Surrogate endpoints in cost-effectiveness analysis for use in health technology assessment

White paper

January 2025

Table of contents

1 Introduction

Surrogate endpoints are biomarkers or other intermediate outcomes that predict a treatment effect on a final clinical outcome. The topic of surrogate endpoints in cost-effectiveness analysis is of relevance to many health technology assessment (HTA) bodies. This is because of their increasing use in regulation which can result in a lack of data on long-term effectiveness when a new health technology enters the market.

Some HTA bodies have published methodological guidance on the acceptability of using surrogate endpoints to inform decision making. However, the level of detail varies widely, and there is limited guidance on validation and considerations for economic modelling. Furthermore, despite the extensive published general methodological guidance on conceptualisation of economic models and evidence synthesis, there remains a lack of clarity on how to address surrogate endpoints specifically when conceptualising an economic model to inform HTA.

In this context, a working group of HTA bodies collaborated to produce recommendations on best practice when using surrogate endpoints in health economic models to inform HTA decision making. The project activity was led and coordinated by the National Institute for Health and Care Excellence (NICE) in the UK. Members of the working group included:

- Canada's Drug Agency (CDA-AMC), previously the Canadian Agency for Drugs and Technologies in Health (CADTH)
- the Institute for Clinical and Economic Review (ICER) in the US
- the Australian Department of Health and Aged Care
- the National Health Care Institute (ZIN) in the Netherlands
- the Institute for Technology Assessment in Health (IETS) in Colombia
- Rubix Health in the US.

Working group members led on different tasks and the results from each task were then used to determine best practices in model conceptualisation for cost-effectiveness analysis involving surrogate endpoints. The final set of

recommendations are designed to be used alongside existing economic modelling guidance when using a surrogate endpoint, and include considerations around:

- definition
- justification
- adoption
- statistical validation
- incorporation
- reporting
- approaches to quantify and present uncertainty.

The recommendations are a consensus from working group members and are intended for all interested parties involved in developing health economic models for HTA decision making. The original inspiration for the paper was the challenges arising from changes in pharmaceutical clinical trials and regulation. However, many of the recommendations in this paper are generalisable to other health technologies, because they were developed by a working group that included members and participants with a breadth of experience in HTA.

The authors of this paper acknowledge the contributions made by all working group members and those from other HTA bodies, and by academic colleagues and representatives who were involved in focus groups to inform the paper or reviewed the paper. A list of contributors is included at the end of the paper.

2 Background

HTA is used by many healthcare systems to inform pricing and reimbursement recommendations of new technologies. There is variation between healthcare systems in exactly how HTA is used, both in terms of the scope of technologies that are evaluated (for example, drugs, devices and interventional procedures) and what methods are used to inform decision making. However, in general, HTA will aim to assess the added value of a new technology (the clinical or relative effectiveness of the new technology against relevant comparators) and the cost implications and or budget impact to inform a decision on whether the technology should be made available by payers. HTA often tends to take a long-term perspective that assesses the future impact in terms of health outcomes and costs of introducing a new technology for an acute or chronic condition in the healthcare system.

Several countries include cost-effectiveness analysis as part of HTA that may use simulation modelling to estimate final outcomes. Such models require evidence about the long-term effectiveness and costs of the new technology and any relevant comparators. However, surrogate endpoints (biomarkers or other intermediate outcomes that predict treatment effect on a final clinical outcome) are being increasingly used in licensing. Over the past 30 years, more than half of new drug and biological treatment approvals by the US Food and Drug Administration (FDA), European Medicines Agency (EMA), and Pharmaceuticals and Medical Devices Agency (PMDA) in Japan were based on trial evidence of treatment effects on a biomarker endpoint as an acceptable surrogate (Ciani et al., 2023). Products granted conditional marketing approval or accelerated approval by the EMA between 2011 and 2018 have predominantly relied on non-validated surrogate endpoints (Schuster Bruce et al., 2019). This means that there is often a lack of data on long-term effectiveness for new health technologies entering the market. This creates challenges for HTA bodies and payers when they need such evidence to assess the clinical and cost effectiveness of technologies, and the challenge is particularly acute when the surrogate endpoint is novel and lacks clear links to the final outcomes of interest. It is not uncommon for decision

makers to find themselves having to make recommendations in the absence of well-established evidence validating surrogate endpoints.

The study by Ciani et al. (2023) (Ciani et al., 2023) reported inconsistencies in how surrogate endpoints are defined and appraised by clinicians, regulators and HTA bodies. Regulatory agencies typically refer only to biomarkers (physical signs or laboratory measures that are indicators of normal biological processes, pathogenic processes or responses to an exposure or intervention) as surrogate endpoints, while treating all other endpoints as final outcomes (U.S. Food and Drug Administration et al., 2016). However, HTA bodies and clinicians have a broader definition of surrogate endpoints that includes both biomarkers and other so-called 'intermediate' outcomes. Ciani et al. (2023) proposed a framework for determining whether a trial outcome is an 'intermediate' or final outcome to provide clarity and support consistency across different settings (Ciani et al., 2023).

There is some published methodological guidance on using surrogate endpoints to inform decision making. But the level of detail varies and there is a limited guidance on using surrogate endpoints for economic modelling and therefore this paper aims to address some of these gaps. (Bujkiewicz et al., 2019; Caro et al., 2012; Roberts et al., 2012).

3 Objectives

The main objective was to produce recommendations on how to address surrogate endpoints in health economic models. The focus on using surrogate endpoints in health economic models was chosen because of the lifetime horizon usually adopted in these models, making instances in which surrogate endpoints are used to estimate treatment effectiveness particularly challenging for decision making.

The project was divided into 4 tasks, each led by one of the collaborating organisations, with the goal of addressing the following questions:

1. What are the regulatory standards for validation of surrogate endpoints? (Task lead: Australian Department of Health and Aged Care)
2. What are the published statistical methods for validation of surrogate endpoints? (Task lead: CDA-AMC)
3. What are the existing standards for validation of surrogate endpoints of HTA bodies? (Task lead: NICE)
4. How have surrogate endpoints been addressed in existing HTA evaluations by different HTA bodies and what lessons can be learnt from these assessments? (Task lead: NICE)

The results from each task were then used to determine best practices in model conceptualisation for cost-effectiveness analysis involving surrogate endpoints.

This report summarises the key findings from each task and presents recommendations and considerations for economic modelling when using surrogate endpoints to inform HTA decision making.

The full methodological details and findings for each task are provided in the [supplementary material](#).

4 Overview of methods and key findings

4.1 Regulatory standards for validation of surrogate endpoints

A scoping review of regulatory assessment guidance documents was carried out to identify regulatory standards for validation of surrogate endpoints. This review searched the EMA, FDA, PMDA and International Council on Harmonisation (ICH) websites. This was supplemented by a systematic search of Ovid MEDLINE for articles published in English within the past 5 years (January 2019 to February 2024) with 38 articles screened and their reference lists reviewed to identify regulatory guidance (further details in the supplementary material).

A total of 8 guidance documents were identified through the review (refer to the supplementary material). Of those, 6 guidance documents primarily discuss the criteria and considerations for biomarker and surrogate qualification, the evidence required for clinical effectiveness, and the validation process for surrogate endpoints in the context of life-threatening and severely debilitating diseases. Two documents served as case references relevant to this research question.

The FDA has 3 categories for surrogate endpoint consideration at different stages: 'candidate', 'reasonably likely', and 'validated'. Table 1 outlines these categories, providing definitions and summarising the use of surrogate endpoints in regulatory decision making.

