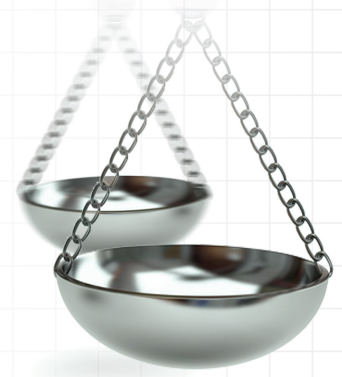




Composite endpoints in health technology assessment: Part 2 – expanding the evidence base with a framework for best modeling practice

Journal of **Comparative Effectiveness Research**

Andrew Briggs^{*1} , Jieliang Chen², David Booth³, Jason A Davis³, Aris Angelis^{1,4} ,

Muthiah Vaduganathan⁵ & Pardeep S Jhund⁶ 

¹Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK

²Global Market Access & Pricing, AstraZeneca R&D, Gaithersburg, MD 20878, USA

³Health Economics & Outcomes Research Ltd, Cardiff, UK

⁴General Secretariat for Strategic Planning, Ministry of Health, Athens, Greece

⁵Cardiovascular Division & Center for Cardiometabolic Implementation Science, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

⁶British Heart Foundation, University of Glasgow, Glasgow, UK

*Author for correspondence: andrew.briggs@lshtm.ac.uk

The first paper of this two-part series critically examined the role of composite endpoints in health technology assessments (HTAs) and outlined strategies for determining whether to employ the composite estimate of treatment effect or disaggregate into the component endpoints of the composite and apply separate treatment effects within a modeling framework. In this second paper, we expand the discussion beyond a pivotal trial and consider the way in which additional evidence from the same indication for different drugs in the same class, or the same drug for different indications, could be employed within HTAs. We offer a continuation of the case study of dapagliflozin for the treatment of heart failure with preserved or mildly reduced ejection fraction, where the evidence base was expanded to consider empagliflozin for the same indication, as well as both dapagliflozin and empagliflozin for heart failure with reduced ejection fraction. We conclude that, where there are multiple drugs in the same class and/or multiple indications for a class of drugs, HTAs should consider the exchangeability of the information that comes from additional evidence beyond the pivotal registration study. Further, we show how consideration of this additional evidence can support the decision concerning whether to disaggregate the components of a composite endpoint or employ the treatment effect from the composite for each endpoint.

Plain language summary

What is this article about? Composite endpoints combine multiple individual outcomes, known as components, into a single measure. They are commonly used in clinical trial analyses, as they can make it easier to detect differences between treatment and control groups; however, they can also present challenges when trying to determine whether new treatments provide value for money. This article highlights how to properly incorporate composite endpoints into economic analyses to ensure accurate and meaningful evaluations of health technologies. The article uses a real-world example of dapagliflozin in heart failure to show the challenges faced when handling composite endpoints and suggests ways to address these issues.

What methodology is described? Using a case study of dapagliflozin for the treatment of heart failure, this article provides a framework to evaluate different methods for including composite endpoint data in economic analyses. It looks at how to handle either composite endpoints or their individual components to assess the effectiveness of treatments, focusing on how additional evidence for different drugs in the same class from the same indication or the same drug for different indications could be used for health technology assessment.

What do the results mean? The article emphasizes the need to look at more than just key trial data when creating models for regulatory approval. It's important to consider evidence from other drugs in

the same class that treat the same or related conditions, as well as evidence from the same drug used for different conditions. Taking a broader approach can lead to more accurate and reliable economic evaluations with less uncertainty. This is particularly important for health technology assessments, where clear and consistent evaluations help guide treatment decisions and reimbursement policies.

First draft submitted: 2 February 2025; Accepted for publication: 7 August 2025; Published online: 10 September 2025

Keywords: composite endpoints • cost-effectiveness • health economic modeling • health technology assessment • sodium-glucose co-transporter 2 inhibitor

Composite endpoints have become increasingly prevalent in clinical trials, offering potential advantages in terms of efficiency of trial conduct and gains in statistical power [1,2]. However, handling composite endpoints within the context of health technology assessments (HTAs), where the timely evaluation of new treatments is crucial for informing reimbursement decisions and ensuring patient access to clinically effective and economically efficient therapies, is not straightforward, and analysts may face several challenges. The complexity of deciding whether to apply the composite estimate of treatment effect or to disaggregate the components within an economic model can have a considerable impact on the outcomes of HTAs, and ultimately the final decisions on pricing and reimbursement. Therefore, this two-paper series aimed to describe the important methodological considerations for handling composite endpoint data in cost-effectiveness models, to ensure that such data are being used appropriately in the context of HTAs.

The first article of this two-paper series explored these potential challenges by establishing a framework for best modeling practices when handling composite endpoints in HTA, involving the construction of a health economic model to support reimbursement decisions [3]. We critically examined how to determine whether to apply a composite estimate of treatment effect or to disaggregate into endpoint components and apply separate treatment effects in the context of an economic evaluation. By illustrating the implications of the various modeling assumptions on treatment effect, we informed a critical discussion that highlighted the importance of accurately capturing uncertainty in evidence used within the decision-making process.

