supported by the Wellcome Trust [108903/B/15/Z]. The sponsor had no role in the study design, data collection, analysis, and decision to publish or preparation of the manuscript.

Conflicts of interest: All authors have no conflicts of interest that are directly relevant to the content of this article.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Code Availability: Not applicable.

Author Contributions: All authors contributed to the conceptualisation and design. JYAC performed the literature search and selection, data extraction, evidence synthesis and drafted the original manuscript. JC and NL contributed insights for quality assessment and interpretation of results, revised the manuscript, and provided supervision. All authors reviewed and approved the final manuscript.

Acknowledgements

The authors would like to express their gratitude to Professor Allan Wailoo and Professor Paul Tappenden from the University of Sheffield for their valuable feedback on the initial version of the work. The authors would also like to extend their appreciation to the Sheffield Centre for Health and Related Research (SCHARR) library service at the University of Sheffield for their guidance on review methodologies.

1. Introduction

Over the last decade, there has been a growing interest in Health Technology Assessments (HTA) in considering treatment sequences alongside the existing focus on evaluating treatments at discrete points in a treatment pathway [1-5]. Given the increasing treatment options [6-10], most diseases involve a pathway of care, wherein the effectiveness and cost-effectiveness of a treatment can be affected by treatments administered before or after it. That is, alternating the order of treatments can affect the overall effectiveness, such as the overall survival (OS). These variations can, in turn, lead to changes in the overall costs of managing a disease. Therefore, to make optimal resource allocation decisions, it is imperative to consider the entire sequences, rather than individual treatments in isolation. Appraising treatment sequences requires the development of economic models that compare sequences. While defining the structure of treatment-sequencing models is recognised as challenging, other aspects, such as selecting an appropriate baseline sequencing-comparator and evidence of treatment effectiveness, are equally critical [11, 12].

To date, only the English [13, 14] and Canadian [15] HTA guidelines explicitly acknowledge the potential need to consider the sequencing of technologies [16], albeit without detailed decision-analytic modelling instructions [13-15]. For instance, the 2022 guideline from the National Institute for Health and Care Excellence (NICE) [14] suggests considering the "care pathway" for the appraised technology and its comparator when appropriate, such as the sequences of treatments and diagnostic tests. While the provided guidance is rather generic, two NICE Decision Support Unit (DSU) briefing papers [17, 18] that informed the 2013 NICE guideline [13] underscores the ongoing challenges HTA agencies face in numerous appraisals involving treatment sequences, especially regarding the selection of treatment sequences and their associated costs. Furthermore, NICE is currently piloting the Pathway Approach for technology appraisals (TAs) in two disease (i.e., renal cell carcinoma and non-small cell lung cancer) [19], which seeks to streamline the process of economic model development and review for treatments within the same disease, potentially influencing future evaluations of treatment sequences in HTA.

Despite existing reviews highlighting challenges in treatment-sequencing evaluation [11, 12], a gap remains in primary research and innovation addressing these challenges. Additionally, the interplay between various challenges and nuances beyond economic model structures is underexplored. Therefore, we aimed to update and consolidate existing literature on methodological and conceptual advancements. We curate a "roadmap" of the relevant literature to provide a holistic view on the current state of treatment-sequencing in HTA, highlighting key challenges and ongoing research areas, laying a foundation for commissioning and developing new initiatives to address the

challenges and develop new methods.

Clinical trials are typically the primary source of evidence used in HTA, but these rarely focus on comparing treatment sequences, or on the effectiveness of a treatment conditional on a specific treatment sequence used prior. Real-world data (RWD) may be more adept at capturing treatment sequences, and there it a trend towards the enhanced use of RWD in HTA—particularly evident in organisations like NICE for England and Wales [20]. Hence, a key focus of our review investigates whether literature exists that discusses the integration of RWD into HTA in the context of treatment sequencing. We explored the potential and related challenges of this integration and how it fits within the broader literature roadmap. NICE published a framework on the use of real-world evidence (RWE) in 2022 [20], which postdates most of the included papers in our study. Consequently, our review aims to contribute fresh perspectives that may enhance the findings of existing studies, by relating their findings to the implications of NICE's RWE framework.

Our paper is structured as follows: we first describe the methods of our review (2. Methods), followed by an overview of the included papers and a narrative synthesis of our findings (3. Results). Finally, we discuss the implications of our findings for HTA and recommend avenues for future research (4. Discussion and 5. Conclusion).

2. Methods

2.1 The pearl-growing approach

In light of the limitations of conventional systematic review methods for methodological literature, our research employed the pearl-growing approach [21, 22]. In particular: (1) A literature search for "treatment sequences" in the absence of established Medical Subject Headings (MeSH) and Embase Subject Headings (Emtree) could lead to biased results. (2) Our objective was to explore methodology advancements that are relevant across diverse diseases. However, during a pilot systematic search (Section 2.1.1), we noted a large number of studies that discussed treatment-sequences in a particular disease area (often multiple sclerosis or rheumatoid arthritis (RA) [23, 24]), without describing relevant methodology beyond sequencing-model structures. This made the conventional systematic review approach highly inefficient. (3) Furthermore, relying exclusively on searching electronic databases like PubMed and EMBASE could result in overlooking crucial ongoing studies and doctoral theses in this emerging topic area. Given these considerations, the pearl-growing approach appeared to be the best suited strategy.

The pearl-growing approach applied here begins with an initial key paper (i.e. the initial “grit”) and expands by

reviewing its reference list and studies that cited it (i.e. growing the "pearl") [25]. This process is iteratively repeated for each added relevant paper, eventually leading to the identification of a core set of pertinent studies. This approach has proved more efficient for reviewing complex evidence [21]. Variations of this technique exist, including using more than one initial paper [22] or utilising the approach to identify a set of comprehensive search terms rather than studies [26]. For transparency, we detailed our review procedures in the following sections.

2.1.1 Identification of the initial key paper

In October 2019, a pilot review was conducted on PubMed to identify an initial key paper, using search terms: ("economic model(s)" OR "economic evaluation(s)") AND ("treatment sequence(s)" OR "sequence(s)"). The study by Viola et al. [12], focusing on the selection of sequencing comparators, was selected as the initial "grit" for its alignment with the inclusion and exclusion criteria in Table 1, and for being the latest relevant publication with treatment sequence in its title.

2.1.2 Reference and citation tracking

The first round of reference and citation tracking began by screening the reference and citation list of the initial key paper using their titles and abstracts based on the same criteria from the pilot review (Table 1). The reference list was extracted from the full-text of Viola et al.'s study [12]. The citations were identified through Google Scholar Search, as it provides a wider range of results compared to conventional electronic databases. Following screening, we evaluated the eligibility of shortlisted records up to December 2019 by reviewing their full-text.

Upon completing the first round of reference and citation tracking, a new set of key papers for the next round was identified. The same procedure was then performed repeatedly for each new set of literature identified in each round. The process concluded when the literature searches saturated, namely when no more relevant articles could be identified. Any pertinent articles that the iterative process overlooked, but were brought to the authors attention, were manually included.

2.2 Review update

To ensure the review remained up-to-date towards publication, an update of citation tracking was performed in February 2023 for the studies included in the initial review to include additional relevant articles.

3. Results

3.1 Overview of the included studies

The review included a total of eleven studies, with six identified during the initial phase and five discovered in the subsequent update. Figure 1 illustrates the inclusion process, and Table 2 details the objectives of each study. In the initial review, besides Viola et al.'s study [12], three studies were included via reference and citation tracking [11, 27, 28], and two doctoral theses from Tosh and Kim were added post-hoc [29, 30]. Tosh's thesis, focusing on model optimisation methods, led to a derivative publication about economic models in RA [3]. Although this subsequent work was cited by Viola et al. [12], we had excluded it due to its disease-centric focus. Kim's doctoral thesis, which neither referenced nor was cited by other identified papers, was included because of its parallels to Tosh's doctoral work. In the third round (not shown in Figure 1 for simplicity), we assessed the references and citations of studies included in the second round as well as those that were manually included. Although we found further studies tangentially related to our review's objective, they had overly broad scopes. Some did not specifically investigate the methodological issues of treatment sequences, but rather broader issues of evaluating chronic diseases involving time-dependent transitions and sequential procedures [31-43]. Others predominantly comprised clinical studies, researching the clinical impacts of sequential treatment strategies and treatment switching, without an emphasis on HTA [44-51]. Further, one study is a conference abstract lacking full-text [52]. Consequently, we excluded these studies, indicating that our literature search has reached saturation. In the updated review, five new studies were included [16, 53-56], all of which cited at least one of the aforementioned six papers. One of the five was a derivative publication [53] of Lewis's doctoral thesis identified in the initial review [28].