Table 1 Surrogate endpoints in regulatory decision making

Item	Description
Clinical endpoint	A clinically meaningful measure of how a patient feels, functions or survives. Regulatory agencies ideally require randomised data on a final patient-relevant endpoint such as overall survival or a morbidity endpoint (for example, stroke or myocardial infarction) (U.S. Food and Drug Administration, 2018).
Biological marker	A physical sign or laboratory measurement that occurs in association with a pathological process and that has putative diagnostic and or prognostic utility (Lesko & Atkinson, 2001; U.S. Food and Drug Administration, 2018). Biomarkers may include molecular, histologic, radiographic or physiological characteristics (U.S. Food and Drug Administration, 2018).
Surrogate endpoint	A biomarker that is intended to serve as a substitute for a clinically meaningful endpoint and is expected to predict the effect of a therapeutic intervention (Lesko & Atkinson, 2001). The relationship between the surrogate endpoint and the clinical benefit may be based on the following hierarchy: randomised data, epidemiologic or non-randomised data, mechanistic or /pathophysiologic reasoning, and other scientific evidence (U.S. Food and Drug Administration, 2018)
Candidate surrogate endpoint category	Has a mechanistic connection to a final patient-relevant endpoint but has insufficient epidemiological (or randomised) data to establish correlation with the final patient-relevant endpoint. (U.S. Food and Drug Administration, 2018).
Reasonably likely surrogate endpoint category	Supported by strong mechanistic and or epidemiological rationale such that an effect on the surrogate endpoint is expected to be correlated with an endpoint intended to assess clinical benefit in clinical trials, but without sufficient clinical (randomised) data to show that it is a validated surrogate endpoint. Decisions around what is a reasonable likely surrogate endpoint are often made through ongoing discussions between the FDA and sponsor evaluating the individual rationale and evidence available. Reasonably likely surrogate endpoints may be used for accelerated approval for medicines (partway through the clinical development programme) and potentially also for approval or clearance of medical devices. In the case of accelerated approval for medicines, post-marketing confirmatory trials are required to verify and describe the anticipated effect on morbidity or mortality or other clinical benefit. Well-known examples are radiographic evidence of tumour shrinkage (response rate) and progression-free survival in certain cancer types. (U.S. Food and Drug Administration, 2018)
Validated surrogate	Has a combination of a clear mechanistic rationale and ideally, data from multiple randomised trials showing that the effect on the surrogate endpoint predicts the effect on the clinical outcome of primary interest.

Item	Description
endpoint category	Validated surrogate endpoints can be used as the basis for regular marketing approval of a medical product without the need for additional studies to demonstrate the clinical benefit.
Situations where validated surrogate endpoints can inform regulatory decision making	To address various challenging situations in which conducting outcome studies against placebo would be considered unethical and direct patient-relevant outcomes would require prolonged studies. For accelerated approval for serious diseases, is considered to expedite access to promising treatments where there is an unmet need. To assess harm (for example, Hy's Law as a predictor of hepatic toxicity) (U.S. Food and Drug Administration et al., 2016)

A [summary table on the FDA website summarises surrogate endpoints used by sponsors as primary efficacy clinical trial endpoints for drug approval or licensure by the FDA](#) (U.S. Food and Drug Administration, 2022). Surrogate endpoints that may be appropriate for use as primary clinical trial endpoints for efficacy that have not been used yet to support applications for approval are also presented in this table. This table represents a good resource for drug developers considering going through the FDA regulatory process. The FDA website highlights that the surrogates in the table are context dependent, depending on the disease, patient population, therapeutic mechanism of action, and whether there are treatments available or not.

As is the case for validation of surrogate endpoints more generally, the concept of regulatory validation of a surrogate endpoint is somewhat vague. There are no standardised regulatory requirements to establish that a surrogate endpoint is a reliable predictor of a patient-relevant clinical endpoint. Regulatory acceptance of a surrogate endpoint usually relies on factors such as biological plausibility, statistical correlations (patient-level and also ideally trial-level), and consensus within the clinical community.

4.2 HTA guidance on use of surrogate endpoints

A targeted review of HTA guidance documents was carried out to identify guidance on using surrogate endpoints to inform HTAs. This review built on

findings of a recently published review reporting results based on searches of HTA guidance run in 2018 (Grigore et al., 2020). A total of 22 guidance documents were reviewed from 18 HTA bodies (which included the European Network for Healthcare Technology Assessment [EUnetHTA]). The guidance from 17 of these HTA bodies included details related to surrogate endpoints and were included in data extraction (further details in the supplementary material).

Our review indicates that there has been little change in guidance on the use of surrogate endpoints in HTA since 2018 (Grigore et al., 2020), despite a number of these bodies having updated their methods guidance. There remains a varying level of detail provided by different HTA bodies in the areas considered in the data extraction (for example, overall acceptability, strength of evidence, validation, transferability and reporting requirements). Additionally, there is little information available on how surrogate endpoints should be considered when conceptualising economic models. This highlights the need for consistent, detailed guidance for those developing economic models to support HTA decision making. Working group discussions confirmed that the identified guidance is in line with what they would expect. This is in terms of consistency around acceptability of surrogate endpoints across HTA body guidance, and lack of clarity on what constitutes a valid surrogate endpoint and how they should be considered in economic models. These findings were also confirmed by the focus group discussions (see section 4.4) and aligned with the findings in a recently published comparative analysis of HTA positioning by the Office of Health Economics in the UK (Radu et al., 2024). An overview of the findings of the review are presented in Table 2.

Table 2 HTA guidance on the use of surrogate endpoints

Item	Description
Overall acceptability of surrogate endpoints	Overall acceptability of surrogate endpoints is mentioned in guidance documents from all the 17 HTA bodies included for data extraction. The following themes were identified:

Item	Description
	• There is a consistent preference for final outcomes to be used.
	• If a final outcome is not available, then a surrogate endpoint can be used.
	• If a surrogate endpoint is to be used, then it should be validated.
	• In all cases, the use of the surrogate endpoints (validated or non-validated) must be justified, and their validity evaluated.
	• If there are no final outcomes and no validated surrogate endpoints, then some HTA bodies suggest submissions can include a non-validated surrogate endpoint. But a non-validated surrogate endpoint may not be accepted as evidence of additional benefit of an intervention regardless of benefits observed.
	• There is a recognition of limitations when assessment of comparative effectiveness of interventions is informed by surrogate endpoints and that these limitations need to be acknowledged, and efforts taken to address them.
Strength of evidence	8 of the 17 HTA bodies mention strength of evidence to support the use of surrogate endpoints in their guidance documents.
	Where bodies provide detail on strength of evidence, usually there is reference to the following levels outlined in Ciani et al. (2017) (Ciani et al., 2017), with preference for effect shown in multiple studies:
	• Level 1: evidence demonstrating that treatment effects on the surrogate endpoint correspond to effects on the patient-centred outcome (from clinical trials); comprises a meta-analysis of several randomised controlled trials; showing that changes in the surrogate can predict commensurate changes in the final outcomes.
	• Level 2: evidence demonstrating a consistent association between the surrogate endpoint and the final patient-centred outcome (from interventional, epidemiological or observational studies).
	• Level 3: only evidence of biological plausibility of an association between the surrogate endpoint and the final patient-centred outcome (from pathophysiological studies and or an understanding of the disease process).
	There is recognition that while the preference may be for level 1, in some instances (for example, for interventions for rare diseases), lower levels of evidence may be acceptable. Some manuals refer to a holistic approach, for example, statistical relationship needs to be accompanied with explanation of biological plausibility.

Item	Description
Validation of surrogate endpoints	7 of the 17 HTA bodies mention validation of surrogate endpoints in their guidance documents. 2 HTA bodies describe a staged validation that follows the levels described in Ciani et al. (2017); starting with establishing biological plausibility (level 3). Then, evaluating if there is a strong correlation between the surrogate endpoint measure and the final outcome measure (level 2). Then finally evaluating whether there is evidence to demonstrate the relationship between the treatment effect in the surrogate and the effect on the final outcome measure, preferably in multiple randomised trials (level 1). Only 3 HTA bodies mention correlation values in relation to demonstration of validity of a surrogacy relationship in their guidance documents: the Institute for Quality and Efficiency in Health Care (IQWiG) in Germany; the National Authority of Medicines and Health Products (INFARMED) in Portugal, which is informed by IQWiG guidance; and the EUnetHTA, which includes multiple EU HTA bodies (EUnetHTA 21, 2023; Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2011, 2022; Perelman et al., 2019). They all state that while there is no clear consensus on values for confirming adequacy of a surrogate, a correlation coefficient that falls in the range between 0.85 and 0.95 is generally cited in the literature to indicate a strong correlation. The guidance from IQWiG and INFARMED states that a correlation is classified as strong if the lower limit of the confidence interval of the correlation coefficient R is 0.85 or above. It is classified as weak if the upper limit of the confidence interval of R is 0.70 or below, and it is classified as medium if the confidence interval of R overlaps, even partially, the interval between 0.70 and 0.85. It also outlines how the conclusion would vary based on quality of the validation study. The guidance from all 3 bodies states that even if there is no high correlation demonstrated, conclusions may still be made around the validity of a surrogate if the surrogate threshold effect (defined in Burzykowski et al. 2005 as the minimum treatment effect on a surrogate that is reliably predictive of a treatment effect on the clinical outcome) is taken into consideration (Burzykowski et al., 2005).
Transferability	5 of the 17 HTA bodies mention transferability of surrogate endpoints in their guidance documents. In the guidance manuals, validation is considered context specific to the population and intervention being assessed. HTA bodies will consider transferability of validity of a surrogate endpoint from previous validation studies to other disease areas and technologies based on the following: • population • interventions and mechanism of action • setting in which data is collected

Item	Description
	• disease and disease stage. If validation studies are only available in different settings, the guidance manuals suggest considering the feasibility of developing a new statistical validation model in a similar context. If a validation study includes different disease entities or interventions, then it should include an assessment of heterogeneity.