Building on this foundation, this second paper expands the framework beyond the scope of a single pivotal trial to consider the integration of additional evidence both from the same therapeutic indication (other drugs within the same class) and for the same drug tested in other patient populations (i.e., a drug with multiple indications). This is particularly relevant in cases where multiple drugs target similar conditions, raising questions about the exchangeability of evidence across drug classes and indications, though it may also be applicable in circumstances where the same drug is used across less similar indications (e.g., as in oncology). This article presents a full methodological framework that considers the totality of evidence for drugs with multiple indications and multiple drugs within the same class, in the context of HTAs whose principal aim is to inform reimbursement decisions within a healthcare system.

To illustrate this methodological framework, we continue our previous case study of dapagliflozin, examined for the treatment of heart failure with preserved or mildly reduced ejection fraction (HFpEF/HFmrEF). Briefly, the first article of this two-paper series presented a case study of dapagliflozin plus standard of care, versus placebo plus standard of care, in patients with HFpEF/HFmrEF to assess the impact of utilizing either the composite endpoint, its components, and the influence of selectively including specific components on model outcomes [3], based on data from the DELIVER trial [4]. In this paper, we expand our consideration of the evidence base to include data on empagliflozin for the same indication, another drug in the sodium-glucose co-transporter 2 inhibitor (SGLT2i) class, as well as evidence for both dapagliflozin and empagliflozin in heart failure with reduced ejection fraction (HFrEF, a different therapeutic indication where both drugs are approved). We explore how, in scenarios where multiple drugs within the same class or across different indications are involved, HTA agencies should carefully consider the extent to which information from these additional evidence sources can enhance the estimates of treatment effect for the indication under consideration. By doing so, decision-makers can better assess the appropriateness of using composite endpoints or disaggregating component effects when evaluating submitted models. This approach not only strengthens the evidence base, but also supports more nuanced and accurate HTA outcomes, ultimately facilitating more informed and effective healthcare decision-making. In the following sections, we introduce the

general methodological framework, followed by the application of that framework to the dapagliflozin case study for HFpEF/HFmrEF.

Framework for considering clinical evidence in HTAs

As highlighted in the first paper of this two-part series, reimbursement authorities provide scant guidance on how to handle composite endpoints when submitting cost-effectiveness models for consideration. Most reimbursement authorities provide detailed guidance on how to approach clinical evidence more generally. For example, the most recent methods manual published by the National Institute for Health and Care Excellence (NICE) in the UK reports a clear preference for randomized control trial (RCT) data to assess relative treatment effectiveness [5], and provides detailed recommendations for the synthesis of such evidence:

“The aim of clinical-effectiveness analysis is to get precise, relevant and unbiased estimates of the mean clinical effectiveness of the technologies being compared. Consider all relevant studies in the assessment of clinical effectiveness and base analyses on studies of the best available quality. Consider the range of typical patients, normal clinical circumstances, clinically relevant outcomes, comparison with relevant comparators, and measures of both relative and absolute effectiveness with appropriate measures of uncertainty. NICE prefers RCTs directly comparing the intervention with 1 or more relevant comparators and, if available, these should be presented in the reference-case analysis.” – NICE 2023, paragraph 3.4.1, underline emphasis added [5]

However, NICE fails to explicitly describe which studies should be considered relevant in the assessment of clinical effectiveness, and so this remains open for some level of interpretation. In the case where there are multiple drugs within a class, the effectiveness of a similar drug within the same class is considered to be relevant, given that alternative drugs in the same class are certainly comparators for one another. Indeed, NICE recommends the use of network meta-analyses for indirect and mixed treatment comparisons for such circumstances [5]. However, it is perhaps less clear as to whether evidence for the same drug tested for different indications is considered relevant. Though, in instances where the mechanism of action and the measured clinical trial outcomes are comparable, we suggest the relative treatment effect should at least be considered as potentially relevant for inclusion.

The first paper considered whether, in a cost-effectiveness modeling exercise to support a reimbursement decision based on an RCT that used a composite endpoint, the relevant treatment effect employed in the model should be estimated on the composite endpoint, or whether the component endpoints should be disaggregated and the independent treatment effects applied [3]. This critical evaluation highlighted that using the composite endpoint was more efficient if either the treatment effects of the components were similar or the component endpoints were valued the same. However, if the components had markedly differing values and the treatment effects on the components were different, then the use of the composite treatment effect would bias the model. This choice on whether to apply the composite treatment effect or to disaggregate into the separate component treatment effects forms the first choice of the expanded framework shown in Figure 1. Alongside the choice to disaggregate the composite endpoint, there is a further option to exclude component treatment effects that do not reach conventional levels of statistical significance, as discussed in the first paper of this series [3].

The second choice of the framework is whether to include evidence on other drugs in the same therapeutic class for the given indication. The third choice is whether to include evidence related to other indications for those drugs. For this third choice, we also consider the possibility that this could apply not just to a separate indication, but also to defined subgroups within a particular indication. Often, as in the case of the heart failure example below, the distinction between subgroups within an indication and a separate indication is blurred owing more to the progress of clinical development than to truly differing disease processes. For example, the use of SGLT2is in adjacent clinical areas such as chronic kidney disease and diabetes, together with the substantial overlap between these clinical populations, has led to the emergence of cardio-renal-metabolic ‘syndrome’ [6].

From Figure 1 we see that the three dichotomous choices lead to a permutation of $2^3 = 8$ scenarios, with an additional scenario 0 to cover the situation included in the first paper where only treatment effects that are significant are included in the health economic modeling.