Table 3 summarises the characteristics of the included studies. The majority of them (55%) were published between 2020-2022 [12, 16, 53-56], with the remainder from 2015-2019 [11, 27-30]. Only slightly more than half of the studies (55%) were published in peer-reviewed platforms [11, 12, 16, 53, 55, 56]. The remaining studies included four PhD theses (36%) [27-30] and a report from the NICE DSU [54]. Among all non-thesis articles (64%, n = 7), three undertook systematic reviews on methods that can be leveraged to shape the methodological framework for economic evaluation of treatment sequences [11, 53, 55], one paired a systematic review with a case study on the selection of baseline sequencing-comparators [12], while the remaining three are non-systematic reviews addressing various issues around treatment sequence evaluation [16, 54, 56].

3.2 Key research themes and their development roadmap

We identified several research themes (Table 3), including sequencing-model structures and problem structuring (82%, n = 9) [11, 12, 16, 27-30, 53, 55, 56]; developing conceptual frameworks for treatment-sequencing-modelling

(36%, n = 4) [11, 28, 53, 55]; methods to tackle the effectiveness of treatment sequences, especially in handling evidence from non-randomised data and indirect comparisons (45%, n = 5) [11, 28, 53, 55, 56]; and the potential for using RWE to inform sequencing analyses (45%, n = 5) [27, 28, 53, 55, 56]. Further developing themes included the potential of model optimisation methods (18%, n = 2) [29, 30] and causal inference methods (18%, n = 2) [54, 56], selection of treatment-sequence comparators (9%, n = 1) [12], and the impact of treatment displacement on resource allocation efficiency and equity (9%, n = 1) [27].

Figure 2 presents a "roadmap" detailing the evolution of treatment-sequencing research. It highlights the intersections and shifts among various research topics, revealing an increasing trend towards discussing statistical methodology in the context of treatment sequences. This led to our focus on statistical methods for estimating treatment sequence effectiveness in subsequent sections. We encourage readers to refer to Figure 2 for a visual representation of our narrative synthesis results (Section 3.3) and the subsequent Discussion (Section 4).

3.3 Present landscape and challenges of treatment-sequencing evaluation in HTA

In this section, we provide a narrative overview of the current practice and primary challenges associated with treatment-sequencing in HTA across research themes, drawing from the selected studies.

3.3.1 HTA guidance for evaluating treatment sequences

Zheng et al. introduced a framework for treatment-sequencing modelling in 2017, providing step-by-step recommendations [11]. Derived from a systematic review of NICE TAs up to TA321 (2014), Zheng et al.'s framework comprises four steps: (1) model conceptualisation, (2) selecting suitable modelling approach, (3) considering appropriate data sources for model inputs and (4) determining computation tools. Zheng et al. indicated that treatment sequences have mainly been incorporated to reflect clinical practice or trial design, determine where to place new treatments in a sequence, or evaluate the impact of placing an additional treatment into existing treatment sequences. They noted an uptrend in TAs involving treatment sequences across various disease areas, including oncology, autoimmune, cardiovascular, neurology/mental health, infectious disease, and diabetes, with oncology featuring the highest number of treatment-sequencing models. A similar pattern was observed in an updated review by Viola et al. extending to TA527 (2018) [12].

Lewis et al. [28, 53] and Huang et al. [55] advanced Zheng et al.'s framework by further unpacking the complexities in determining the effectiveness of treatment sequences and suggesting potential solutions. Lewis et al. underscored the importance of leveraging quantitative evidence synthesis methods and scrutinising simplifying

assumptions [28, 53]. Meanwhile, Huang et al. identified specific challenges inherent to oncology treatment-sequencing and proposed solutions to improve the estimation of treatment sequence effectiveness.

Faria's review offered guidance on problem structuring for economic evaluation [16], with a section highlighting the importance of incorporating relevant sequences in the identification of pertinent decision options. Faria noted challenges in parameterising sequences in economic evaluation due to limited direct evidence on the effectiveness of treatments across all positions of sequences.

3.3.2 Scarce clinical evidence for treatment sequences and necessary statistical adjustments in evidence synthesis

Several studies highlighted the challenges of data scarcity in evaluating the effectiveness of treatment sequences [16, 28, 53, 55]. Such limitations largely stem from the lack of randomised controlled trials (RCTs) comparing treatment sequences, frequently necessitating the coalescence of evidence of discrete treatment effects (i.e. line-of-therapy (LOT)-specific effect) from different sources.

Lewis et al. identified that while network meta-analysis/meta-analysis (NMA/MA) is frequently utilised in HTA for synthesising treatment effectiveness evidence, there existed a gap in HTA guidance when it comes to adapting these methods for evaluating treatment sequences [53]. They highlighted that the feasibility of conducting an NMA/MA with RCTs comparing treatment sequences was hindered due to the lack of such trials. Furthermore, while it might be feasible to conduct an NMA/MA using observational studies that compare treatment sequences, this approach's viability could be considerably influenced by the inherent biases of observational studies [53]. Driven by these observations, Lewis et al. underscored the potential of adapting meta-analytic methods to reflect the "position effect" in deriving LOT-specific treatment effectiveness, including using meta-regression, stratification, or subgroup analyses in NMA/MA to account for the impact of patient's treatment history. Although Lewis et al.'s review of meta-analytic methods was non-disease-specific, only five out of 23 studies were cancer-related, while over half (n = 13) focused on RA.

Huang et al., conversely, outlined issues related to data scarcity in modelling oncology treatment sequences [55]. Key challenges include (1) strategies to adapt the effectiveness of a LOT based on its position within a sequence; (2) examining the interplay between the timing of progression, discontinuation of a LOT, and initiation of the subsequent LOT, and their impact on the modelling results; and (3) the availability of head-to-head comparisons evidence for a specific LOT or the entire treatment sequence. Huang et al. reviewed 46 oncology treatment-sequencing models from NICE TAs and PubMed to determine how the aforementioned challenges have been tackled. They stressed the need

to include both the time to treatment discontinuation and disease progression for LOT-effect in a treatment-sequencing model, or provide justification for any exceptions. Furthermore, Huang et al. believed treatment-free gaps should be included unless inapplicable, such as in late-stage cancers with minimal treatment gaps.

Huang et al. noted the prevalent practice of combining data from multiple clinical trials across different LOTs to model the effectiveness of a treatment sequence (41 models (89%)), such as merging progression-free survival (PFS) from an earlier LOT trial (e.g. PFS of treatment A in a first-line setting) and OS from a later LOT trial (e.g. OS of treatment B at in second-line setting) to model the OS for an entire treatment sequence (e.g. OS of treatment sequence A→B). However, none of these "mergers of evidence" adjusted the estimated effectiveness based on the characteristics of patients from different data sources (i.e., the misalignment of patient characteristics between LOTs), primarily due to the absence of individual patient-level data. In contrast, they identified three models estimated the OS of an entire treatment sequence using information from a single trial [57-59], and two models used RWD to estimate the OS of an entire treatment sequence [60, 61]. Given that RWE and single-trial evidence for treatment sequences are often unavailable by the time of the HTA submissions, Huang et al. recommended combining trial evidence from different LOTs with necessary adjustments as a more pragmatic approach [55].

Huang et al. found that no models applied head-to-head effectiveness evidence for comparing the entire treatment sequences, consistent with findings from Lewis et al [53]. Merely 24% of models applied methods to adjust for indirect treatment comparisons (ITC), which is exclusively limited to comparisons within a single LOT instead of the whole sequence. Huang et al. recommended making ITC adjustments for each LOT in the sequence and further adjusting for patient characteristics at the initiation of subsequent treatments, whenever possible. Overall, Huang et al. concluded that that there is substantial room for improvement in estimating treatment sequence effectiveness.

3.3.3 Simplifying assumptions in response to data scarcity

Despite Lewis et al. highlighting the use of advanced meta-analytic methods for generating LOT-specific evidence, many NMA/MA did not employ these methods [53]. Lewis et al. found that, such omissions often led to the need for additional simplifying assumptions when integrating such evidence into treatment-sequencing models [53], potentially causing biases and uncertainties. They summarised a taxonomy of these assumptions, such as whether a treatment's effect is dependent on its position or previous treatments and whether any modifications to the treatment effect should be made depending on its position or disease duration. Lewis et al. underscored the importance of carefully assessing the simplifying assumptions in sequencing models. We encourage readers to refer to Lewis et al.'s

study for details.

3.3.4 Modelling approaches

Decision-analytic modelling approaches for treatment sequencing are well-established [11]. Among the 63 treatment-sequencing models Zheng et al. analysed, cohort state-transition models (e.g. Markov models, Semi-Markov models) were most prevalent, followed by discrete event simulation (DES), individual state-transition models (e.g. microsimulations), and decision-trees. Several crucial factors can influence the choice of modelling approach and the construction of health states, such as patient heterogeneity, number of LOTs, and type of clinical outcomes [11]. Lewis et al. further detailed the advantages and disadvantages of each approach with a comparison table in their paper [53].

Although partitioned survival models are prevalent in oncology [62], Zheng et al. suggested that such models were not utilised for treatment sequences due to their inherent methodology and limitations [11]. Contrasting this, an updated review by Viola et al. highlighted several oncology NICE TAs employing partitioned survival models for treatment sequences [6]. Lewis et al. also identified comparable applications [53].

