



Canada's Drug Agency  
L'Agence des médicaments du Canada  
Drugs, Health Technologies and Systems. Médicaments, technologies de la santé et systèmes.



NICE National Institute for  
Health and Care Excellence



Zorginstituut Nederland

## Surrogate endpoints in cost-effectiveness analysis for use in health technology assessment: supplementary material

This supplementary material provides the full methodological details and findings for each task outlined in the [main report](#) that aim to develop a set of recommendations on how to address surrogate endpoints in health economic models.

## Table of contents

A.	Review of regulatory standards for validating surrogate endpoints.....	4
A.1	Objectives.....	4
A.2	Methods.....	4
A.2.1	Search strategy .....	4
A.2.2	Document selection .....	4
A.3	Findings.....	5
A.3.1	Definitions.....	10
A.3.2	Categories of surrogate endpoints.....	10
A.3.3	Use of surrogate endpoints in regulatory decision making .....	11
A.4	Summary .....	12
A.5	References .....	12
A.6	Additional materials .....	14
A.6.1	Search outputs .....	14
B.	Review of HTA standards for surrogate endpoints .....	19
B.1	Objectives.....	19
B.2	Methods.....	19
B.2.1	Literature review .....	19
B.2.2	Data extraction .....	20
B.2.3	Outline of analysis .....	21
B.3	Findings.....	21
B.3.1	Overall acceptability of surrogate endpoints .....	24
B.3.2	Strength of evidence.....	24
B.3.3	Validation process, methods and values .....	25
B.3.4	Transferability .....	25
B.3.5	Reporting requirements .....	26
B.3.6	Considerations in economic modelling .....	27
B.4	Summary .....	27
B.5	References .....	28
B.6	Additional materials .....	33
C.	Review of statistical methods for validation of surrogate endpoints ....	50
C.1	Objectives.....	50
C.2	Methods.....	50

Surrogate endpoints in health technology assessment – Supplementary material

C.2.1	Literature review .....	50
C.2.2	Selection criteria .....	50
C.2.3	Search strategy .....	51
C.2.4	Selection of documents .....	52
C.2.5	Data extraction, analysis and presentation .....	52
C.3	Findings .....	53
C.4	Discussion .....	61
C.5	Summary .....	63
C.6	References .....	64
C.7	Supplementary materials .....	68
C.7.1	Search strategy .....	68
C.7.2	Excluded studies list .....	70
D.	Surrogate endpoints in cost-effectiveness analysis for use in health technology assessment – a qualitative analysis .....	74
D.1	Objectives .....	74
D.2	Methods .....	74
D.2.1	Outline of analysis .....	75
D.3	Results .....	76
D.3.1	Experience of using surrogate endpoints in economic modelling for HTA	76
D.3.2	Challenges of using surrogate endpoints in economic modelling for HTA	78
D.3.3	Guidance on the use of surrogate endpoints in economic modelling in HTA decision making .....	80
D.3.4	Suggested recommendations for future use of surrogate endpoints in economic modelling for HTA .....	85
D.3.5	Recommendations for further work .....	87
D.4	Summary .....	88
D.5	References .....	88

## **A. Review of regulatory standards for validating surrogate endpoints**

### **A.1 Objectives**

A scoping review was undertaken to identify the standards for validating surrogate endpoints in the context of regulatory decision making key concepts and definitions, and to identify and map the standards used by regulatory agencies for the validation of surrogate endpoints (Munn et al., 2018).

### **A.2 Methods**

#### **A.2.1 Search strategy**

To identify the regulatory standards used in validating surrogate endpoints, regulatory guidance documents were searched using Google Scholar and a direct search on the websites of the United States (US) Food and Drug Administration (FDA), the European Medicines Agency (EMA), and International Council on Harmonisation (ICH) website, using the search terms: {endpoint or surrogate or biomarker}, “regulatory standard” and {endpoint or surrogate or biomarker}, “regulatory guidance” and {accelerated or conditional or provisional} and “regulatory standard” and {accelerated or conditional or provisional}. To supplement the direct web search of the regulatory guidance, we conducted a systematic search of Ovid MEDLINE for articles published in English within the past 5 years (January 2019 to February 2024). Search terms: “regulatory guidance” and {endpoint or surrogate or biomarker} were used to filter the titles and the abstracts. Upon identification of abstracts that mentioned these search terms, the corresponding full texts were examined to identify guidance by regulatory agencies. If the full texts referenced such regulatory guidance, those documents were then pursued through the reference list provided.