Case study continuation – dapagliflozin for HFpEF/HFmrEF

In this section, we continue the case study from the first paper that considered the comparative effectiveness and cost-effectiveness analysis of dapagliflozin, an SGLT2i, for the treatment of HFpEF/HFmrEF. This reflects the decision problem faced by NICE in the UK when it embarked on its recently published recommendation for

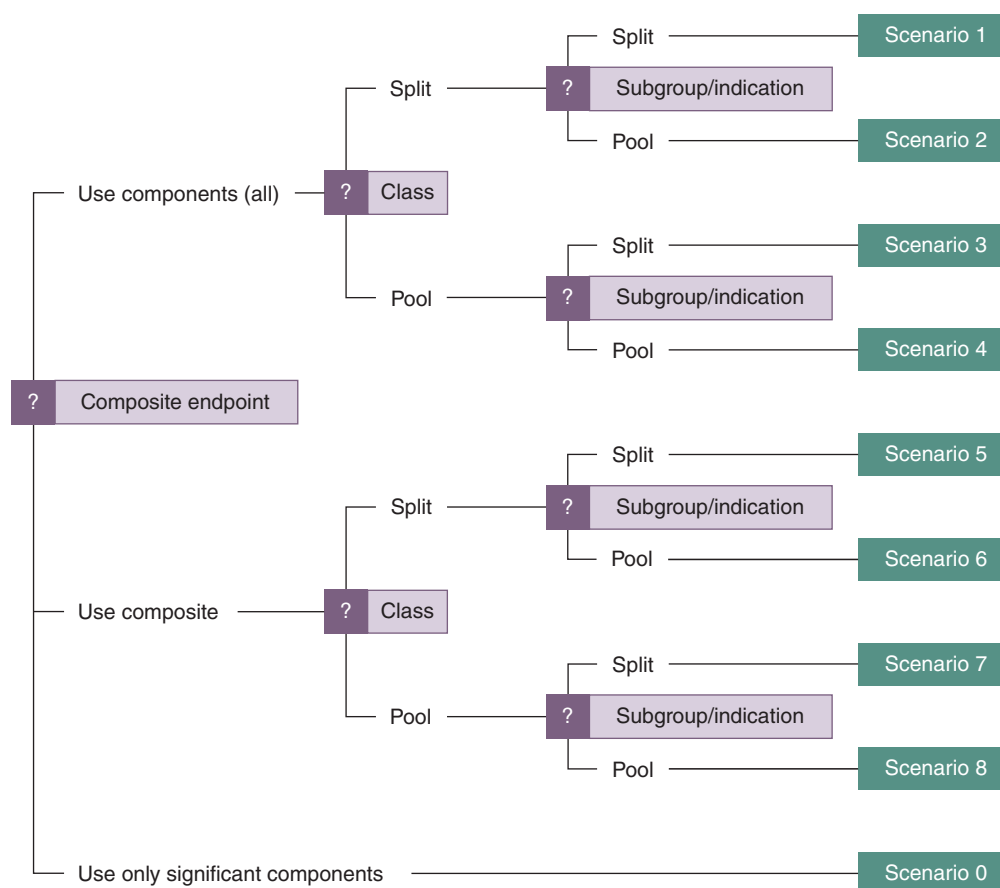


Figure 1. Framework for choices concerning the disaggregation of the composite treatment effect into component treatment effects, whether to pool across drugs in the same class and whether to pool over subgroups/indications.

dapagliflozin in this particular indication [7]. The key study informing this reimbursement decision was the Phase III, placebo-controlled DELIVER trial in adult patients with HFpEF or HFmrEF [4]. The primary outcome of the DELIVER trial was a composite endpoint of worsening heart failure, defined as either an unplanned urgent visit or hospitalization for heart failure (HHF), or cardiovascular death (CV death). The results of the DELIVER trial with respect to these outcomes are presented in scenarios 1 and 5 in Figure 1, alongside scenario 0, where the nonsignificant impact on the CV death outcome is assumed to have no treatment effect.

Applying the framework to treatment effects

Here, we expand the previous presentation to show the pivotal trials for both dapagliflozin and empagliflozin in both the HFrEF and HFpEF/HFmrEF indications. These pivotal trials are DAPA-HF [8] for dapagliflozin, EMPEROR-reduced [9] for empagliflozin in HFrEF, and DELIVER [4] and EMPEROR-preserved [10] for dapagliflozin and empagliflozin, respectively, in HFpEF/HFmrEF. Figure 2 presents the treatment effect estimates for dapagliflozin and empagliflozin from the four pivotal trials by the composite endpoint (time to first HHF or CV death) and by the disaggregated component endpoints of the composite, obtained from published literature [4,8–10]. Scenario 1 in Figure 2 shows the pivotal trial data by its component endpoints, while scenario 5 shows the composite endpoint treatment effect from each trial.