3.3.5 Selection of treatment sequences in a decision problem

Zheng et al. noted that most NICE TAs predominately compared treatment sequences that reflect clinical practice, rather than optimal placement or excluding suboptimal treatment sequences [11]. Viola et al. investigated the impact of baseline sequence-comparator selection on the incremental cost-effectiveness results by conducting a cost-effectiveness case study with a treatment-sequencing Markov model featuring four hypothetical treatments [12]. Their findings revealed that non-cost-effective treatments were never part of an optimal (i.e. most cost-effective) sequence. Viola et al. proposed excluding non-cost-effective treatments from the baseline comparator or placing them later in the sequence to improve health resource allocation efficiency. However, they also acknowledged the limited generalisability of their case study due to the number of treatments and length of the sequences included. Furthermore, their focus was on maximising resource allocation efficiency without posing any constraints on the positions of treatments in a sequence, which could occur in real-world practice due to marketing authorisations. Nevertheless, Viola et al. underscored the need for judiciously selecting baseline sequencing comparators. They cautioned that overlooking the cost-effectiveness of individual treatments when defining the baseline sequencing-comparator could lead to misleading results [12].

3.3.6 Impact of treatment displacement on resource allocation efficiency and equity

Haywood's doctoral thesis delved into the issue of treatment displacement. He defined treatment displacement

as where new treatments cause existing treatments to be shifted to later LOTs [27]. If not properly addressed in economic evaluations, Haywood argues that this can result in biased decision-making. Specifically, Haywood argued that continuing to pay the previously agreed-upon price for displaced treatments is unlikely to prove cost-effective. This is primarily due to the uncertainty surrounding the effectiveness of these treatments once they are displaced and used at a later line of therapy, with a high likelihood that effectiveness will be reduced. Consequently, paying the previously agreed-upon price of displaced treatments when they are used at a later LOT can lead to allocative inefficiency. Given the lack of clarity on reduced effectiveness, Haywood proposed adjusting the price of displaced treatments downwards to better reflect their anticipated diminished effectiveness and restore their cost-effectiveness.

To support his argument, Haywood conducted modelling studies to assess the impact of treatment displacement in several cancers [27], including breast, colorectal, and non-small cell lung cancer. The scarcity of clinical evidence on sequencing led him to undertake a *de novo* RWD analysis using Australian Pharmaceutical Benefits Scheme (PBS) data to understand local treatment patterns, subsequently guiding what treatments and the number of LOTs to be incorporated into his economic models and their associated costs. However, for LOT-specific effectiveness estimates (i.e., PFS), Haywood chose to apply estimates from literature and perform meta-analysis instead of conducting RWD analysis. Haywood's systematic review and meta-analyses of existing evidence revealed that the displacement of a treatment may lead to decreased effectiveness, treatment duration, and increased toxicity per unit time. Haywood's economic modelling results showed that displacement of an existing treatment resulted in an increased incremental cost-effectiveness ratio (ICER), with a reduction in the price of these treatments being necessary to restore the cost-effectiveness. Overall, he recommended carefully considering the impact of displacement on currently subsidised treatments in cancer treatment funding to ensure equity and cost-effectiveness to avoid underestimating the total costs and overestimating the total benefits of a new treatment being introduced to clinical practice [27].

3.3.7 Model computation and tools

Tosh and Kim's theses were among the earliest works on innovative treatment-sequencing methodologies [29, 30]. Both emphasised the significance of computational optimisation methods for sequencing-models, particularly in scenarios where the decision problem involves a large number of sequencing-comparators ($n > 1000$). Tosh investigated the use of simulation optimisation methods to identify optimal or near-optimal RA treatment sequences, driven by the inconsistencies in existing economic models for comparing RA treatment sequences [3], which may lead to potentially inaccurate cost-effectiveness estimates. He approached the issue as a combinatorial discrete simulation

optimisation problem, experimenting with methods such as simulated annealing and genetic algorithms. While Tosh found the simulation optimisation via simulated annealing (SOSA) promising, its time-consuming nature prompted calls for further research on its generalisability. Concurrently, Kim studied similar methods for treatment sequences in primary hypertension. The nuance in Kim's model lay in its emphasis on individualised treatment sequences, where each patient in the model could potentially receive the same "treatment strategy" but with different treatment sequences. For instance, patients exceeding a certain threshold for a particular characteristic would follow one sequence, while those who fell below the threshold would follow a different one. This contrasted with Tosh's approach, which examined the average effect of each treatment sequences on the entire population.

Zheng et al.'s review [11] summarised software that have been used in modelling treatment sequences, including those in Excel VBA, R, Arena and C. While both Tosh and Kim identified VBA and TreeAge as options for modelling treatment sequences, they opted for Simul8 and Matlab in their respective doctoral these [29, 30].

3.3.8 Potential of RWE in informing sequencing analyses

RWD has emerged as a valuable tool in supporting treatment-sequencing decisions [11, 12, 27, 53, 55]. Evidence derived from RWD (i.e., RWE) may be used to inform local treatment patterns, sequence selection, and shape model structures [11, 12, 27, 53, 55]. RWE can also be leveraged to capture the effectiveness of the entire treatment sequences or LOT-specific effectiveness [39, 42]. There are, however, significant challenges that undermine the potential of such application. One primary limitation is the delay in RWD's availability [11, 53, 55]. Typically, data collection commences only after a new drug is introduced to the market, leading to delays in its inclusion in RWD to be used for HTA submissions. Additionally, studies warned of the inherent issues in RWD that could result in biased findings (e.g. selection bias, confounding) [53] and underscored the importance of applying statistical methods to mitigate biases in RWD analyses [27, 53]. Despite these challenges, studies advocated for a further understanding of the pros and cons of RWE in treatment sequencing decisions in HTA [53, 63].

A recent review by Simpson et al [56] highlighted the latest advances in using RWE to evaluate sequencing in HTA. They spotlighted Spelman et al.'s approach of harnessing RWD from multiple countries to tackle the issue of the unmeasured confounding in treatment-sequencing comparative effectiveness studies [64]. Spelman et al.'s study focused on evaluating different treatment schedules for relapsing-remitting multiple sclerosis (RRMS). It is highlighted that, even when earlier observational studies compare "similar" patients based on their measured characteristics [65-67], unmeasured confounders may still exist between patients receiving different treatment patterns

in a single health care setting. Spelman et al., conversely, compared patients from Sweden and Denmark, where the treatment options of RRMS patients were comparable but the primary recommended treatment strategy diverged (starting with highly effective disease-modifying therapies versus treatment escalation, respectively) [64]. By controlling for various patient characteristics and comparing the clinical outcomes between the two countries, they proxied the impact of different treatment strategies. This method resembles using geography as an instrumental variable, which has faced criticism in oncology [68]. Spelman et al., nonetheless, introduced an innovative approach to comparative effectiveness research [64], which could prove valuable for future HTA submissions involving treatment sequences. Spelman et al.'s study was not directly included in our review because it focused solely on clinical findings, while Simpson et al.'s review [56] shed light on its relevance to HTA.

3.3.9 Relevance of causal inference methods in sequencing

A recent NICE DSU report [54] critically reviewed methods for evidence synthesis on clinical effectiveness in HTA, updating methodologies developed since the 2013 NICE guidance [13]. The report by Welton et al. features a section underscoring the potential of causal inference methods, especially in addressing the issue of unwanted treatment-switching in clinical trials, referencing a prior NICE DSU technical support document (TSD) (TSD 16) [69]. Welton et al. highlight the significance of applying causal inference tools in HTA to delineate the disease-treatment pathway, facilitating covariate selection within statistical methods for synthesising evidence of treatment effectiveness.

We included Welton et al.'s report for its insight into the parallels between treatment-switching in trials and treatment sequences, suggesting a new research direction. Welton et al. proposed that statistical methods capable of handling the dynamic treatment changes in RCTs (regardless of whether unintended or permitted) could potentially offer insights into treatment-sequencing. Thus, they recommend joint research to co-develop these methods in tandem with treatment-sequencing modelling. We interpreted this as an opportunity to leverage causal inference principles in developing statistical methods to estimate sequencing effects using RCT data containing treatment-switching, in line with the heavy reliance on these principles in the NICE DSU TSD 16 guidance [69]. While recognising the value of causal inference methods, our review excluded several epidemiological and statistical studies on dynamic or sequential treatment strategies due to their non-HTA focus [44, 45, 51]. However, these studies may provide further insights into evaluating sequencing effects using causal inference methods, warranting the need for a separate, in-depth review. To highlight this point, we retain Welton et al.'s paper in our roadmap.

4. Discussion

Our review underscores the escalating interest in evaluating treatment sequences in HTA and summarises the emerging research themes (Table 3, Figure 2). These themes have arisen in response to the multifaceted challenges of treatment sequences, encompassing conceptualising the decision problem, choosing suitable approaches for constructing treatment-sequencing decision models, identifying and deriving appropriate clinical effectiveness evidence for treatment sequences in the face of data scarcity, scrutinising effectiveness assumptions and statistical methods for adjustments, and optimising model computations. We found that interdisciplinary research offers promising solutions to these challenges. For example, methods from statistical and epidemiological research were utilised to handle evidence from non-randomised data and perform adjusted indirect comparisons [53, 55, 70, 71], and methods from operational research and computer science were applied for model optimisation [29, 30].