Reporting requirements	Reporting requirements are mentioned by 3 of the 17 HTA bodies. The following need to be reported:
	
	• final outcome that the surrogate endpoint replaces
	• rationale for use of the surrogate endpoint
	• biological or clinical rationale for the link between the surrogate endpoint and the final outcome
	• validation of the surrogate endpoint
	    o level of evidence for association
	    o strength of association
	    o certainty of association
	• transferability: alignment of studies used in validation with population, intervention (mechanism of action) and disease concerned in the submission
	• additional sources of uncertainty.

Considerations in economic modelling	5 of the 17 HTA bodies mention economic modelling considerations with some detailed guidance provided as summarised below by 2 HTA bodies: • NICE guidance (National Institute for Health and Care Excellence (NICE), 2023) highlights that uncertainty associated with the relationship between surrogate endpoints and the final outcomes should be quantified and presented in addition to being explored through scenario and probabilistic sensitivity analyses. • The PBAC guidance (Pharmaceutical Benefits Advisory Committee (PBAC), 2016) states that when using a proposed surrogate endpoint in an economic evaluation, it must be transformed into a final outcome. This transformation can also be used to calibrate economic models.

4.3 Statistical methods for validation of surrogate endpoints

A scoping review of published literature on statistical approaches for validating surrogate endpoints was carried out, with searches completed in December 2023 (full details in the supplementary material). The review focused on statistical methods outlined in review articles and guidance

documents published in English within the past 5 years (January 2019 to December 2023). Reference lists were scanned to identify additional papers that met the selection criteria. Of 1,286 unique records identified by the database searches, a total of 347 abstracts that were published after 2018 were screened. Of those, 29 full texts were reviewed with 3 review articles finally included (Elliott, 2023; Weir & Taylor, 2022; Zhuang & Chen, 2020). An additional review article (Ensor et al., 2016) did not meet our eligibility criteria (published pre-2019) but was included based on its relevance.

Statistical methods identified in the review articles, and their underlying assumptions, considerations and linked articles are presented in Table 3. Similar to the findings from the HTA guidance review reported in section 4.2, few criteria or thresholds were identified for determining a 'good surrogate' based on statistical measures. There is no consensus across HTA guidance and evaluation frameworks about the use of proposed criteria for asserting the acceptability of a surrogate endpoint (Thorlund et al., 2024). Concept papers, perspectives and commentaries report that conclusions about the validity of a surrogate endpoint require judgements about the quality of the relationship between the surrogate endpoint and final outcome and the uncertainty around the predicted treatment effect on the final outcome (Christensen et al., 2024; Ciani et al., 2017, 2022; Dawoud et al., 2021). It is important to note that the statistical methods described in Table 3 were developed for application to the assessment of level 1 evidence (randomised controlled trial [RCT] data with drug intervention, surrogate endpoint and final outcome in a data set). Application of these methods to real-world or other data sets would require careful consideration, particularly related to the effects of confounding between treatment and the final outcome, and between treatment and the surrogate endpoint.

As outlined in section 4.2, the context of the validation needs to be considered when using surrogate endpoints validated in previous studies. Context is of additional concern when employing methods for multiple trials since differences across trials (for example, patients' key clinical characteristics, mechanism of action of the treatments, length of follow up) can undermine the

internal validity of the evaluation. Thus, methods for multiple trials should generally be applied to settings with multiple large, high-quality RCTs that evaluate the same treatment in the same context, but such evidence is rarely available.

The synthesis of individual patient data rather than aggregate data from RCTs is the preferred approach for surrogate endpoint validation (Thorlund et al., 2024). However, individual patient data from relevant trials is often inaccessible. Apart from the 'methods for multiple trials' in Table 3 (applied in an aggregate data setting), all methods were described in the context of individual patient data.

Table 3 Methods for statistical validation of surrogate endpoints

Method name(s) and review source	Related reference article(s)	Description	Important considerations
Foundational or first methods to discuss validation of surrogate endpoints	–	–	–
Prentice criteria (Elliott, 2023; Ensor et al., 2016; Weir & Taylor, 2022; Zhuang & Chen, 2020)	(Prentice, 1989)	**Surrogacy evaluation** Model criterion for defining a surrogate endpoint. The criterion ensures that a test of the relationship between a treatment and surrogate endpoint is a valid test of the relationship between the treatment and the clinical outcome. No quantitative measure for evaluating surrogacy is provided for this method.	Considered to be an overly strict criteria for evaluating a surrogate endpoint as a perfect surrogate endpoint. Specifically developed for the setting in which both the surrogate endpoint and the final outcome are time-to-event endpoints. **Key assumptions** • No unmeasured confounding between the surrogate and the final outcome.
Proportion of treatment effect explained (logistic regression setting) (Elliott, 2023; Ensor et al., 2016; Zhuang & Chen, 2020)	(Freedman et al., 1992; Lin et al., 1997)	**Surrogacy evaluation** Quantifies the relationship using the PTE which is based on the ratio between 2 estimates of the treatment effect on the final outcome with and without adjustment for the surrogate endpoint. **Proposed criteria for assessment** Freedman et al. (Freedman et al., 1992) suggest that the lower limit of the 95% confidence interval for PTE should	Extends the Prentice Criteria to a binary final outcome setting using a logistic regression approach (extensions to other settings also exist). PTE is not a true proportion because it does not strictly lie between 0 and 1 except under strict conditions. PTE is highly variable except in situations in which the treatment effects on the final outcome are large in magnitude. **Key assumptions**

Method name(s) and review source	Related reference article(s)	Description	Important considerations
		exceed a critical value, such as 0.5 or 0.75, before declaring a surrogate as adequate.	• No unmeasured confounding between the surrogate and the final outcome. • Treatment effect is primarily mediated by the surrogate endpoint. • No interaction between treatment and surrogate. • Both the unadjusted and adjusted model (adjusted for surrogate endpoint) are correct, which is generally not possible. In some situations, it is impossible for either model to be true.
Proportion of treatment effect explained (a generalised version) (Elliott, 2023; Ensor et al., 2016; Zhuang & Chen, 2020)	(Deslandes & Chevret, 2007; Wang et al., 2020; Wang & Taylor, 2002)	**Surrogacy evaluation** Conceptually, the approach to evaluation is the same as described above. The PTE is defined by a ratio where the numerator is the change in the treatment effect on the final outcome due to the change in the surrogate induced by the treatment. The denominator is overall treatment effect for the final outcome.	Generalised version of the PTE (Freedman et al., 1992; Lin et al., 1997) described above that is defined independently from the specified modelling approach. The approach has also been extended to settings with multiple surrogate endpoints. PTE is not a true proportion unless the biological mechanism adheres to several strict conditions, for example, it may not be a true proportion if the surrogate endpoint can be influenced by adverse events due to treatment. PTE is highly variable except in situations in which the treatment effects on the final outcome are large in magnitude. **Key assumptions** • No unmeasured confounding between the surrogate and the final outcome.
Methods for multiple trials	–	–	–

Method name(s) and review source	Related reference article(s)	Description	Important considerations
Meta-analytic approach, meta-regression (Zhuang & Chen, 2020)	(Boissel et al., 1992; Daniels & Hughes, 1997; Hughes et al., 1995; Lin et al., 1997)	**Surrogacy evaluation** Quantifies the relationship based on the APEP, a weighted average of the predicted error within each trial. Smaller values of APEP compared to a clinically meaningful treatment effect on the final outcome provide support for a valid surrogate endpoint.	Requires multiple trials on the same treatment with the same surrogate and clinical outcome. Can be formulated using both a frequentist and Bayesian framework. The approach has been adapted for both continuous and binary settings. **Key assumptions** • The relationship between the treatment, surrogate endpoint, and the clinical outcome are the same across multiple trials. • Participants across trials are exchangeable. • No unmeasured confounding between the surrogate and the clinical outcome.
Meta-analytic approach, relative effects (Elliott, 2023; Ensor et al., 2016; Weir & Taylor, 2022; Zhuang & Chen, 2020)	(Alonso et al., 2004; Burzykowski & Buyse, 2006; Buyse & Molenberghs, 1998; Qu & Case, 2007)	**Surrogacy evaluation** Can be viewed as an extension of the single-trial approach to evaluation based on RE [see single trial: relative effect and adjusted association, (Buyse & Molenberghs, 1998), below]. The first formulation of this approach quantified surrogacy using a trial level R^2 given by the proportion of the variation in the total effect explained by the trial-level random effect. An alternative measure for evaluation is the LRF which can be interpreted as the amount of information gained about the clinical outcome after accounting for the surrogate endpoint.	Requires multiple trials on the same treatment with the same surrogate and final outcome. In the absence of a large number of high-quality studies that meet these criteria, the approach can result in substantial estimation error and severe loss of precision for the treatment effect on the final outcome. Evaluation metrics can be calculated under random effects or a fixed effects model formulation, but the random effects formulation is generally computationally burdensome and thus a fixed effects approach is most common. Both frequentist and Bayesian frameworks have been implemented. The method has been extended to multiple settings for the type of surrogate and final outcomes including time-to-event, bivariate, multivariate, and repeated measures approaches.