To consider the class effect for HFpEF/HFmrEF, pooled meta-analysis data of treatment effects across the DELIVER and EMPEROR-preserved trials were obtained from literature [11,12], with scenario 3 corresponding to the disaggregated endpoints and scenario 7 corresponding to the composite endpoint (Figure 2). Notably, although the treatment effect on CV death was not significant in both DELIVER and EMPEROR-preserved, the numerical value of the treatment effect was comparable in both trials, and when pooled by class, the endpoint almost reached statistical significance, suggesting the lack of a significant treatment effect is likely due to insufficient power of the

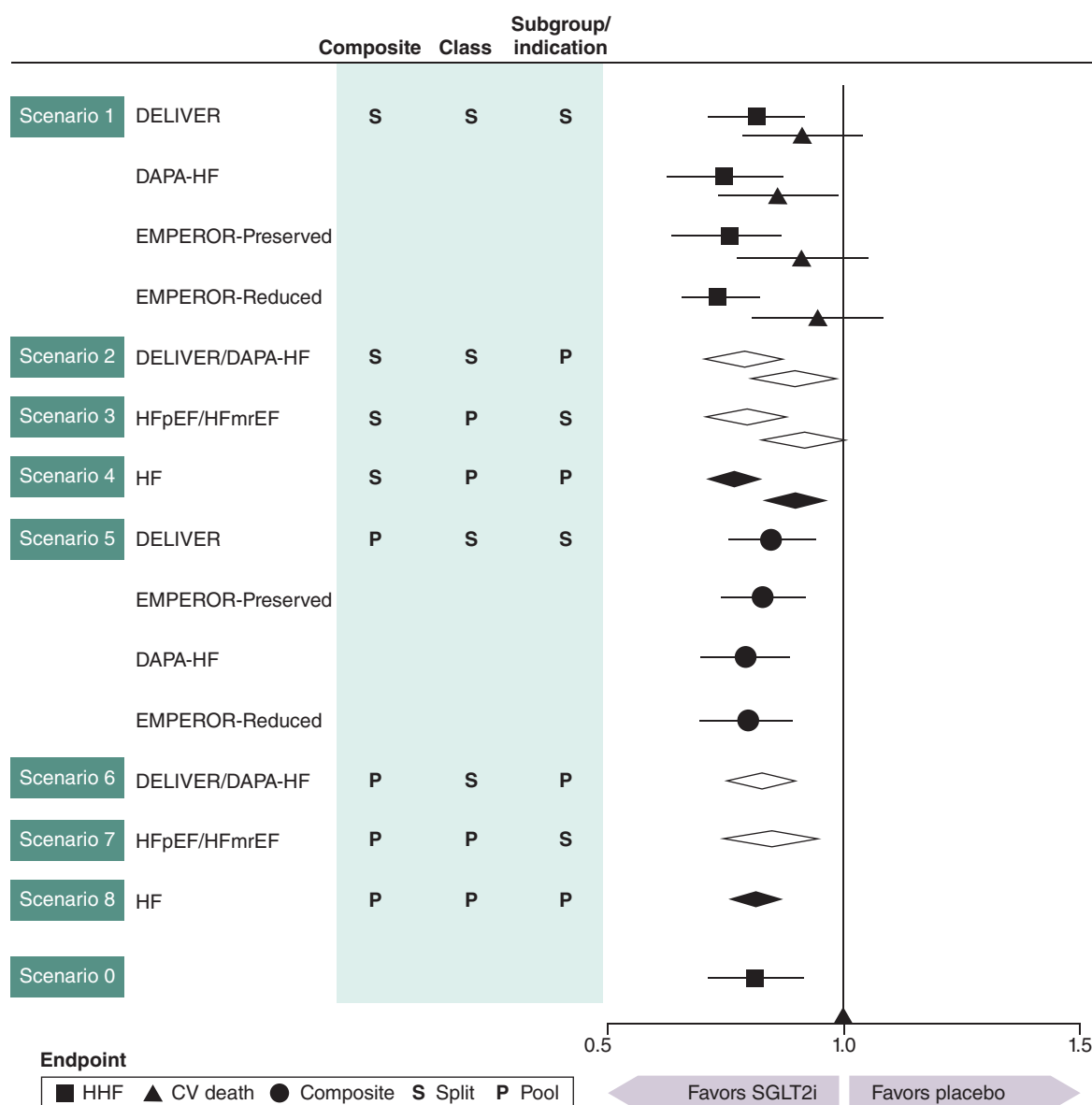


Figure 2. Treatment effect estimates based on the framework described in Figure 1 using the pivotal trials for both dapagliflozin and empagliflozin in both the heart failure with reduced ejection fraction and heart failure with preserved or mildly reduced ejection fraction indications. Urgent HF visits are included in the composite outcome definitions for dapagliflozin. Diamond symbol markers depict pooled estimates across trials, with their relative position indicating either CV death or HHF components (as per the disaggregated rows above); open diamonds correspond to disaggregated HF indications of HF (HFrEF and HFpEF/HFmrEF on separate lines) while filled diamond symbols correspond to aggregation across HF indications. CV: Cardiovascular; HFrEF: Heart failure with reduced ejection fraction; HFpEF/HFmrEF: Heart failure with preserved or mildly reduced ejection fraction; HHF: Hospitalization for heart failure; SGLT2i: Sodium-glucose co-transporter 2 inhibitor.

individual studies [13]. Alternatively, pooled meta-analysis data for dapagliflozin in the DAPA-HF and DELIVER trials were sourced from literature to represent treatment effects across both HFrEF and HFpEF/HFmrEF [12], with scenario 2 for the component endpoints and scenario 6 for the composite endpoint. As the CV death component endpoint was significant in HFrEF in the DAPA-HF trial, the pooled treatment effect on CV death was also significant in scenario 2.

Overall, the consistency of the relative treatment effects across the four pivotal trials for dapagliflozin and empagliflozin suggests that pooling across both drugs in class and across both indications may be appropriate.