4.1 Implications of existing treatment-sequencing research for HTA

Early interest in treatment sequencing effects in specific diseases [17, 18, 72-74] led to NICE including advice on integrating sequential treatments into comparators in their 2013 methods guide [13, 17, 18]. Before Zheng et al. introduced the non-disease specific treatment-sequencing modelling framework [11], modelling guidance provided only general considerations about incorporating subsequent treatments [13, 15, 75-77]. Earlier sequencing-modelling frameworks were either only presented as conference abstracts [52] or had a broader focus on event sequencing beyond treatments, such as the sequencing of diagnostic procedures or whole disease modelling [36, 37]. The recently updated 2022 NICE manual highlights the importance of "care pathways" [14], covering the sequence of treatments, tests, and other relevant technologies. In this revision, NICE explicitly requires including all diagnostic technologies in a sequence, but does not stipulate the same for "treatment" sequences. Further, NICE is piloting the Pathway Approach to streamline the review of treatments for the same diseases, potentially impacting future HTA involving treatment sequences [19].

In estimating the effectiveness of treatment sequences, existing studies primarily recommend refining the derivation of LOT-specific effects and collating evidence from various LOTs, given the difficulties in acquiring data for entire sequences during the appraisal process [53, 55]. Such approach hinges on leveraging statistical methods to make adjustments for position effects, and to approximate head-to-head comparisons when collating evidence. Lewis et al.'s findings on adapting meta-analytic methods to account for position effects [53] seem relevant to NICE's guidance for meta-analytic methods (e.g. NICE DSU TSDs 1, 3, and 4) [78-80], while Huang et al.'s recommendations

[55] seem to relate closely to NICE DSU TSD 18 [70], focusing on population-adjusted indirect comparisons (i.e., matching-adjusted indirect comparison (MAIC) and simulated treatment comparisons (STC)). However, none of these guidelines discuss these similarly rooted methods in the context of treatment sequence evaluation. Furthermore, Huang et al.'s work seems to be anchored to state-transition models, raising questions about the applicability of their insights to partitioned survival models [81], which are prevalent in oncology. Importantly, it remains unclear whether there are other challenges unique to partitioned survival models yet to be identified. For example, "shoehorning" a sequence into a partitioned survival model might require alternative methods to collate evidence from multiple LOTs. For instance, combining survival curves from varied sources for different LOTs into a single OS curve for a sequence can lead to problems, such as unrealistic crossing of cumulative treatment durations and OS curves.

While none of the studies we examined focused on specific methods for generating RWE on treatment sequence effectiveness, several emphasised the importance of applying statistical methods to mitigate biases in RWD analyses [27, 53]. These insights mirror those in NICE DSU TSD 17 about using observational data to estimate treatment effectiveness [71]. However, TSD 17 does not touch on the relevance of these techniques in the context of treatment sequences. Zheng et al. underscored RWE's role beyond deriving effectiveness model inputs—also in corroborating model predictions and assessing uncertainties through sensitivity and scenario analyses. This aligns with the concept of evidence triangulation, prevalent in fields like epidemiology [82, 83]. While evidence triangulation has its merits, its utility as a validation tool might be constrained without clear acceptability criteria. Further, it is ambiguous whether Zheng et al. also implied using RWE as targets for model calibration [84]. Should this be the case, the uncertainty inherent in sequencing models could be exacerbated, especially when RWE—known for its susceptibility to confounding bias—is not validated. Specifically, a recent study highlights that the use of calibration targets that are not well-matched between models and data can result in biased outcomes [85].

In summary, for sequencing effect estimation, we believe it is crucial to incorporate relevant statistical methods for indirect comparisons, evaluating position or sequence effects, and adjusting for potential confounding and selection biases. Further reviews and research are needed to understand the strengths and weaknesses of available statistical methods in the context of treatment-sequencing. Moreover, understanding how these methods should be tailored for varying data sources, model types and outcomes requires further exploration. We therefore provide actionable directions for future research in Sections 4.2.1 to 4.2.5.

Determining what sequences to incorporate into a decision problem necessitates striking a balance among

several factors [12], including conducting an exhaustive review of treatment sequencing clinical evidence [28], considering local clinical guidelines [11], taking into account real-world practice and marketing authorisations [11, 16, 27], optimising resource allocation efficiency [12, 27], and computational feasibility [29, 30]. While Viola et al. recommend removing non-cost-effective treatments from a baseline sequencing-comparator to maximise resource allocation efficiency [12], it is noteworthy that NICE typically defines its comparator of interest as the most commonly prescribed treatment sequence rather than the "most cost-effective standard treatment". Lewis raised concerns about the NICE Committee's approach, noting that the their decisions frequently stem from deliberations without comprehensive review of clinical evidence on sequencing [28]. Furthermore, while the majority of appraisals assess fewer than 10 treatment sequences, Tosh and Kim examined exceptions with significantly more sequences [29, 30]. The impact of computational feasibility on sequence inclusion remains unknown. To date, there are no explicit guidelines prioritising any specific factor for sequence inclusion.

Regarding the impact of displaced treatments, Haywood's proposal to adjust their costs may be contentious [27]. Displaced treatments typically have shorter durations and reduced effectiveness [27], leading to inherent cost reduction. Hence, implementing additional price cuts could potentially over-penalise. Secondly, reducing the price of displaced treatments in fast-evolving disease areas may raise fairness concerns and discourage new treatment development, especially for late-stage treatments. In particular, new treatments may emerge quickly, so that the "less new ones" are displaced within a short time. Such dynamic might become apparent through initiatives like NICE's Pathway Project [19]. Thirdly, having treatments with varying prices at different LOTs presents complexities. A potential approach for addressing this issue is to implement mandatory pre-planned re-evaluation when displacement occurs and treatments begin to be used at a later LOT. This could lead to price adjustments where necessary. The potential role of prospective RWE in monitoring the treatment effectiveness and displacement has not been explicitly discussed. These normative issues extend greatly beyond economic evaluation and warrant further discussion.

4.2 Future research recommendation

4.2.1 Exploring consideration of treatment sequences in HTA beyond sequencing-models

The extent to which treatment-sequencing comparisons that do not necessitate explicit sequencing model structures remains uncertain as most existing reviews are geared towards treatment-sequencing economic models [11, 12, 53, 55]. For example, the manufacturer in NICE TA387 [58] used the OS of abiraterone from the COU-AA-302 trial [86] (abiraterone as a first-line treatment in metastatic castration-resistant prostate cancer) to represent the OS of

an entire treatment sequence (abiraterone → docetaxel → best supportive care) in the model, while an underlying three-line model was maintained for the purpose of calculating the costs of subsequent treatments. Therefore, technically, no conventional sequencing structure was required to model the OS. This approach heavily relies on the assumption that the subsequent treatments received by patients in the COU-AA-302 trial are representative of those in England's clinical practice.

Conversely, some HTAs may unintentionally compare the effects of treatment sequences without a sequencing model because of utilising effectiveness data that inherently include sequencing details. For example, when the aim is to compare treatments as if no patients had undergone any unintended subsequent treatment, it becomes problematic to use RCT data with unintended treatment-switching without making appropriate adjustments [69, 87]. Exploring the prevalence of these approaches in HTA across disease types is warranted, as they can affect the necessary assumptions required in populating the treatment effectiveness in economic models. While biases in the unadjusted unintended treatment-switching scenario are well-recognised [69], uncertainties in the first scenario—where the effect of a first-line treatment is assumed to represent the effect of the entire treatment sequence (e.g. TA387 [58])—may have been overlooked, warranting further investigation.

4.2.2 Further research on the derivation of clinical effectiveness for treatment sequences

Data scarcity remains a significant challenge when evaluating treatment sequences in HTA [11, 53, 55, 56]. Despite existing studies outlining potential methods for deriving treatment sequence effectiveness [53, 55], there may be a need for an updated overview of the current practice. The meta-analytic methods summarised by Lewis et al. appear to be a viable approach for deriving LOT-specific effectiveness [53]. However, their recommendation primarily stems from RA studies before 2013 and may not be as applicable in oncology due to the complexities of survival meta-analysis [88, 89]. Huang et al., conversely, exclusively focused on oncology studies and recommended applying ITC adjustment methods when combining data from various trials [55].