Method name(s) and review source	Related reference article(s)	Description	Important considerations
		Other measures have also been proposed such as the PIG. **Proposed criteria for assessment** Thresholds for determining a 'good surrogate' based on the metrics described above are difficult to determine and are recommended to be based on the specific context of the surrogacy evaluation. An STE is a measure that was developed to assist in evaluating the strength of the surrogate which measures the minimum value of a treatment effect on a surrogate endpoint for which the predicted effect on the final outcome would be statistically significantly different from zero.	Evaluation based on trial level R^2 has been criticised for lacking a clear interpretation outside of the setting in which both the surrogate endpoint and final outcome are continuous. A noted advantage of relying on the LRF for evaluation is that it provides a uniform interpretation across applications. **Key assumptions** - The relationship between the treatment, surrogate endpoint, and the final outcome are the same across multiple trials. - Participants across trials are exchangeable. - No unmeasured confounding between the surrogate and the final outcome. However, the multiple trial setting allows for sensitivity analysis for this assumption by estimating the probability that a subsequent trial will yield an effect on the surrogate and final outcome in the same direction. - No interaction between treatment and surrogate.
Information theoretic approach (Ensor et al., 2016; Weir & Taylor, 2022)	(Alonso & Molenberghs, 2007)	**Surrogacy evaluation** Quantifies the relationship through an R^2 measure that is interpretable as the proportion of the uncertainty in the final outcome at the individual level that is removed by adjusting for the surrogate endpoint. There are many ways to quantify this measure, one of which is the LRF used in the meta-analytic approach based on relative effects (see above).	Requires multiple trials on the same treatment with the same surrogate and final outcome. Both frequentist and Bayesian frameworks have been implemented. The method has been extended to multiple settings for the type of surrogate and final outcomes including time-to-event, bivariate and repeated measures. **Key assumptions**

Method name(s) and review source	Related reference article(s)	Description	Important considerations
			• The relationship between the treatment, surrogate endpoint, and the final outcome are the same across multiple trials. • Participants across trials are exchangeable. • No unmeasured confounding between the surrogate and the final outcome.
Causal inference methods	–	–	–
Natural indirect effects (Elliott, 2023; Ensor et al., 2016; Zhuang & Chen, 2020)	(Robins & Greenland, 1992)	**Surrogacy evaluation** Quantifies relationship as the NIE defined as the average difference in the final outcome for a fixed treatment due to the effect of the treatment on the surrogate endpoint. This is estimated relative to the total effect which is the sum of the NIE and the NDE (effect on outcome due solely to treatment). This is equivalent to the generalised PTE in certain situations.	Both frequentist and Bayesian frameworks have been implemented. The method has been extended to multiple settings for the type of surrogate and final outcomes including both linear and non-linear model settings. **Key assumptions** • No unmeasured confounding between the surrogate and the final outcome. • No interaction between treatment and surrogate endpoint (extensions to avoid this assumption have been developed).
Principle stratification, causal effect predictiveness (Elliott, 2023; Ensor et al., 2016; Weir & Taylor, 2022; Zhuang & Chen, 2020)	(Conlon et al., 2014; Frangakis & Rubin, 2002; Gilbert & Hudgens, 2008; Li et al., 2010, 2011)	**Surrogacy evaluation** Expresses the strength of a surrogate as a function of the values of the surrogate under different treatments referred to as the CEP surface. **Proposed criteria for assessment**	Initially formulated under the setting where both the final outcome and surrogate endpoint are binary. Extensions to other select settings have also been developed. Can be challenging to implement in practice due to complexity of estimation based on stringent and unverifiable assumptions. Approaches to estimation can generally be described as a trade-off between precision and plausibility of

Method name(s) and review source	Related reference article(s)	Description	Important considerations
		To assess the strength of the surrogate based on the CEP, various approaches have been proposed based on summary measures of the CEP such as the EAE and the EDE (Gilbert & Hudgens, 2008) as well as graphical approaches (Conlon et al., 2014). One metric, the PAE (Gilbert & Hudgens, 2008), which depends on both the EAE and EDE, has been proposed to have a cutoff value of 0.5 with values below indicating the surrogate is not useful.	assumptions, for example, strong and less plausible assumptions are often required for making definitive conclusions while estimation based on weak and plausible assumptions often results in imprecision of the assessment. Concerns have been raised for whether observed strength of an identified surrogate in a trial is transferable to other trials. Solutions to this concern have been developed but are currently not practically implementable. **Key assumptions** - No unmeasured confounding between the surrogate and the final outcome. Approaches have been developed to resolve this assumption but are currently not practically implementable.
Other relevant methods	–	–	–
Single trial: relative effect and adjusted association (Ensor et al., 2016; Weir & Taylor, 2022; Zhuang & Chen, 2020)	(Buyse & Molenberghs, 1998)	**Surrogacy evaluation** Quantifies relationship by an RE that is interpretable as the slope of a regression of the treatment on the final outcome against the treatment effect on the surrogate endpoint.	Formulated in the linear regression setting (for example, continuous measures of the surrogate and final outcome). **Key assumptions** - No unmeasured confounding between the surrogate and the final outcome. - Regression model is accurate based on 1 trial only, which is an untestable assumption.
Non-parametric approach	(Parast et al., 2016; Price et al., 2018; Wang et al., 2020)	**Surrogacy evaluation** Multiple approaches have been described, each of which generally focus	Unlike other methods described in this table, this approach generally avoids model-based (parametric) assumptions. The approach can be extended to combine multiple surrogate

Method name(s) and review source	Related reference article(s)	Description	Important considerations
(Elliott, 2023)		on evaluation of a surrogate based on the optimisation of mean squared error. One approach quantified the strength of the relationship by calculating the proportion of the overall treatment effect explained by the surrogate endpoint.	endpoints. The approach has been reported to avoid potential bias due to unmeasured confounders between the surrogate endpoint and final outcome.

Abbreviations: APEP, average predicted error of the predicted effect; CEP, causal effect predictiveness; EAE, expected associate effect; EDE, expected disassociate effect; LRF, likelihood reduction factor; NDE, natural direct effect; NIE, natural indirect effect; PAE, proportion associative effect; PIG, proportion of information gain; PTE, proportion of treatment effect explained; RE, relative effect; STE, surrogate threshold effect.

4.4 Experience of HTA bodies with using surrogate endpoints to inform decision making and lessons learnt

A qualitative study was carried out to understand the experience of HTA bodies and associated organisations, through 3 focus groups. A total of 29 participants took part in the study between May and June 2024. There were participants from 20 HTA bodies or associated organisations representing countries across the following geographical regions: Europe, the Americas and Asia-Pacific regions. Key themes identified during the focus group included:

- experience of surrogate endpoints in economic modelling for HTA
- challenges of using surrogate endpoints in economic modelling for HTA
- guidance on the use of surrogate endpoints in economic modelling for HTA decision making
- recommendations for future use of surrogate endpoints in economic modelling for HTA.

The themes are summarised in Table 4. Further details on methods and findings are outlined in the supplementary material.