Fully pooled estimates derived from the literature are shown for the component endpoints (scenario 4) and for the composite endpoint (scenario 8) in [Figure 2](#) [11,12]. Of particular note in scenario 4 is that, as well as the upper limit of the confidence interval (CI) for the CV death component endpoint not crossing 1, the two component endpoints were shown to have treatment effects that are clearly distinct from one another. By incorporating evidence beyond the pivotal DELIVER trial, results support the nonsignificant treatment effect on CV death observed in the DELIVER trial and suggest this finding may be due to the study being underpowered to test disaggregated component endpoints.

Applying the framework to cost-effectiveness results

A cost-effectiveness model, produced to guide NICE's reimbursement decision for dapagliflozin [7,14], was populated with input parameters derived from published studies that informed the treatment effect estimates shown in [Figure 2](#). The natural logarithms of the published relative risk (RR) for HHF and hazard ratio (HR) for CV death were calculated and used to inform the coefficient of treatment in the Weibull model generalized estimating equation of HHF and the Weibull survival model for CV death. Uncertainty of the coefficient of treatment was assessed using the variance, derived from the reported 95% CIs for the RR for HHF and HR for CV death. In all scenarios, non-CV death was held constant, while the effect of treatment on CV mortality was varied, allowing an evaluation of its impact within the model. Probabilistic sensitivity analyses were conducted over 1000 iterations for each scenario, producing estimates for the incremental net-monetary benefit (NMB), incremental cost-effectiveness ratio (ICER) and expected value of perfect information (EVPI), based on the average of the 1000 runs.

The cost-effectiveness estimates in terms of NMB are presented in [Figure 3](#). Results indicated that the most favorable cost-effectiveness outcomes were observed in scenarios 5–8, where the composite estimate of treatment effect was applied to both HHF and CV death outcomes in the model. Notably, less favorable outcomes from a cost-effectiveness perspective were observed when the independent treatment effects for each of the component endpoints were applied, as demonstrated in scenarios 1–4. The least favorable outcomes were estimated for scenario 0, where the lack of a significant treatment effect on CV death was equated with no treatment benefit on that outcome. These results were expected given that CV death is the most serious outcome, thus has the potential to generate the most health gains and healthcare cost-savings when prevented. Despite the lack of significance of the treatment effect on the CV death endpoint in the DELIVER and EMPEROR-preserved trials, the pooled results show the CV death endpoint was closer to reaching statistical significance ($p = 0.052$) at conventional levels [11]. Therefore, extending consideration of treatment effects beyond a single pivotal trial in this case counters the argument that lack of significance should be interpreted as lack of effect. Further, a fully pooled result (scenario 4) provides strong evidence that treatment effects for the component endpoints are not the same, suggesting the lack of significant effects is likely due to studies being underpowered [13].

Overall, the point estimates of scenarios 1–4 are not very different numerically, and none result in a significant NMB. Nevertheless, as one would expect, the uncertainty intervals on NMB shrink once the additional evidence is taken into consideration. [Table 1](#) shows the cost-effectiveness results in more detail with the NMB and 95% uncertainty interval results from [Figure 3](#), together with the estimated ICER and the EVPI. The EVPI is a measure of decision uncertainty that combines the probability of incorrect decision making with the loss of net-benefit that would result from incorrect decision making; higher values are suggestive that additional evidence may be required before a decision can be reached. Although the impact of the different scenarios on the point estimates of cost-effectiveness was modest, the reduction in uncertainty, as illustrated by the width of the uncertainty interval and the magnitude of the EVPI, shows that the fully pooled scenario 4 had less uncertainty in cost-effectiveness outcomes than scenario 1, which relies only on the pivotal DELIVER trial.

Discussion & limitations

The first part of this two-paper series examined the increasing use of composite endpoints in clinical trials from the perspective of health economic modeling analysts preparing a submission to a reimbursement/HTA authority [3]. We noted that where the component endpoints of a composite endpoint have different values, these will be reflected as separate outcomes in the health economic model, but that the treatment effect parameters attached to those outcomes could either be represented by the composite treatment effect in the clinical trial or from the independent treatment effects estimated from each component. We noted that only when the treatment effects are truly different and the values attached to the component endpoints differ will the use of the composite treatment effect result in bias. Using the case study of dapagliflozin for the treatment of HFpEF/HFmrEF, we argued that the evidence from