Both Lewis et al. and Huang et al. focus on methods for deriving LOT-specific effectiveness [53, 55], but these approaches heavily rely on simplifying assumptions about effectiveness and population alignment between LOTs. Exploring advanced statistical methods to refine the "degradation effect" parameter (a prevalent strategy in autoimmune disease TAs [90-93]) could potentially mitigate uncertainties stemming from oversimplified assumptions. This approach could be particularly useful for sequences with a significantly higher number of LOTs, where

determining LOT-specific effects for each LOT is challenging (e.g. autoimmune HTAs). This approach is suitable where a degradation effect assumption can be made, but less applicable in situations where subsequent treatments may be more effective due to specific prior events, such as treatment-induced mutations that enhance the effectiveness of later therapies. Furthermore, exploring alternative data sources and statistical methodologies to assess the comparative effectiveness of complete treatment sequences remains an underexplored area, and could prove valuable. Reviewing statistical methods originally developed for other purposes but are potentially applicable to treatment sequence evaluation (e.g. those adapted for tackling treatment-switching in RCTs [54, 69]) from both HTA and non-HTA domains, could inspire new strategies to address sequencing challenges in HTA.

4.2.3 The role of RWE and causal inference in assessing treatment sequence effectiveness

Despite the limited RWD availability for new treatments during appraisals, considerable research emphasises the significance of exploring RWE's potential in informing treatment-sequencing effectiveness [11, 27, 53, 55, 56]. RWD holds importance because of its ability to capture patient treatment trajectories over time. By leveraging causal inference principles, biases in RWD analysis could be mitigated [71, 94, 95]. While existing HTA guidelines provide methods suitable for non-sequencing scenarios [54, 71], tailored modifications may be required for treatment-sequencing. For instance, assessing the comparative effectiveness of two complete treatment sequences is complex due to time-varying confounding, potentially necessitating advanced adjustment techniques, such as g-methods [96, 97]. Statistical methods highlighted by Lewis et al. and Huang et al. implicitly tap into the importance of causal inference [53, 55]—for instance, adjusted ITC is used to ensure fair comparisons between two groups. This aligns with HTA's goal to compare the counterfactual outcomes of alternative interventions: one without the new treatment and one with it in the healthcare system. Given that the concept of counterfactual outcomes resonate with causal inference principles, we advocate for their explicit integration into treatment-sequencing HTA, especially when utilising RWE. This perspective largely mirrors the views in the recently released NICE RWE framework [20].

We identified several niches where RWE can support treatment-sequencing HTA. Each necessitates tailored study designs and statistical approaches to address biases arising from non-randomised treatments. These areas include: (1) deriving the effectiveness of a specific LOT or a segment of the treatment sequence, whether in the control or treatment arm; (2) deriving the overall effectiveness of the treatment sequence in the control arm; (3) deriving the comparative effectiveness between two (or more) complete treatment sequences.

Application (1) is commonly used to populate the effectiveness of later-line treatments in the model when trial

evidence is constrained by short follow-up durations. Conceptually, methodologies described by Lewis et al. and Huang et al. can be considered to address indirect comparisons and ensure population alignment across LOTs [53, 55]. However, further nuances and challenges of "plugging-in" RWE into a treatment sequence assessment warrant further exploration. Further, this form of RWE has also been employed as calibration targets for modelling early detection interventions where evidence from long-term follow-up is scarce [98]. Biases can, however, arise if a mismatch between the model and data exists [85]. Additionally, this type of RWE can aid in evidence triangulation in oncology, particularly when scrutinising the plausibility of the extrapolated survival curve of the appraised treatment beyond trial periods. Comparing the late-stage survival curve from RWE against trial extrapolations can reveal any unrealistic extrapolations, given that such extrapolations are greatly influenced by the choice of parametric models [99, 100]. However, a caveat is that estimates derived from RWE on late-stage survival might underestimate the effects of the appraised treatment, since this evidence often comes from data gathered before the introduction of newer frontline treatments. Recent studies have explored integrating RWE into trial-based survival extrapolations [101]. Future work in this field may benefit from delving deeper into the nuances of treatment sequencing for such applications.

Application (2) represents a unique case of an external (synthetic) control arm, which received considerable attention in HTA and regulatory bodies [102, 103], including the NICE RWE framework [20], even though none of these have specifically addressed the treatment-sequencing scenario. This application's complexity arises from the need to blend data from multiple sources (typically RWD with a trial arm) and employing statistical methods to enable direct comparisons of complete treatment sequences. While the statistical methods may conceptually resemble those used for RWD analysis, adaptions for extending inferences may be necessary [104, 105]. Predictable challenges include unmeasured confounding, the absence of time-varying covariates, and missing information on subsequent treatments in trials. While it seems conceptually viable, a renewed review may be needed to identify existing application and associated challenges, and further application studies are imperative to assess such application's feasibility.

Application (3) is specifically pertinent in the re-evaluation of appraisals or reimbursement decisions for treatments previously funded through alternative funds (e.g. NICE Cancer Drugs Fund (CDF)) [1, 106, 107]. In such instances, data on previously appraised treatments might become available, presenting an opportunity to harness RWE for direct comparisons of treatment sequences. This is especially relevant in determining the optimal sequencing or positioning of treatments. Notably, clinical trials rarely assess the effectiveness of a new treatment across various

treatment lines (e.g., standard of care → new treatment versus new treatment → standard of care), whereas such sequences can co-exist in real-world settings and may serve as relevant comparators in an appraisal.

The Target Trial emulation (TTE) approach in epidemiology [63] seems to be a promising tool for comparing two complete treatment sequences in RWD [108]. The TTE framework [63], rooted in the principle of causal inference, has garnered substantial attention for its potential in explicitly designing observational studies to estimates causal effects and mitigate issues such as confounding, selection bias, and immortal time bias. The U.S. Food and Drug Administration (FDA) has funded projects aimed at examining the Target Trial emulation's feasibility to answer clinical questions for regulatory purposes [109, 110], and the NICE RWE framework endorsed the TTE framework and recommends application in HTA wherever relevant [20].

NICE's RWE framework briefly mentioned the viability of using emulated Target Trial to compare dynamic treatment strategies, such as treatment sequences, without specifying operational details. In our view, treatment sequences can be categorised into different types of treatment strategies [95]: (1) time-related static treatment strategies (where patients transition to the next-line of treatment at fixed intervals) [111]; (2) dynamic treatment strategies (where the timing of patients transitioning to the next-line of treatment is based on specific events (e.g. disease progression) [112, 113], with each patient assigned to a specific type of treatment sequence within each treatment arm); and (3) dynamic treatment strategies with individualised treatment sequences (where the timing of patients transitioning to the next-line of treatment is based on patient characteristics, with the possibility of each patient being assigned to different a treatment sequence based on specific events and additional patient characteristics (e.g. biomarker status)) [114]. Each of these strategies may require overlapping but distinct analytical approaches. Additionally, instead of emulating Target Trials for the comparison of dynamic treatment strategies, which are more relevant to Applications 2 and 3, a recent clinical epidemiology study by Bujkiewicz et al. performed a bivariate NMA on line-specific emulated Target Trials from the British RA Register [115]. This approach aimed to bridge the disconnected networks in NMAs when assessing the effectiveness of first and second-line RA therapies, contributing to improving insights for Application 1.

Given the complexities mentioned above, further studies are, therefore, needed to assess the feasibility and validity of utilising Target Trial emulation to compare treatment sequences suitable for HTA purposes under different scenarios [108]. This also echoes a recent review that underscores the need for future research on the application of RWE with Target Trial emulation in HTA re-evaluation [116].

4.2.4 Role of other sources of data harbouring treatment sequencing information

Apart from RWD, other data sources may also harbour comprehensive sequencing information but remain underexplored in prior research. These include (1) trials (RCTs or single-arm trials) with extended-follow up and collect information on non-randomised subsequent treatments; (2) RCTs that randomise patients to receive different treatment sequences; (3) sequential multiple assignment randomised trials (SMARTs) [51, 114], which involve multiple stages of randomisation at points when treatment changes are indicated, unlike conventional trials where patient randomisation occurs only at baseline. Applying estimates from source (2) is relatively straightforward and requires no adjustment if the treatment sequences being compared align with those in a specific decision problem, although such alignment is uncommon. Utilising estimates from source (3), however, necessitates the use of specialised analytical techniques specific to SMARTs [44, 117]. Jamie Robins, known for his contribution in causal inference methods, noted that the key difference between dynamic treatments in real-world settings and SMARTs lies in the randomisation probabilities—unknown in the former while predetermined in the latter [118]. Statistical methods developed for SMARTs share similarities with causal inference methods for real-world dynamic treatments (e.g. g-methods, not limited to sequential treatments), and could be relevant for both assessing treatment sequences in RWD and trials that include sequencing data. Utilising data from source (1) may also require causal inference-guided statistical methods to mitigate confounding bias [54], and facilitate adjustments for indirect comparisons when integrating it with other data sources. Given its prevalence over sources (2) and (3), source (1) holds substantial promise for offering timely insights into treatment sequencing involving the appraised treatment during an appraisal, setting it apart from RWD's role. Thus, methods for utilising data from source (1) for HTA should be investigated.