Table 4 Summary of key themes identified from focus groups with HTA bodies and associated organisations on their experience of using surrogate endpoints

Themes	Details
Experience of surrogate endpoints in economic modelling for HTA	–
Defining surrogates	• There is a lack of clarity on defining surrogate endpoints and there is a need for a standardised approach or framework to align on this.
Frequency of encountering surrogates in evaluations	• Participants across the focus groups agreed that there is a general increase in use of surrogate endpoints regardless of how they were defined with a surrogate endpoint informing treatment effectiveness in more than 50% of evaluations.
Awareness of surrogate in the pipeline through early engagement with industry	• Participants from bodies that offer early scientific advice to technology developers highlighted how this provided a mechanism for awareness of surrogate endpoints that may be encountered a few years in the future.
Challenges of using surrogate endpoints in economic modelling for HTA	–
Lack of guidance for technology developers around use of surrogates to inform decision making	• Most of the HTA bodies have limited guidance on how to manage surrogate endpoints in their methods manuals. • There was consensus among participants that it would be useful to have more detailed guidance for evaluating clinical effectiveness based on surrogate endpoints and for its translation to the assessment of cost effectiveness.
Evaluating economic models that use surrogate endpoints	• Participants reported that: ○ in many instances, economic models are designed around the surrogate endpoint and not the condition, making their validation difficult

Themes	Details
	○ the assumption of surrogacy is sometimes embedded in multiple transitions making testing the impact of its removal complex. - As can occur in trials without surrogates, there was sometimes a disconnect between how the trial was designed or what was observed in the trial and what is being modelled.
Assessing the uncertainty arising due to the use of surrogate endpoints	- Participants highlighted that, given the opportunity cost of a decision based on a surrogate endpoint, it is crucial that the degree of uncertainty is quantified by approaches like: ○ undertaking scenario analyses including best and worst values for the surrogate endpoint, removing the surrogate endpoint entirely from the model and utilising a more established final outcome ○ value of information analysis.
Validating a surrogate endpoint	- Evidence regarding the validity of a surrogate outcome is often collected to meet regulatory approval requirements. The HTA body will still need to see and understand this evidence for their own decision making but are unlikely to draw any contradictory conclusions regarding its validity. - Participants suggested that it was key to understand what evidence and previous decisions had been made in the same technology, for the same disease or using the same surrogate endpoint. - Many participants felt that the requirements of some validation processes are complex, time-consuming and costly for technology developers and HTA bodies evaluating the evidence.
Alignment in expectations between regulators, HTA bodies and technology developers	- There can be a lack of alignment in expectations between regulators, HTA bodies and technology developers regarding the level of information required for decision making, especially relating to the surrogate endpoint. - The lack of alignment can make it difficult to properly understand the evidence around the surrogacy relationships.
Guidance on using surrogate endpoints in economic modelling in HTA decision making	–
What does future guidance need to look like?	Future guidance should be: Adaptable: to reflect the differences across HTA bodies remit, resource and capacity.

Themes	Details
	• Flexible: to apply to different situations in which surrogate endpoints may be encountered. • Consistent: guidance needs to be consistent within HTA bodies (for example, can be used across different types of evaluation) and across HTA bodies.
Patient needs and decision context	• It is important for guidance to consider what matters most to people living with the conditions and how meaningful the change based on a surrogate endpoint is from their perspective when using it in economic modelling. • The evidence available for rare conditions is less likely to meet the standards of evidence needed for other conditions and this needs to be acknowledged in guidance issued.

Standards for validation of surrogate endpoints	• Participants were keen to have more guidance on how surrogate endpoints should be validated, including:
	    o appropriate methods
	    o how the results should be presented
	    o standards on how the relationship between the surrogate endpoint and final outcome should be estimated at an individual and a trial level.
	
	• Participants from HTA bodies wanted a requirement for the submission of a systematic review to support assumptions around surrogacy of outcome.

Transparent reporting of evidence	• Participants wanted better reporting of evidence from technology developers that would help them to understand the surrogate relationships and how the economic models had been built. • They suggested that models should be set up in a way that offers flexibility to explore uncertainties related to the surrogate endpoint. • There was also a suggestion that transparent reporting is required from HTA bodies and modellers on how they are using the evidence from a surrogate endpoint, how it has been validated and how this might affect the results. This could also include the procedure of modelling, the assumptions, the sensitivity analyses performed, and their limitations and their strengths.
Database of surrogate endpoints	• Multiple participants suggested that a 'living library' of surrogate endpoints and their available evidence that is shared across HTA bodies could be useful alongside a guidance document.

Themes	Details
	• It was noted that this could be challenging because of the different contexts in which surrogate endpoints can be applied.
Recommendations for future use of surrogate endpoints in economic modelling for HTA	–

Recommendations for developing guidance on surrogate endpoints	• Participants wanted guidance that:
	    ○ is flexible and versatile so it can be applied to a wide range of scenarios, for example, using a tiered approach or an algorithm that accounts for variability in factors such as availability of evidence
	    ○ reflects the varying structures and resources of HTA bodies globally
	    ○ considers the different launch patterns of technology developers that mean some HTA bodies receive submissions later than others and hence receive different data due to changes in data availability over time
	    ○ is mindful of areas that the HTA bodies identified as particularly challenging and where they may accept lower levels of evidence about the surrogacy relationship, for example, for new technologies aimed at rare diseases.

Recommendations for developing an economic model using a surrogate endpoint	• Participants wanted recommendations in the development of an economic model that uses a surrogate endpoint to include:
	• **Model conceptualisation, design and structure**
	    ○ The model conceptualisation should be based around the disease area rather than the surrogate.
	    ○ The economic model should take precedent into account but remain flexible, avoiding reliance on previous approaches and decision making if they are no longer appropriate or applicable to the current situation.
	    ○ The use of a surrogate endpoint in an economic model should be justified and supported by data including how the context of the clinical assessment supports the validity of the surrogacy relationship.
	    ○ The model should be structured to accommodate updates with future data cuts, allowing for the integration of mature evidence on surrogate endpoints, or replacing them with final outcomes as they become available.

Themes	Details

	o The model should ideally be built so that the effect of the surrogate endpoint can be removed to allow a full exploration of its impact on the results. For example, the effect of treatment on final outcomes by the surrogate endpoint and the duration of that effect should be parameters which can be explored within the model.
	
	• **Validation of the surrogate endpoint should include**:
	  o Clinical expertise to understand the clinical relevance of the surrogate endpoint.
	  o A comprehensive review of the evidence on the relationship between the final outcome and the surrogate endpoint.
	  o Real-world evidence collection could be used as an option to validate the surrogacy relationship.
	
	• **Assessing uncertainty in the economic model**
	  o Test assumptions around relationship of surrogate endpoint to the final outcome in scenario analyses, for example, best and worst-case scenarios.
	  o Consider the use of advanced techniques such as value of information analysis to explore the benefit of reducing key uncertainties associated with the use of a surrogate endpoint.

Recommendations for reporting an economic model using a surrogate endpoint	• Participants recommended the transparent reporting of an economic model using a surrogate endpoint to understand the evidence and uncertainties associated with the use of a surrogate endpoint in economic modelling.
Recommendations for further work	• Participants highlighted the potential usefulness of a database of surrogate endpoints and the evidence to support its use. This would help bodies to validate their surrogate endpoints and potentially reduce the workload associated with the validation step.

5 Recommendations

This section makes general considerations and recommendations for model conceptualisation when considering surrogate endpoints. The recommendations are intended to complement existing best practices for model conceptualisation (Bujkiewicz et al., 2019; Caro et al., 2012; Roberts et al., 2012), drawing on the findings from the various activities conducted throughout this project. The recommendations are aimed at all interested parties involved in developing health economic models for HTA decision making. This includes, but is not limited to, technology developers, HTA decision makers and academic groups that support them. Recommendations are summarised in Table 5.

5.1 General recommendations

5.1.1 Standardised definitions for surrogate endpoints

- There is a need to use a standardised framework for defining surrogate endpoints across HTA bodies. The framework proposed by Ciani et al. (Ciani et al., 2023), which outlines criteria for categorisation of trial endpoints including biomarkers and other intermediate outcomes as surrogate endpoints, can be used.

- Efforts should also be made to align HTA definitions of surrogate endpoints with widely accepted evidence-based medicine approaches like those proposed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Working Group (The GRADE Working Group, 2013).

5.1.2 Considerations for designing studies to assess the effectiveness and safety of technologies

- If there are clinical measures to assess the impact of interventions on the disease pathway of the condition of interest, include these as endpoints in the trial even when surrogate endpoints are being considered.

- Technology developers should use existing opportunities (for example, scientific advice and early dialogue) to engage with regulators and HTA

bodies at an early stage when developing evidence generation plans to ensure that outcomes selected, and statistical analyses planned, align with the expectations of these bodies.

- When using surrogate endpoints in clinical trials, follow the following reporting guidance:
  - SPIRIT-Surrogate for trial protocols (Manyara et al., 2024a)
  - CONSORT-Surrogate for trial reports (Manyara et al., 2024b).