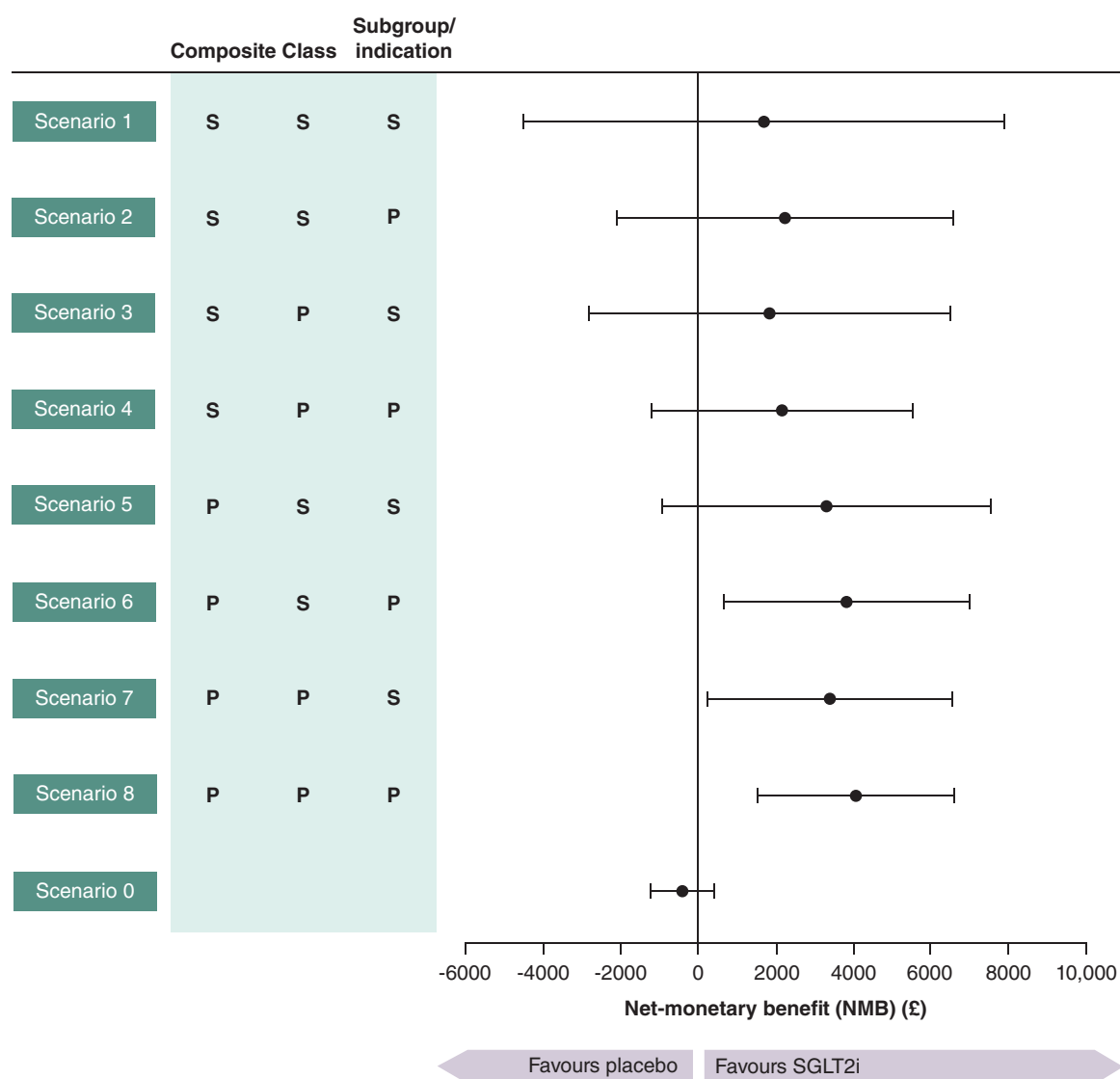


Figure 3. Net-monetary benefit with 95% uncertainty intervals for each scenario of the framework assuming £20,000 per quality-adjusted life year as the decision-making threshold.

NMB: Net-monetary benefit; SGLT2i: Sodium-glucose co-transporter 2 inhibitor.

Table 1. Cost-effectiveness results from each Scenario from the framework with uncertainty measures.

Scenario	ICER	NMB	(95% UI)	EVPI
Scenario 1	£10,256	£1697	(-£4510 to £7904)	£609
Scenario 2	£8791	£2262	(-£2079 to £6603)	£182
Scenario 3	£9636	£1857	(-£2827 to £6541)	£305
Scenario 4	£8847	£2154	(-£1225 to £5534)	£87
Scenario 5	£7636	£3306	(-£952 to £7563)	£66
Scenario 6	£6991	£3804	(£638 to £6970)	£5
Scenario 7	£7557	£3357	(£205 to £6510)	£11
Scenario 8	£6710	£4052	(£1541 to £6562)	£0
Scenario 0	£27,152	-£426	(-£1227 to £376)	£36

EVPI: Expected value of perfect information; ICER: Incremental cost-effectiveness ratio; NMB: Net-monetary benefit; UI: Uncertainty interval.

the pivotal DELIVER trial was suggestive that the treatment effects on HHF and CV death were different and that given the different QALY values and health service costs associated with these events, the use of independent treatment effects was recommended. Further, despite the lack of significance for the CV death component endpoint, we advised against omitting the treatment effect on CV death as the clinical study was not powered for detecting a difference in the component endpoints.

In this second paper, the framework was extended to consider evidence from different drugs within the same therapeutic class for the same indication and from the same drug with multiple indications. We proposed eight possible scenarios within the framework to demonstrate how the related evidence may be combined, and added a ninth scenario for when nonsignificant treatment effects on component endpoints were disregarded. We illustrated, in a continuation of the SGLT2i for HFpEF/HFmrEF case study, how the incorporation of other related evidence to the model impacted treatment effect estimates for HHF and CV death outcomes, and subsequently the respective treatment cost-effectiveness estimates.