On a separate note, recent advancements in adaptive trial designs could also potentially be adapted to improve the evaluation of treatment sequence effectiveness [119-122]. While SMARTs focus on developing pre-defined multi-stage adaptive interventions, the design of adaptive trials modify trials over time based on interim data. With the design of dynamic protocols, advancements in adaptive trials may provide transferable statistical methods for assessing sequencing effectiveness. Specifically, adaptive trials differ from traditional ones by continuously adjusting randomisation probabilities over time. Informally, these (time-varying) adapted randomisation probabilities sit between the known probabilities of receiving treatments in SMARTs and the unknown probabilities in RWD analysis. Thus, statistical methods for analysing adaptive trials theoretically share similarities with those previously mentioned for SMARTs and may offer additional insights in addressing challenges related to sequence analysis.

4.2.5 Model computation, utility, costs and equity

While the potential of computational optimisation methods for treatment sequencing decisions attracted early attention [29, 30], no further developments have been seen in recent years. This may be associated with a tendency for HTA to focus on individual treatment effects as oppose to optimising treatment sequences. Further, these early studies were hampered by inadequate evidence on LOT-specific effects [29, 30], necessitating the use of crude simplifying assumptions that undermined the value of these methods. Nonetheless, advancements in statistical methods to better identify these effects may improve the potential value of these computational methods (see Section 4.2.2). A thorough review of recent appraisals might shed light on the recent use of these methods. Although not widely incorporated into routine HTA processes, relevant codes developed by Tosh and Kim are available online for adaption [29, 30]. Notably, methods developed in Tosh and Kim's theses were highlighted by The Professional Society for Health Economics and Outcomes Research (ISPOR) Task Force as exemplars when a decision problem is framed as a constrained optimisation problem that accounts for a set of decisions over time [123].

Finally, several other areas have received limited research attention but may be crucial for further exploration, including equity, utility, and costs within treatment sequence evaluation. Utility, in particular, may encounter similar data scarcity issues as treatment sequences since patient data at specific time points while on a certain treatment is often lacking.

4.3 Strengths and limitations of the review

Our study has several strengths. Firstly, our review provides a comprehensive summary of studies focusing on treatment sequencing within HTA, highlighting both challenges and advancements, including early-stage developments from doctoral theses. The compilation reveals emerging research areas and underscores the value of interdisciplinary research. Secondly, our roadmap elucidates how these studies interrelate, enabling researchers with specific interests to swiftly locate pertinent papers and associated references. Thirdly, we bridged the gap between the insights drawn from the selected studies and their relevance to existing HTA guidance [70, 71, 78-80] as well as the newly issued NICE RWE framework [20]. This highlights current gaps in HTA guidelines pertaining to tailored methods for treatment-sequencing, and offers actionable research areas for enhancement, aiming to complement the existing guidelines.

Despite its strengths, our study also comes with limitations. First, by focusing on reviews, methodological, and conceptual papers, the scope of our review may be deemed as limited. However, given the frequent citations to Zheng

et al.'s work [11] across the majority of the identified literature, we are confident that we have captured the field's pivotal paper and its associated literature. Secondly, with our non-disease-specific and HTA-oriented focus, we might miss other relevant methodological papers that brings additional insights. However, for readers interested in specific modelling studies or innovative methods from other fields, these can likely be found in the references of the papers we included. Thirdly, we may not fully capture all disease-specific nuances of their interplay with each of the research topics. Readers are encouraged to further explore the implications of subject knowledge. Lastly, we may have missed articles discussing treatment sequences if they did not directly mention such concepts in their titles or abstracts.

5. Conclusion

Our research highlights the multifaceted challenges in evaluating treatment sequences in HTA, including problem structuring, modelling approaches, selection of treatment-sequencing comparators, model optimisation and data scarcity in evaluating the effectiveness of treatment sequences. Each of these aspects presents complex, interrelated issues that warrant further investigation. This field is rapidly evolving, with the issue of data scarcity in treatment sequences being particularly salient, causing decision uncertainties, arising from inherent biases in data sources, effectiveness simplifying assumptions, and the appropriateness of the adjustment methods utilised. We found that interdisciplinary research offers promising solutions to these challenges. This includes applying causal inference principles from epidemiology and statistical research to manage evidence from non-randomised data and perform adjusted indirect comparisons. Additionally, insights from operational research and computer science contribute to model optimisation.

Several avenues for future research are identified. Firstly, given the remarkable potential of RWD in capturing treatment sequences, it is imperative to investigate appropriate strategies for leveraging RWE in treatment sequence evaluation. In particular, it is prudent to explore how emerging causal inference methods, such as Target Trial emulation, can be employed to harness RWD for generating comparative effectiveness of treatment sequences. This is especially pertinent in the context of re-evaluating treatment appraisals or synthesising an external treatment-sequencing control arm. Secondly, it is equally crucial to delve into the application of causal inference methods in analysing clinical trials that harbour sequencing information. This approach holds promise for providing timely estimations of the appraised treatment's effectiveness within a sequence, offering valuable insights during the appraisal process. Thirdly, under-researched areas should be explored, including the generalisability of model

optimisation methods, equity, utility, and costs in treatment sequence evaluation. Finally, an updated review of treatment-sequencing economic evaluation applications may provide valuable insights into the evolving field.

References

1. NICE. Position statement: consideration of products recommended for use in the Cancer Drugs Fund as comparators, or in a treatment sequence, in the appraisal of a new cancer product. Accessed: Mar 23, 2020. Available at: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal-guidance/cancer-drugs-fund/CDF-comparator-position-statement.pdf

2. Claxton K. Exploring uncertainty in cost-effectiveness analysis. Pharmacoeconomics. 2008; 26: 781-98.

3. Tosh J, Stevenson M, Akehurst R. Health Economic Modelling of Treatment Sequences for Rheumatoid Arthritis: A Systematic Review. Curr Rheumatol Rep. 2014; 16: 447.

4. Benedict A, Stevenson M, Sorensen S. ISPOR Europe 2017: Modeling Separate Lines of Treatment Versus Treatment Sequences in Cancer (Workshop 8). Accessed: June 20, 2020. Available at: https://www.ispor.org/docs/default-source/presentations/1221.pdf?sfvrsn=6f1f29a1_1

5. Al M, Jin H, Robins J, et al. Virtual ISPOR Europe 2020: Novel Methods for Modeling Treatment Sequencing (Spotlight session 3). Accessed: Nov 19, 2020. Available at: https://www.ispor.org/conferences-education/conferences/past-conferences/ispor-europe-2020/program/spotlight-sessions

6. Kaufman HL, Atkins MB, Subedi P, et al. The promise of Immuno-oncology: implications for defining the value of cancer treatment. Journal for immunotherapy of cancer. 2019; 7: 1-11.

7. Chehayeb RJ, Hood A, Wang X, et al. Treatment Sequencing Patterns and Associated Direct Medical Costs of Metastatic Breast Cancer Care in the United States, 2011 to 2021. JAMA Network Open. 2022; 5: e2244204-e04.

8. Roux J, Grimaud O, Leray E. Use of state sequence analysis for care pathway analysis: The example of multiple sclerosis. Stat Methods Med Res. 2019; 28: 1651-63.

9. Kim AH, Jang JE, Han J. Current status on the therapeutic strategies for heart failure and diabetic cardiomyopathy. Biomed Pharmacother. 2022; 145: 112463.

10. Kern DM, Cepeda MS, Defalco F, et al. Treatment patterns and sequences of pharmacotherapy for patients diagnosed with depression in the United States: 2014 through 2019. BMC Psychiatry. 2020; 20: 1-10.

11. Zheng Y, Pan F, Sorensen S. Modeling Treatment Sequences in Pharmacoeconomic Models. Pharmacoeconomics. 2017; 35: 15-24.

12. Viola MG, Diamantopoulos A. Determining the Baseline Strategy in a Cost-Effectiveness Analysis with Treatment Sequences. Appl Health Econ Health Policy. 2020; 18: 17-29.

13. NICE. NICE Process and methods: Guide to the methods of technology appraisal 2013 (PMG 9). Accessed: June 2, 2020. Available at: https://www.nice.org.uk/process/pmg9/chapter/foreword

14. NICE. NICE health technology evaluations: the manual (PMG 36). Accessed: Aug 17, 2023. Available at: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation

15. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies: Canada — 4th Edition. Accessed: March 11, 2023. Available at: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf

16. Faria R. Problem Structuring in Economic Evaluation. In: Oxford Research Encyclopedia of Economics and Finance. Oxford University Press, 2021.

17. NICE. Briefing paper for methods review working party on choosing comparators. Accessed: Jan 26, 2023. Available at: https://www.sheffield.ac.uk/media/34080/download?attachment

18. NICE. Briefing paper for methods review working party on treatment sequences and downstream costs Accessed: Jan 26, 2023. Available at: https://www.sheffield.ac.uk/media/34080/download?attachment

19. NICE. Taking a proportionate approach to technology appraisals. Accessed: Jan 26, 2024. Available at: https://www.nice.org.uk/about/what-we-do/proportionate-approach-to-technology-appraisals

20. National Institute for Health and Care Excellence. NICE real-world evidence framework [ECD 9]. Accessed: June 30, 2023. Available at: https://www.nice.org.uk/corporate/ecd9/chapter/overview

21. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ. 2005; 331: 1064-65.