5.2 Recommendations when considering a surrogate endpoint in economic modelling to inform HTA decision making

5.2.1 Health economic model structure for estimating the cost effectiveness of a technology when using a surrogate endpoint

- Models should be built to reflect the disease pathway and key clinical events, not around a surrogate endpoint.
- The model structure should be driven by the decision problem or research question and not determined solely by data availability.
- The model should be built so that the effect of treatment on final outcomes by the surrogate and the duration of that effect are modifiable parameters that can be explored within the model. Additionally, it should be possible to remove the effect of the surrogate to allow a full exploration of its impact on the results.
- The model should also be designed to allow the incorporation of new information on the surrogacy relationship as it becomes available, for example, with later data cuts.
- Decision makers will look at health economic models that have already been developed and consider the new model in light of these. Previous health economic models using surrogate endpoints to estimate cost effectiveness of technologies for a condition should not be replicated if a more appropriate approach to modelling could be taken.

5.2.2 Selection and incorporation of outcomes for estimating treatment effectiveness in economic models

- If data has been collected as part of the trial on a final outcome that reflects the impact of the intervention on the disease, then that final outcome should be used in the base case (main analysis) of cost-effectiveness analysis to inform decision making. A clear justification should be provided in instances where this approach is not followed.

- If a surrogate endpoint for estimating treatment effectiveness will be used in the economic model, then its use should be justified, and the surrogacy relationship evaluated appropriately as described in section 5.2.3.

- When considering a health economic model, decision makers will discuss how it relates to clinical practice. Study endpoints used in health economic models should be practical to collect in real-world practice to reflect cost effectiveness of interventions accurately and to allow further data collection to inform decision making.

5.2.3 Validation of surrogate endpoints

- Strength of association:
    - Describe the evidence for surrogacy based on the 3 levels outlined in Ciani et al. (2017): biological plausibility of relation between surrogate endpoint and final outcome, evidence demonstrating a consistent association between the surrogate endpoint and final patient-centred outcome, and evidence demonstrating effect based on meta-analysis of several RCTs (Ciani et al., 2017).
    - If level 1 evidence is available, quantify the surrogacy relationship using an appropriate statistical approach (for example, those in Table 3 or a validated method yet to be developed). Report on the plausibility of the underlying assumptions and limitations of the approaches used.
    - If level 1 evidence is not available to evaluate the surrogacy relationship, make efforts to explore and establish the biological plausibility of the surrogate endpoint, and at a minimum, assumptions regarding the

relationship between changes in the surrogate outcome and in final outcomes described and justified.

- The correlation coefficient value(s) from the statistical validation should be reported. However, it should be noted that several measures have been proposed (refer to Table 3 for examples) and there is no consensus for any of these measures on the strength of the relationship needed for a surrogate to be considered valid.

- Expert validation:
  - The input of clinical experts on the plausibility of the surrogate endpoint as a predictor of final outcomes should be sought.
  - The input of people living with conditions under consideration should be sought regarding whether a surrogate endpoint is meaningful or not.

- Transferability of surrogacy to other contexts:
  - If considering applying evidence of surrogacy based on a previous validation study, the alignment of the following should be established:
    ◊ population
    ◊ interventions and mechanism of action
    ◊ setting in which data are collected
    ◊ disease and disease stage.
  - If a validation study includes different populations, settings, disease entities or interventions, then it should include an assessment of heterogeneity.
  - These steps should be repeated when later data cuts become available, to assess the validity of the surrogate considering further evidence.

5.2.4 Assessing uncertainty in the economic model

- At a minimum, the following should be explored and presented alongside the base-case analysis:
  - probabilistic sensitivity analyses including parameters that capture the uncertainty in the surrogate relationship
  - scenario analysis (for example, testing assumptions around relationship of surrogate to final outcome considering best and worst-case scenarios) with justification for the range selected.

- Consider the use of advanced techniques such as value of information analysis to explore the benefit of reducing key uncertainties associated with the use of a surrogate endpoint and identify situations in which further evidence is or is not justified.
- If the uncertainty in the cost-effectiveness analysis is so large that the analysis cannot be used for decision making, HTA bodies should engage with technology developers to establish a way forward, for example, through further evidence collection.

5.2.5 Reporting an economic model using a surrogate

When reporting the evidence of a surrogate relationship and its use in an economic model, use clear, non-technical language. Using the CONSORT-Surrogate extension (Manyara et al., 2024b), the following should be reported in submissions that include surrogate endpoints (EUnetHTA 21, 2023):

- The final outcome that the surrogate endpoint replaces.
- Rationale for use of the surrogate endpoint.
- Biological or medical rationale for link between final and surrogate endpoint.
- Validation of the surrogate endpoint:
  - level of evidence for association (Ciani et al., 2021)
  - strength of association
  - certainty of association.
- Transferability: alignment of studies used in validation with population, intervention (mechanism of action) and disease concerned in the submission.
- Additional uncertainties:
  - If the surrogacy relationship was not investigated in the pivotal trial, it may be noted that the true relationship between the change in the surrogate endpoint and risk of incurring the final outcome is unknown.
  - If the literature and or clinical expert opinion suggest the relationship is biologically plausible, but there is no evidence quantifying the effect, it may be noted that while the surrogacy relationship is plausible, there is

uncertainty regarding the quantification of the associated risk change as it pertains to patients with the condition under review (refer to Table 3 for approaches to quantification).

Table 5 Overview of recommendations for health technology developers

Recommendation	Details
Designing studies using surrogates	–
Choose appropriate clinical endpoints	Use widely accepted and robust clinical measures where available even where more novel outcomes are considered. If clinical measures exist to assess the impact of interventions on the disease pathway in the condition of interest, then these should be included as endpoints in the trial even when surrogate endpoints are being considered.
Engage with regulators and HTA bodies early	Engagement at an early stage will ensure that outcomes selected, and statistical analyses planned, align with the expectations of these bodies.
Report study design and results appropriately	The SPIRIT-Surrogate for trial protocols (Manyara et al., 2024a) and CONSORT-Surrogate for trial reports (Manyara et al., 2024b) should be used.
Choosing an appropriate model structure	–
Build the model to reflect the disease pathway and key clinical events	The model should not be built around a surrogate or intermediate endpoint.
Choose a model that is driven by the decision problem or research question	The model structure should not be solely determined by data availability.
Appropriately consider previous relevant economic analyses	Previous analyses can inform the model structure but should not be replicated if a more appropriate approach to modelling could be taken.
Selecting final outcomes for economic model	–
Choose an appropriate final outcome for the base-case economic analysis	If data is available on a widely accepted final outcome, this should be used in the base-case cost-effectiveness analysis.
Justify choice of a surrogate endpoint over final outcome if selected	Choice of surrogate endpoint should be justified and validated appropriately.

Recommendation	Details
	Study endpoints used in health economic models, using surrogate endpoints, should reflect cost effectiveness of interventions accurately and allow further data collection to inform decision making.
Validating a surrogate endpoint	–
Establish validity of surrogate endpoint	Use a phased approach based on the 3 levels outlined in Ciani et al. (2017) (Ciani et al., 2017).
Use an appropriate statistical validation approach	Quantify the surrogacy relationship using an appropriate statistical approach, while noting the considerations, underlying assumptions, and limitations of each approach. Report the correlation coefficient alongside HTA-recommended ranges to contextualise results. If level 1 evidence is not available, use available evidence and expert opinion to explore the biological plausibility and the relationship between the surrogate endpoint and final outcome.
Gain expert validation of surrogate	From people living with the condition on whether a surrogate endpoint is meaningful or not where appropriate. From clinical experts on the plausibility of the surrogate endpoint as a predictor of final outcomes.
If using evidence of surrogate relationship based on previous validation studies, provide sufficient information to assess the alignment	Include: • population • interventions and mechanism of action • setting in which data is collected • disease and disease stage. If a validation study includes different populations, settings, disease entities or interventions, then it should include an assessment of heterogeneity.
Assessing uncertainty in the economic model	–
As a minimum, present additional analyses to explore uncertainty	Include parameters relating to the surrogate endpoint in a probabilistic sensitivity analysis.

Recommendation	Details
	Test assumptions around relationship of surrogate to final outcome in scenario analyses.
Consider advanced methods	Value of information analysis.
Reporting	–
Fully report approach to undertaking evaluation of surrogacy relationship to inform economic models in submissions to HTA bodies	The final outcome that the surrogate replaces. Rationale for using the surrogate endpoint. Biological or medical rationale for link between final and surrogate endpoint. Validation of surrogate: - level of evidence for association (Ciani et al., 2017) - strength of association - certainty of association. Transferability: Alignment of studies used in validation with population, intervention (mechanism of action) and disease concerned in the submission. Additional uncertainties.