The application of our framework demonstrated the importance, from an HTA modeling perspective, of considering all evidence relating to the relative treatment effects of an intervention, rather than relying only on the evidence from a single or small number of clinical trials. Expanding the evidence base to other drugs in class and between similar populations that use the same outcome borrows ‘strength’ and potentially allows more robust decision-making. This is consistent with NICE methods guidance to consider all the relevant evidence. Also of note, in relation to the expanded indication, is that NICE had already given a positive reimbursement decision to both dapagliflozin and empagliflozin for the HFpEF indication. Given the close correspondence of the study designs of the pivotal RCTs between the two SGLT2is across both indications, it is natural in relation to this case study to consider that the relative treatment effects might be exchangeable and therefore pooling across studies to be appropriate. Our analysis has illustrated that the key advantage of pooling evidence is the resulting reduction in uncertainty, rather than any substantial changes to the point estimates of cost-effectiveness (c.f. scenarios 1 and 4). Although we have approached this framework from the perspective of HTA decision-making for bodies such as NICE in the UK, we believe that the approach considered here may also be relevant to the development of clinical guidelines and for drug regulatory bodies. Indeed, the EMA accepted a pooled analysis of the DAPA-HF and DELIVER trials [12]; [Figure 2\(f\)](#) in the label for dapagliflozin [15]. By contrast, the FDA did not accept the pooled analysis and instead considered the HFpEF population covered by DAPA-HF and HFmrEF/HFpEF populations covered by DELIVER as separate indications [16].

A key limitation of the framework is the need for substantial subjective judgment. When extending the evidence base beyond the key pivotal trial(s), consideration needs to be given to the exchangeability of the additional information, both from a design perspective and an empirical perspective. For example, it is important to consider whether trial designs are comparable, and whether treatment effects vary between drugs of the same class or whether they have similar chemical structures [17–20]. It is also necessary to review whether the indications are similar and whether the expected mechanism of action for the drug is the same across different indications. For our case study, we would argue that the case for exchangeability of evidence is high, especially given that the expanded indication is along the continuum of ejection fraction in HF. In this example, the two indications are more akin to a subgroup analysis on a dichotomization of ejection fraction. Indeed, recent clinical studies have pooled data across the two indications and presented continuous analysis of the ejection fraction covariate [12,21]. It is possible that in other applications of the framework presented here, the exchangeability of the evidence base may be less clear, and further critical evaluation of the evidence may be required. Nevertheless, all evidence should be considered together, and the framework presented in this publication provides a mechanism for doing so.

Specific to the HF case study presented here, an HTA body may decide to consider only the pivotal clinical trial (DELIVER) and not rely on pooling data across class and/or indication. In that case, however, the consideration of the expanded evidence would at least allow them to conclude that scenario 0, where the treatment effect on CV death is denied due to lack of statistical significance, should not be entertained.

To summarize, this article highlights the importance of considering other relevant information beyond pivotal trial data when preparing models for regulatory approval, particularly where evidence is available from different drugs (in the same class) with the same therapeutic indication and from the same drug in other therapeutic indications. The framework and worked case study presented in this two-paper series provides a foundation for future researchers to consider when handling composite endpoints within the context of HTAs.

Executive summary

- Composite endpoints are formed by aggregating several individual clinical endpoints into a single outcome.
- The use of composite endpoints to assess treatment effect is becoming more common in clinical trials, likely due to their benefits in terms of trial efficiency and improvements in statistical power.
- According to many widely referenced guidelines, all relevant evidence relating to a decision problem should be considered when assessing relative treatment effectiveness.
- As reimbursement authorities do not provide specific guidance on the handling of composite endpoints to inform cost-effectiveness analyses, there are inconsistencies between the approaches/methodologies followed.
- The first paper of this two-part series presented a case study assessing various scenarios using either the composite endpoint or its disaggregated components, as well as the selective inclusion of certain component effects to explore potential challenges.
- This second article presents a full methodological framework to critically assess how additional evidence for different drugs in the same class from the same indication or the same drug for different indications could be employed within health technology assessments, where appropriate.
- Developing on the case study of dapagliflozin for the treatment of heart failure with preserved or mildly reduced ejection fraction presented in the first paper, the evidence base was expanded to include empagliflozin for the same indication, as well as both dapagliflozin and empagliflozin for heart failure with reduced ejection fraction.
- Aggregation of treatment efficacy data (use of a composite endpoint, across members of a drug class for the same indication or related indications or both) tended to reduce uncertainty in cost-effectiveness results compared with corresponding disaggregated analyses.

Author contributions

A Briggs and J Chen conceived the study. A Briggs performed statistical analyses and generated data visualizations. A Briggs, J Chen, D Booth, JA Davis, A Angelis, M Vaduganathan and PS Jhund contributed to data interpretation, manuscript preparation and critical review of the manuscript.

Financial disclosure

This research was funded by AstraZeneca.

Competing interests disclosure

A Briggs has acted as consultant to various commercial companies that have products funded through the NHS. J Chen is an employee of AstraZeneca. D Booth and JA Davis are employees of Health Economics and Outcomes Research Ltd, who received funding from AstraZeneca in relation to this work. A Angelis is working in the Hellenic Ministry of Health (the opinions in this paper do not reflect the views of the Ministry of Health) and he was a 2021-2022 scholar at the National Institute for Health and Care Excellence. M Vaduganathan has received research grant support, served on advisory boards, or had speaker engagements with American Regent, Amgen, AstraZeneca, Bayer AG, Baxter Healthcare, BMS, Boehringer Ingelheim, Chiesi, Cytokinetics, Fresenius Medical Care, Idorsia Pharmaceuticals, Lexicon Pharmaceuticals, Merck, Milestone Pharmaceuticals, Novartis, Novo Nordisk, Pharmacosmos, Relypsa, Roche Diagnostics, Sanofi, and Tricog Health, and participates on clinical trial committees for studies sponsored by AstraZeneca, Galmed, Novartis, Bayer AG, Occlutech and Impulse Dynamics. PS Jhund reports speakers' fees from AstraZeneca, Novartis, ProAdWise Communications; advisory board fees from AstraZeneca, Boehringer Ingelheim; research funding from AstraZeneca, Boehringer Ingelheim, Analog Devices Inc., Roche Diagnostics PSJ's employer the University of Glasgow has been remunerated for clinical trial work from AstraZeneca, Bayer AG, Novartis and Novo Nordisk. Director, Global Clinical Trial Partners (GCTP). The authors have no other competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript apart from those disclosed.