22. Schlosser RW, Wendt O, Bhavnani S, et al. Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review. Int J Lang Commun Disord. 2006; 41: 567-82.

23. Hernandez L, O'Donnell M, Postma M. Modeling Approaches in Cost-Effectiveness Analysis of Disease-Modifying Therapies for Relapsing-Remitting Multiple Sclerosis: An Updated Systematic Review and Recommendations for Future Economic Evaluations. Pharmacoeconomics. 2018; 36: 1223-52.

24. Jansen JP, Incerti D, Mutebi A, et al. Cost-effectiveness of sequenced treatment of rheumatoid arthritis with targeted immune modulators. J Med Econ. 2017; 20: 703-14.

25. Badampudi D, Wohlin C, Petersen K. Experiences from using snowballing and database searches in systematic literature studies. Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering (EASE). Nanjing, China: ACM Digital Library, 2015.

26. Zwakman M, Verberne LM, Kars MC, et al. Introducing PALETTE: an iterative method for conducting a literature search for a review in palliative care. BMC Palliat Care. 2018; 17: 82.

27. Haywood P. Increasing complexity of cancer care: How displaced treatments impact efficiency, cost-effectiveness and equity [Doctoral Thesis]. University of Technology Sydney, 2018.

28. Lewis R. Quantitative evidence synthesis methods for the assessment of the effectiveness of treatment sequences for clinical and economic decision-making [Doctoral Thesis]. Bangor University, 2019.

29. Tosh J. Simulation optimisation to inform economic evaluations of sequential therapies for chronic conditions: a case study in rheumatoid arthritis [Doctoral Thesis]. University of Sheffield, 2015.

30. Kim E. Sequential drug decision problems in long-term medical conditions: a case study of primary hypertension [Doctoral Thesis]. University of Sheffield, 2015.

31. Hawkins N, Sculpher M, Epstein D. Cost-effectiveness analysis of treatments for chronic disease: using R to incorporate time dependency of treatment response. Med Decis Making. 2005; 25: 511-9.

32. Saramago P, Manca A, Sutton AJ. Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. Value Health. 2012; 15: 639-49.

33. Tordrup D, Chouaid C, Cuijpers P, et al. Priorities for Health Economic Methodological Research: Results of an Expert Consultation. Int J Technol Assess Health Care. 2017; 33: 609-19.

34. Heeg BMS. Developing health economic models of chronic diseases for reimbursement purposes [Doctoral Thesis]. University of Groningen: Groningen, The Netherlands. 2015.

35. Brixner DI, Watkins JB. Can CER be an effective tool for change in the development and assessment of new drugs and technologies? J Manag Care Pharm. 2012; 18: S06-S11.

36. Tappenden P, Chilcott J, Brennan A, et al. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. Value Health. 2012; 15: 1127-36.

37. Lord J, Willis S, Eatock J, et al. Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. Health Technol Assess. 2013; 17.

38. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006; 15: 1295-310.

39. Rauner MS, Gutjahr WJ, Heidenberger K, et al. Dynamic policy modeling for chronic diseases: metaheuristic-based identification of pareto-optimal screening strategies. Oper Res. 2010; 58: 1269-86.

40. Cressman S, Browman GP, Hoch JS, et al. A Time-Trend Economic Analysis of Cancer Drug Trials. Oncologist. 2015; 20: 729-36.

41. de Mello-Sampayo F. The timing and probability of treatment switch under cost uncertainty: an application to patients with gastrointestinal stromal tumor. Value Health. 2014; 17: 215-22.

42. Vernon JA, Hughen WK. A Primer on Dynamic Optimization and Optimal Control in Pharmacoeconomics. Value Health. 2006; 9: 106-13.

43. Chick SE, Inoue K. New two-stage and sequential procedures for selecting the best simulated system. Oper Res. 2001; 49: 732-43.

44. Murphy SA, Oslin DW, Rush AJ, et al. Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. Neuropsychopharmacology. 2007; 32: 257-62.

45. Cain LE, Saag MS, Petersen M, et al. Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. Int J Epidemiol. 2016; 45: 2038-49.

46. Messner DA. Evaluating the comparative effectiveness of treatment sequences in oncology: a new approach. Journal of Comparative Effectiveness Research. 2015; 4: 537-40.

47. Dekker J, Van HL, Hendriksen M, et al. What is the best sequential treatment strategy in the treatment of depression? Adding pharmacotherapy to psychotherapy or vice versa? Psychother Psychosom. 2013; 82: 89-98.

48. Weintraub M, Waldman E, Koplewitz B, et al. A sequential treatment algorithm for infants with stage 4s neuroblastoma and massive hepatomegaly. Pediatr Blood Cancer. 2012; 59: 182-84.

49. Trappe R, Oertel S, Leblond V, et al. Sequential treatment with rituximab followed by CHOP chemotherapy in adult B-cell post-transplant lymphoproliferative disorder (PTLD): the prospective international multicentre phase 2 PTLD-1 trial. The lancet oncology. 2012; 13: 196-206.

50. Grigor C, Capell H, Stirling A, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. The Lancet. 2004; 364: 263-69.

51. Wang L, Rotnitzky A, Lin X, et al. Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer. Journal of the American Statistical Association. 2012; 107: 493-508.

52. Briggs A, Sidhu M, Baker T. A framework for analysing treatment sequences: Incorporating time dependent transitions through partitioned survival analysis. Value Health. 2014; 17: A204-A05.

53. Lewis RA, Hughes D, Sutton AJ, et al. Quantitative Evidence Synthesis Methods for the Assessment of the Effectiveness of Treatment Sequences for Clinical and Economic Decision Making: A Review and Taxonomy of Simplifying Assumptions. Pharmacoeconomics. 2020: 1-37.

54. Welton N, Phillippo D, Owen R, et al. CHTE2020 sources and synthesis of evidence; update to evidence synthesis methods. Accessed: Aug 29, 2022. Available at: https://rees-france.com/wp-content/uploads/2020/12/CHTE-2020_synthesis-of-evidence.pdf

55. Huang M, Ramsey S, Xue W, et al. Conceptual Framework and Methodological Challenges for Modeling Effectiveness in Oncology Treatment Sequence Models. Pharmacoeconomics. 2021: 1-12.

56. Simpson A, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to HTA: part 5. Journal of Comparative Effectiveness Research. 2022; 11: 213-15.

57. NICE. Enzalutamide for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated (TA 377). Accessed: Apr 30, 2021. Available at: https://www.nice.org.uk/guidance/ta377

58. NICE. Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated (TA 387). Accessed: Apr 30, 2021. Available at: https://www.nice.org.uk/guidance/ta387

59. Pan F, Reifsnider O, Zheng Y, et al. Modeling clinical outcomes in prostate cancer: application and validation of the discrete event simulation approach. Value Health. 2018; 21: 416-22.

60. Parikh RC, Du XL, Robert MO, et al. Cost-effectiveness of treatment sequences of chemotherapies and targeted biologics for elderly metastatic colorectal cancer patients. Journal of Managed Care & Specialty Pharmacy. 2017; 23: 64-73.

61. Rognoni C, Ciani O, Sommariva S, et al. Cost–effectiveness analysis of treatments involving radioembolization in intermediate-stage hepatocellular carcinoma. Journal of Comparative Effectiveness Research. 2018; 7: 209-21.

62. Woods BS, Sideris E, Palmer S, et al. Partitioned survival and state transition models for healthcare decision making in oncology: where are we now? Value Health. 2020; 23: 1613-21.

63. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016; 183: 758-64.

64. Spelman T, Magyari M, Piehl F, et al. Treatment escalation vs immediate initiation of highly effective treatment for patients with relapsing-remitting multiple sclerosis: data from 2 different national strategies. JAMA neurology. 2021; 78: 1197-204.

65. Simonsen CS, Flemmen HØ, Broch L, et al. Early high efficacy treatment in multiple sclerosis is the best predictor of future disease activity over 1 and 2 years in a Norwegian population-based registry. Front Neurol. 2021; 12: 693017.

66. Brown JWL, Coles A, Horakova D, et al. Association of initial disease-modifying therapy with later conversion to secondary progressive multiple sclerosis. JAMA. 2019; 321: 175-87.

67. He A, Merkel B, Brown JWL, et al. Timing of high-efficacy therapy for multiple sclerosis: a retrospective observational cohort study. The Lancet Neurology. 2020; 19: 307-16.

68. Vertosick EA, Assel M, Vickers AJ. A systematic review of instrumental variable analyses using geographic region as an instrument. Cancer Epidemiol. 2017; 51: 49-55.

69. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. Accessed: June 2, 2020. Available at: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD16_Treatment_Switching.pdf

70. Phillippo D, Ades T, Dias S, et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. 2016.

71. Faria R, Hernadez Alava M, Manca A, et al. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data. Accessed: 2020, June 2. Available at: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD17-DSU-Observational-data-FINAL.pdf

72. National Institute for Clinical Excellence. Capecitabine and oxaliplatin in the adjuvant treatment of stage III (Dukes' C) colon cancer (Technology Appraisal 100). Accessed: Jun 30, 2020. Available at: https://www.nice.org.uk/guidance/ta100

73. Barton P, Jobanputra P, Wilson J, et al. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. Health Technol Assess. 2004; 8: iii, 1-91.