6 Discussion

Surrogate endpoints are increasingly being used to inform HTA decision making and reimbursement. It is important to have guidance on how to use these endpoints when assessing the clinical and cost effectiveness of technologies, particularly when the surrogate endpoint is novel without clear links to the final outcome. This includes the appropriateness of their use in trial design and economic modelling to estimate cost effectiveness (Bujkiewicz et al., 2019).

Findings from the focus groups and the reviews of regulatory and HTA guidance that we carried out confirmed that there is a lack of consistent, detailed guidance on approaches for considering surrogate endpoints across HTA and regulatory agencies for decision making. They also showed that, among HTA bodies, there has not been much change in methodological guidance since 2018 (Grigore et al., 2020; Member State Coordination Group on Health Technology Assessment, 2024). The results of our scoping review of statistical methods show that best practices about the statistical approach to validation of surrogate endpoints have not yet been established.

We have compiled an inventory of common statistical methods that may be applied in the validation of surrogate endpoints. The descriptions of these methods, accompanying references, and important considerations may be useful to health technology developers and HTA bodies who need to make decisions when available evidence is only for a surrogate endpoint, or to inform future work on best practices for validation of surrogate endpoints. To date, there has not been widespread agreement across HTA guidelines and evaluation frameworks about the use of proposed criteria for asserting the acceptability of a surrogate endpoint (Thorlund et al., 2024). It has instead been suggested that drawing conclusions about the validity of a surrogate endpoint requires judgements about the quality of the relationship between the surrogate and final outcome and the uncertainty around the predicted treatment effect on the final outcome (Christensen et al., 2024; Ciani et al.,

2017, 2022; Dawoud et al., 2021; Member State Coordination Group on Health Technology Assessment, 2024).

We recognise that in some cases, it may be challenging to demonstrate validity of the surrogate relationship based on data from multiple trials. However, we would encourage technology developers to interrogate the surrogacy relationship extensively and provide the best available data linking the surrogate endpoints to final outcomes before using them in submissions.

7 Conclusion

In this paper, we provide guidance for economic model conceptualisation to inform HTA decision making when using surrogate endpoints to estimate treatment effectiveness. This guidance is to be used alongside other established best practices for economic modelling in HTA (Bujkiewicz et al., 2019; Caro et al., 2012; Roberts et al., 2012).

8 References

Alonso, A., & Molenberghs, G. (2007). Surrogate marker evaluation from an information theory perspective. *Biometrics*, *63*(1), 180–186. https://doi.org/10.1111/j.1541-0420.2006.00634.x

Alonso, A., Molenberghs, G., Burzykowski, T., Renard, D., Geys, H., Shkedy, Z., Tibaldi, F., Abrahantes, J. C., & Buyse, M. (2004). Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics*, *60*(3), 724–728. https://doi.org/10.1111/j.0006-341X.2004.00222.x

Boissel, J. P., Collet, J. P., Moleur, P., & Haugh, M. (1992). Surrogate endpoints: a basis for a rational approach. *Eur J Clin Pharmacol*, *43*(3), 235–244. https://doi.org/10.1007/bf02333016

Bujkiewicz, S., Achana, F., Papanikos, T., Riley, R. ., & Abrams, K. . (2019). *NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints*. https://www.sheffield.ac.uk/nice-dsu/tsds/full-list

Burzykowski, T., & Buyse, M. (2006). Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat*, *5*(3), 173–186. https://doi.org/10.1002/pst.207

Burzykowski, T., Molenberghs, G., & Buyse, M. (2005). *The evaluation of surrogate endpoints.* . Springer.

Buyse, M., & Molenberghs, G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*, *54*(3), 1014–1029.

Caro, J. J., Briggs, A. H., Siebert, U., Kuntz, K. M., & Force, I.-S. M. G. R. P. T. (2012). Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Medical Decision Making : An International Journal of the Society for*

Medical Decision Making, 32(5), 667–677.
https://doi.org/10.1177/0272989X12454577

Christensen, R., Ciani, O., Manyara, A., & Taylor, R. (2024). Surrogate
endpoints: a key concept in clinical epidemiology. *Clinical Epidemiology*,
1(167).

Ciani, O., Buyse, M., Drummond, M., Rasi, G., Saad, E. D., & Taylor, R. S.
(2017). Time to review the role of surrogate end points in health policy:
state of the art and the way forward. *Value in Health*, *20*(3), 487–495.
https://doi.org/10.1016/j.jval.2016.10.011

Ciani, O., Grigore, B., Blommestein, H., de Groot, S., Möllenkamp, M., Rabbe,
S., & Daubner-Bendes, R Taylor, R. (2021). Validity of Surrogate
Endpoints and Their Impact on Coverage Recommendations: A
Retrospective Analysis across International Health Technology
Assessment Agencies. *Medical Decision Making.*, *41*(4), 439–452.

Ciani, O., Grigore, B., & Taylor, R. S. (2022). Development of a framework
and decision tool for the evaluation of health technologies based on
surrogate endpoint evidence. *Health Economics*, *31 Suppl 1*, 44–72.
https://doi.org/https://dx.doi.org/10.1002/hec.4524

Ciani, O., Manyara, A., Davies, P., Stewart, D., Weir, C., Young, A., Blazeby,
J., Butcher, N., Bujkiewicz, S., Chan, A., & Dawoud, D. (2023). A
framework for the definition and interpretation of the use of surrogate
endpoints in interventional trials. *EClinicalMedicine.*, *1*(65).

Conlon, A. S., Taylor, J. M., & Elliott, M. R. (2014). Surrogacy assessment
using principal stratification when surrogate and outcome measures are
multivariate normal. *Biostatistics*, *15*(2), 266–283.
https://doi.org/10.1093/biostatistics/kxt051

Daniels, M. J., & Hughes, M. D. (1997). Meta-analysis for the evaluation of
potential surrogate markers. *Stat Med*, *16*(17), 1965–1982.
https://doi.org/10.1002/(sici)1097-0258(19970915)16:17<1965::aid-

sim630>3.0.co;2-m

Dawoud, D., Naci, H., Ciani, O., & Bujkiewicz, S. (2021). Raising the bar for using surrogate endpoints in drug regulation and health technology assessment. *BMJ*, *374*, n2191. https://doi.org/10.1136/bmj.n2191

Deslandes, E., & Chevret, S. (2007). Assessing surrogacy from the joint modelling of multivariate longitudinal data and survival: application to clinical trial data on chronic lymphocytic leukaemia. *Statistics in Medicine*, *26*(30), 5411–5421. https://doi.org/https://doi.org/10.1002/sim.3142

Elliott, M. R. (2023). Surrogate endpoints in clinical trials. *Annual Review of Statistics and Its Application*, *10*, 75–96.

Ensor, H., Lee, R. J., Sudlow, C., & Weir, C. J. (2016). Statistical approaches for evaluating surrogate outcomes in clinical trials: a systematic review. *Journal of Biopharmaceutical Statistics*, *26*(5), 859–879. https://doi.org/https://dx.doi.org/10.1080/10543406.2015.1094811

EUnetHTA 21. (2023). *Individual Practical Guideline Document. D4.4-Outcomes (Endpoints)*.

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*, 21–29.

Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Stat Med*, *11*(2), 167–178. https://doi.org/10.1002/sim.4780110204

Gilbert, P. B., & Hudgens, M. G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics*, *64*, 1146–1154.

Grigore, B., Ciani, O., Dams, F., Federici, C., de Groot, S., Möllenkamp, M., Rabbe, S., Shatrov, K., Zemplenyi, A., & Taylor, R. S. R. (2020). Surrogate endpoints in health technology assessment: an international review of methodological guidelines. *Pharmacoeconomics*, *38*(10), 1055–1070. https://doi.org/10.1007/s40273-020-00935-1

Hughes, M. D., Dergruttolla, V., & Welles, S. L. (1995). Evaluating surrogate markers. *J Acquir Immune Defic Syndr Hum Retrovirol*, *10*.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2011). *Aussagekraft von surrogatendpunkten in der onkologie. [Validity of surrogate parameters in oncology (Rapid report)]. IQWiG-Berichte 80*.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2022). *Allgemeine Methoden Entwurf für Version 7.0 [English translation: General Methods Draft version 7.0]*.