Writing disclosure

Editorial support was provided by Health Economics and Outcomes Research Ltd, funded by AstraZeneca

Data sharing statement

Data are available upon reasonable request to the corresponding author.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. Shi L, Lindsell CJ, Liu D. Trends in use of composite endpoints in clinical trials: a comparison between acute heart failure trials and COVID-19 trials. *J. Clin. Transl. Sci.* 8(1), e55 (2024).
2. Armstrong PW, Westerhout CM. composite end points in clinical research: a time for reappraisal. *Circulation* 135(23), 2299–2307 (2017).
3. Briggs A, Angelis A, Chen J *et al.* Composite endpoints in health technology assessment: part 1 – an illustration of best modelling practice. *J. Comp. Eff. Res.* 14(10), e240117 (2025).
4. Solomon SD, McMurray JJV, Claggett B *et al.* Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. *N. Engl. J. Med.* 387(12), 1089–1098 (2022).
5. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual (2023). Available from: <https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741> (Accessed: 14 November 2024).
- **An extensive guide published by the UK government agency that outlines their preferred modeling/analytical approaches.**
6. Ndumele CE, Ranganaswami J, Chow SL *et al.* Cardiovascular–kidney–metabolic health: a Presidential Advisory from the American Heart Association. *Circulation* 148(20), 1606–1635 (2023).
7. National Institute for Health and Care Excellence. Dapagliflozin for treating chronic heart failure with preserved or mildly reduced ejection fraction. *Final Draft Guidance*, (2023). Available from: <https://www.nice.org.uk/guidance/ta902/documents/final-appraisal-determination-document> (Accessed: 14 November 2024).
8. McMurray JJV, Solomon SD, Inzucchi SE *et al.* Dapagliflozin in patients with heart failure and reduced ejection fraction. *N. Engl. J. Med.* 381(21), 1995–2008 (2019).
9. Packer M, Anker SD, Butler J *et al.* Cardiovascular and renal outcomes with empagliflozin in heart failure. *N. Engl. J. Med.* 383(15), 1413–1424 (2020).
10. Anker SD, Butler J, Filippatos G *et al.* Empagliflozin in heart failure with a preserved ejection fraction. *N. Engl. J. Med.* 385(16), 1451–1461 (2021).
11. Vaduganathan M, Docherty KF, Claggett BL *et al.* SGLT-2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials. *Lancet* 400(10354), 757–767 (2022).
- **Important source of evidence in present study as a meta-analysis of treatment effects in relevant disease area by class.**
12. Jhund PS, Kondo T, Butt JH *et al.* Dapagliflozin across the range of ejection fraction in patients with heart failure: a patient-level, pooled meta-analysis of DAPA-HF and DELIVER. *Nat. Med.* 28(9), 1956–1964 (2022).
- **Important source of evidence in present study as a pooled, patient-level data meta-analysis of treatment effects for a single drug across related indication.**
13. Kondo T, Henderson AD, Docherty KF *et al.* Why have we not been able to demonstrate reduced mortality in patients with HFmrEF/HFpEF? *J. Am. Coll. Cardiol.* 84(22), 2233–2240 (2024).
14. Booth D, Davis JA, McEwan P *et al.* The cost-effectiveness of dapagliflozin in heart failure with preserved or mildly reduced ejection fraction: a European health-economic analysis of the DELIVER trial. *Eur. J. Heart Fail.* 25(8), 1386–1395 (2023).
15. Electronic medicines compendium. Forxiga 10 mg film-coated tablets SmPC (2024). Available from: <https://www.medicines.org.uk/emc/product/7607/smpc> (Accessed: 14 November 2024).
16. DAILYMED. FARXIGA- dapagliflozin tablet, film coated (2024). Available from: <https://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=40cc4f5d-bcf6-4016-9d1e-838bbc4bdddc> (Accessed: 14 November 2024).
17. Furberg CD, Herrington DM, Psaty BM. Are drugs within a class interchangeable? *Lancet* 354(9185), 1202–1204 (1999).
18. Furberg CD, Pitt B. Are all angiotensin-converting enzyme inhibitors interchangeable? *J. Am. Coll. Cardiol.* 37(5), 1456–1460 (2001).
19. Kennedy HL, Rosenson RS. Physicians' interpretation of "class effects": a need for thoughtful re-evaluation. *J. Am. Coll. Cardiol.* 40(1), 19–26 (2002).
20. McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' Guides to the Medical Literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 282(14), 1371–1377 (1999).
- **Proposed guidelines by the Evidence-Based Medicine Working Group for the definition of a drug class effect.**
21. Butler J, Packer M, Filippatos G *et al.* Effect of empagliflozin in patients with heart failure across the spectrum of left ventricular ejection fraction. *Eur. Heart J.* 43(5), 416–426 (2022).