74. Tran-Duy A, Boonen A, van de Laar MA, et al. A discrete event modelling framework for simulation of long-term outcomes of sequential treatment strategies for ankylosing spondylitis. Ann Rheum Dis. 2011; 70: 2111-18.

75. Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes (4th ed.). Oxford: Oxford University Press, 2015.

76. Tappenden P. Conceptual modelling for health economic model development. Accessed: March 30, 2023. Available at: https://eprints.whiterose.ac.uk/74464/1/HEDSDP1205.pdf

77. Kaltenthaler E, Tappenden P, Paisley S, et al. NICE DSU Technical Support Document 13: Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. Accessed: March 30, 2023. Available at: https://www.ncbi.nlm.nih.gov/books/NBK425832/

78. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. Accessed: March 30, 2023. Available at: https://www.ncbi.nlm.nih.gov/books/NBK395870/pdf/Bookshelf_NBK395870.pdf

79. Dias S, Sutton AJ, Welton NJ, et al. NICE DSU Technical Support Document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Accessed: March 30, 2023. Available at: https://www.ncbi.nlm.nih.gov/books/NBK395886/pdf/Bookshelf_NBK395886.pdf

80. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. Accessed: March 30, 2023. Available at: https://www.ncbi.nlm.nih.gov/books/NBK310372/pdf/Bookshelf_NBK310372.pdf

81. Woods B, Sideris E, Palmer S, et al. NICE DSU Technical Support Document 19: partitioned survival analysis for decision modelling in health care: a critical review. Accessed: July 4, 2023. Available at: https://www.sheffield.ac.uk/media/34205/download?attachment

82. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2017; 45: 1866-86.

83. Hammerton G, Munafò MR. Causal inference with observational data: the need for triangulation of evidence. Psychol Med. 2021; 51: 563-78.

84. Vanni T, Karnon J, Madan J, et al. Calibrating Models in Economic Evaluation. Pharmacoeconomics. 2011; 29: 35-49.

85. Mandrik O, Thomas C, Whyte S, et al. Calibrating Natural History of Cancer Models in the Presence of Data Incompatibility: Problems and Solutions. Pharmacoeconomics. 2022; 40: 359-66.

86. Ryan CJ, Smith MR, Fizazi K, et al. Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naive men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. The Lancet Oncology. 2015; 16: 152-60.

87. Latimer N, White I, Tilling K, et al. Improved two-stage estimation to adjust for treatment switching in randomised trials: g-estimation to address time-dependent confounding. Stat Methods Med Res. 2020; 29: 2900-18.

88. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. BMC Med Res Methodol. 2012; 12: 1-16.

89. Jansen JP, Incerti D, Trikalinos TA. Multi-state network meta-analysis of progression and survival data. Stat Med. 2023.

90. NICE. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor (part review of NICE technology appraisal guidance 36, review of NICE technology appraisal guidance 126 and 141) (TA 195). Accessed: Mar 23, 2020. Available at: https://www.nice.org.uk/guidance/ta195

91. National Institute for Health and Care Excellence. Rituximab in combination with glucocorticoids for treating anti-neutrophil cytoplasmic antibody-associated vasculitis (Technology Appraisal 308). Accessed: Jun 30, 2020. Available at: https://www.nice.org.uk/guidance/ta308

92. National Institute for Health and Care Excellence. Apremilast for treating moderate to severe plaque psoriasis (Technology Appraisal 419). Accessed: Jun 30, 2020. Available at: https://www.nice.org.uk/guidance/ta419

93. National Institute for Health and Care Excellence. Ixekizumab for treating moderate to severe plaque psoriasis (Technology Appraisal 442). Accessed: Jun 30, 2020. Available at: https://www.nice.org.uk/guidance/ta442

94. Holland PW. Statistics and causal inference. Journal of the American Statistical Association. 1986; 81: 945-60.

95. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: FL. Chapman & Hall/CRC, 2020.

96. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling. 1986; 7: 1393-512.

97. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol. 2017; 46: 756-62.

98. Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. Med Decis Making. 2011; 31: 625-41.

99. Bell Gorrod H, Kearns B, Stevens J, et al. A Review of Survival Analysis Methods Used in NICE Technology Appraisals of Cancer Treatments: Consistency, Limitations, and Areas for Improvement. Med Decis Making. 2019; 39: 899-909.

100. Latimer NR, Adler AI. Extrapolation beyond the end of trials to estimate long term survival and cost effectiveness. BMJ medicine. 2022; 1.

101. Bullement A, Stevenson MD, Baio G, et al. A systematic review of methods to incorporate external evidence into trial-based survival extrapolations for health technology assessment. Med Decis Making. 2023: 0272989X231168618.

102. Jaksa A, Louder A, Maksymiuk C, et al. A Comparison of Seven Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies. Value Health. 2022; 25: 1967-76.

103. Burger HU, Gerlinger C, Harbron C, et al. The use of external controls: To what extent can it currently be recommended? Pharm Stat. 2021; 20: 1002-16.

104. Colnet B, Mayer I, Chen G, et al. Causal inference methods for combining randomized trials and observational studies: a review. arXiv preprint. Accessed: June 30, 2023. Available at: https://doi.org/10.48550/arXiv.2011.08047

105. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. Eur J Epidemiol. 2019; 34: 719-22.

106. NHS. Cancer Drugs Fund. Accessed: May 22, 2020. Available at: https://www.england.nhs.uk/cancer/cdf/

107. Kang J, Cairns J. "Don't think twice, It's all right": using additional data to reduce uncertainty regarding oncologic drugs provided through managed access agreements in England. PharmacoEconomics-Open. 2022: 1-15.

108. Chang J-Y, Chilcott J, Latimer N. Leveraging real-world data to assess treatment sequences in health economic evaluations: a study protocol for emulating target trials using the English Cancer Registry and US Electronic Health Records-Derived Database. HEDS Discussion Paper. 2024; 24.01: 1-61.

109. Franklin JM, Patorno E, Desai RJ, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. Circulation. 2021; 143: 1002-13.

110. RCT Duplicate. RCT duplicate: randomized controlled trials duplicated using prospective longitudinal insurance claims: applying techniques of epidemiology Accessed: Jan 15, 2021. Available at: www.rctduplicate.org

111. Regan MM, Neven P, Giobbie-Hurder A, et al. Assessment of letrozole and tamoxifen alone and in sequence for postmenopausal women with steroid hormone receptor-positive breast cancer: the BIG 1-98 randomised clinical trial at 8·1 years median follow-up. The Lancet Oncology. 2011; 12: 1101-08.

112. Eichelberg C, Vervenne WL, De Santis M, et al. SWITCH: A Randomised, Sequential, Open-label Study to Evaluate the Efficacy and Safety of Sorafenib-sunitinib Versus Sunitinib-sorafenib in the Treatment of Metastatic Renal Cell Cancer. Eur Urol. 2015; 68: 837-47.

113. Retz M, Bedke J, Bogemann M, et al. SWITCH II: Phase III randomized, sequential, open-label study to evaluate the efficacy and safety of sorafenib-pazopanib versus pazopanib-sorafenib in the treatment of advanced or metastatic renal cell carcinoma (AUO AN 33/11). Eur J Cancer. 2019; 107: 37-45.

114. Lei H, Nahum-Shani I, Lynch K, et al. A" SMART" design for building individualized treatment sequences. Annu Rev Clin Psychol. 2012; 8: 21-48.

115. Bujkiewicz S, Singh J, Wheaton L, et al. Bridging disconnected networks of first and second lines of biologic therapies in rheumatoid arthritis with registry data: bayesian evidence synthesis with target trial emulation. J Clin Epidemiol. 2022; 150: 171-78.

116. Gomes M, Latimer N, Soares M, et al. Target trial emulation for transparent and robust estimation of treatment effects for health technology assessment using real-world data: opportunities and challenges. Pharmacoeconomics. 2022: 1-10.

117. Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. Biometrics. 2002; 58: 48-57.

118. Robins J. Interview with Jamie Robins. Observational Studies. 2022; 8: 37-76.

119. Zang Y, Lee JJ. Adaptive clinical trial designs in oncology. Chinese clinical oncology. 2014; 3.

120. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med. 2018; 16: 29.

121. Mahajan R, Gupta K. Adaptive design clinical trials: Methodology, challenges and prospect. Indian J Pharmacol. 2010; 42: 201-7.

122. Thorlund K, Haggstrom J, Park JJ, et al. Key design considerations for adaptive clinical trials: a primer for clinicians. BMJ. 2018; 360.

123. Crown W, Buyukkaramikli N, Thokala P, et al. Constrained optimization methods in health services research—an introduction: report 1 of the ISPOR optimization methods emerging good practices task force. Value Health. 2017; 20: 310-19.