Lesko, L., & Atkinson, A. (2001). Use of Biomarkers and Surrogate Endpoints in Drug Development and Regulatory Decision Making: Criteria, Validation, Strategies. *Annu Rev Pharmacol Toxicol*, *41*, 347–366. https://www.annualreviews.org/content/journals/10.1146/annurev.pharmtox.41.1.3

Li, Y., Taylor, J. M., & Elliott, M. R. (2010). A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics*, *66*(2), 523–531. https://doi.org/10.1111/j.1541-0420.2009.01303.x

Li, Y., Taylor, J. M. G., Elliott, M. R., & Sargent, D. J. (2011). Causal assessment of surrogacy in a meta-analysis of colorectal cancer trials. *Biostatistics*, *12*(3), 478–492. https://doi.org/10.1093/biostatistics/kxq082

Lin, D. Y., Fleming, T. R., & De Gruttola, V. (1997). Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med*, *16*(13), 1515–1527. https://doi.org/10.1002/(sici)1097-0258(19970715)16:13<1515::aid-sim572>3.0.co;2-1

Manyara, A., Davies, P., Stewart, D., Weir, C., Young, A., Blazeby, J., Butcher, N., Bujkiewicz, S., Chan, A., & Dawoud, D. (2024a). Reporting of surrogate endpoints in randomised controlled trial protocols (SPIRIT-Surrogate): extension checklist with explanation and elaboration. *BMJ*, *9*(386).

Manyara, A., Davies, P., Stewart, D., Weir, C., Young, A., Blazeby, J., Butcher, N., Bujkiewicz, S., Chan, A., Dawoud, D., & Offringa, M. (2024b). Reporting of surrogate endpoints in randomised controlled trial reports (CONSORT-Surrogate): extension checklist with explanation and elaboration. *BMJ*, *9*(386).

Member State Coordination Group on Health Technology Assessment. (2024). *Guidance on outcomes for joint clinical assessments*. https://health.ec.europa.eu/document/download/a70a62c7-325c-401e-ba42-66174b656ab8_en?filename=hta_outcomes_jca_guidance_en.pdf

National Institute for Health and Care Excellence (NICE). (2023). *NICE health technology evaluations: the manual*. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741

Parast, L., McDermott, M. M., & Tian, L. (2016). Robust estimation of the proportion of treatment effect explained by surrogate marker information. *Stat Med*, *35*(10), 1637–1653. https://doi.org/10.1002/sim.6820

Perelman, J., Soares, M., Mateus, C., Duarte, A., Faria, R., Ferreira, L., Saramago, P., Veiga, P., Furtado, C., Caldeira, S., Teixera, M. do C., & Sculpher, M. (2019). *Methodological Guidelines for Economic Evaluation Studies.*

Pharmaceutical Benefits Advisory Committee (PBAC). (2016). *The Pharmaceutical Benefits Advisory Committee Guidelines. Appendix 5: Translating comparative treatment effects of proposed surrogate measures to target clinical outcomes.* https://pbac.pbs.gov.au/appendixes/appendix-5.html

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*, *8*(4), 431–440. https://doi.org/10.1002/sim.4780080407

Price, B. L., Gilbert, P. B., & van der Laan, M. J. (2018). Estimation of the

optimal surrogate based on a randomized trial. *Biometrics*, *74*(4), 1271–1281. https://doi.org/10.1111/biom.12879

Qu, Y., & Case, M. (2007). Quantifying the effect of the surrogate marker by information gain. *Biometrics*, *63*(3), 953–958. https://doi.org/10.1111/j.1541-0420.2007.00852_1.x

Radu, P., Kumar, G., Cubi-Molla, P., Garau, M., Bell, E., Pan, J., Gilardino, R., Van Bavel, J., Brandtmüller, A., Nelson, K., & Goodall, M. (2024). Navigating change: a comparative analysis across health technology assessment agencies on their positions on five key topics. *OHE Contract Research Report*. https://www.ohe.org/publications/how-have-hta-agencies-evolved-their-methods/

Roberts, M., Russell, L. B., Paltiel, A. D., Chambers, M., McEwan, P., & Krahn, M. (2012). Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force–2. *Medical Decision Making*, *32*(5), 678–689.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*(2), 143–155. https://doi.org/10.1097/00001648-199203000-00013

Schuster Bruce, C., Brhlikova, P., Heath, J., & McGettigan, P. (2019). he use of validated and nonvalidated surrogate endpoints in two European Medicines Agency expedited approval pathways: a cross-sectional study of products authorised 2011–2018. *PLoS Medicine*, *16*(9).

The GRADE Working Group, 2013. (2013). GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. In H. Schünemann, J. Brożek, G. Guyatt, & A. Oxman (Eds.), *GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013*. https://guidelinedevelopment.org/handbook

Thorlund, K., Shephard, C., Machado, L., Bourgouin, T., Hudson, L., Ting, E.,

Dempster, W., & Bick, R. (2024). Adapting health technology assessment agency standards for surrogate outcomes in early stage cancer trials: what needs to happen? *Expert Rev Pharmacoecon Outcomes Res*, *24*(3), 331–342. https://doi.org/10.1080/14737167.2024.2302431

U.S. Food and Drug Administration. (2018). *Surrogate Endpoint Resources for Drug and Biologic Development*. https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development

U.S. Food and Drug Administration. (2022). *Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure*. Development Resources. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure

U.S. Food and Drug Administration, Bethesda: National Institutes of Health, & Silver Springs. (2016). *6 BEST (Biomarkers, EndpointS, and other Tools) resource FDA-NIH biomarker working group*. https://www.ncbi.nlm.nih.gov/books/NBK326791/

Wang, X., Parast, L., Tian, L. U., & Cai, T. (2020). Model-free approach to quantifying the proportion of treatment effect explained by a surrogate marker. *Biometrika*, *107*(1), 107–122. https://doi.org/10.1093/biomet/asz065

Wang, Y., & Taylor, J. M. (2002). A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics, 58*(4), 803–812. https://doi.org/10.1111/j.0006-341x.2002.00803.x

Weir, C. J., & Taylor, R. S. (2022). Informed decision-making: statistical methodology for surrogacy evaluation and its role in licensing and reimbursement assessments. *Pharmaceutical Statistics*, *21*(4), 740–756. https://doi.org/https://dx.doi.org/10.1002/pst.2219

Zhuang, R., & Chen, Y. Q. (2020). Measuring surrogacy in clinical research: with an application to studying surrogate markers for HIV treatment-as-

prevention. Statistics in Biosciences, 12(3), 295–323. https://doi.org/https://dx.doi.org/10.1007/s12561-019-09244-4

9 Contributors to the report

9.1 Task leads

Fatima Salih, National Institute for Health and Care Excellence, UK

Laura Flight, National Institute for Health and Care Excellence, UK

Zoe Garrett, National Institute for Health and Care Excellence, UK

Shane Collins, National Institute for Health and Care Excellence, UK

Farah Husein, Canada's Drug Agency, Canada

Mary-Ellen Hogan, Canada's Drug Agency, Canada

Eric Morenz, Canada's Drug Agency, Canada

Matthew Bryan, Canada's Drug Agency, Canada

David Kaunelis, Canada's Drug Agency, Canada

Michelle Gates, Canada's Drug Agency, Canada

Michael Coory, Australian Department of Health and Aged Care, Australia

Phyo Thu Zar Aung, Australian Department of Health and Aged Care, Australia

Dan Ollendorf, Institute for Clinical and Economic Review, US

Marina Richardson, Institute for Clinical and Economic Review, US

9.2 Working group members

Yelan Guo, National Institute for Health and Care Excellence, UK

Dalia Dawoud, National Institute for Health and Care Excellence, UK

Jacoline Bouvy, National Institute for Health and Care Excellence, UK

Jeanette Kusel, National Institute for Health and Care Excellence, UK

Fay McCracken, National Institute for Health and Care Excellence, UK

Saskia Knies, National Health Care Institute, Netherlands

Reva Efe, National Health Care Institute, Netherlands

Nicole Mittmann, Canada's Drug Agency, Canada

Karen Lee, Canada's Drug Agency, Canada

Lucia Fiestas Navarrete, Canada's Drug Agency, Canada

Sean Tunis, Rubix Health, US

Adriana Robayo, Institute for Technology Assessment in Health, Colombia

10 Acknowledgements

We would like to thank the focus group participants who gave up time from their busy schedules to take part in the study. Their insights were invaluable and greatly appreciated.

We would also like to thank members from the Medical Research Council and National Institute for Health and Care Research (MRC-NIHR) Trials Methodology Research Partnership (TMRP) surrogate endpoints specialist interest group who provided feedback on the report.