#### **A.2.2 Document selection**

Two reviewers (MC and PA) performed the direct search of regulatory websites and the database search, and a single reviewer (PA) manually screened the reference lists of relevant articles. The two reviewers discussed

Surrogate endpoints in health technology assessment – Supplementary material

and agreed the final selection criteria. Documents were selected for inclusion if they contained specific guidance from regulatory agencies on the evaluation and application of surrogate endpoints in the context of regulatory decision making. Documents that did not contain direct guidance from regulatory authorities were considered not relevant to this review.

### **A.3 Findings**

There was a total of 38 hits from the database search, of those 3 were included and 35 were excluded. Across the database search and the regulatory website search, a total of 8 documents were included in the review (Table A-1). Of those, 6 guidance documents primarily discuss the criteria and considerations for biomarker or 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, while 2 documents served as case references pertinent to this research question.

**Table A-1. Overview of regulatory guidance documents**

Country or region	Agency	Document reviewed	Key information considered or extracted	Source of document reviewed
US	Food and Drug Administration (FDA)	Surrogate Endpoint Resources for Drug and Biologic Development, 2018 (U.S. Food and Drug Administration, 2018c)	Definitions and categories of surrogate endpoints.	FDA website
US	Food and Drug Administration (FDA)	Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff, 2018 (U.S. Food and Drug Administration, 2018a)	There are no set quantitative criteria for determining whether the relationship between the biomarker and the clinical outcome is sufficiently strong to support biomarker qualification. Criteria based on parameters used to quantify the relationship, such as the threshold values for sensitivity and specificity, and the presence of a gradient (for example, clinical performance change as function of biomarker quantity) can provide confidence that a finding is likely to be relevant, reliable, and statistically robust. Additional considerations that support the biomarker's association with the clinical outcome should also be assessed, such as whether there is a strong biological rationale supporting the role of the biomarker in the proposed context of use and whether the findings are supported by more than one investigation or analysis set or there are multiple lines of evidence (for example, experimental models and human studies).	Huang Y & Yuan J (2024). Improvement of assessment in surrogate endpoint and safety outcome of single-arm trials for anticancer drugs. Expert Review of Clinical Pharmacology, 17(5-6), 477-487. (Huang & Yuan, 2024)
US	Food and Drug Administration (FDA)	Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry, 2018 (U.S. Food and Drug Administration, 2018b)	To establish the clinical effectiveness, at least 2 adequate and well-controlled clinical investigations are needed. The FDA may consider data from 1 adequate and well-controlled clinical investigation and confirmatory evidence as substantial evidence, if the FDA determines that such data and evidence are sufficient to establish effectiveness.	Zhang J, Pilar MR, Wang X, Liu J, Pang H, Brownson RC, He J (2020). Endpoint surrogacy in oncology phase 3 randomised

Country or region	Agency	Document reviewed	Key information considered or extracted	Source of document reviewed
				controlled trials. British Journal of Cancer, 123(3), 333-334. (Zhang et al., 2020)
US	Food and Drug Administration (FDA)	Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance for Industry, 2019 (U.S. Food and Drug Administration, 2019)	<p>In contrast to traditional approval, accelerated approval can be based on a demonstrated effect on a surrogate endpoint that is reasonably likely to predict a clinical benefit but where there are not sufficient data to show that it is a validated surrogate endpoint. Effects on intermediate clinical endpoints can also be a basis for accelerated approval. For drugs granted accelerated approval, the FDA requires post-approval trials to verify the predicted clinical benefit.</p> <p>The FDA exercises its broad scientific judgement in applying the evidentiary approval standards to drugs for life-threatening and severely debilitating diseases, especially where there is no satisfactory alternative therapy. In addition, the accelerated approval regulations built upon this recognition by acknowledging that reliance on a surrogate endpoint “almost always introduces some uncertainty into the risk/benefit assessment, because clinical benefit is not measured directly and the quantitative relation of the effect on the surrogate to the clinical effect is rarely known”. Together these regulations recognise the importance of facilitating the development of, and access to, safe and effective treatment options for life-threatening and severely debilitating diseases with unmet medical needs. This approach has been reinforced by the FDA’s interactions with patients and their caregivers who describe their willingness to accept less certainty about effectiveness in return for earlier access to much needed medicines. For example, for a life-threatening disease without any available treatment, the FDA might accept the results of adequate and well-controlled investigations with less rigorous designs, such as a historically controlled study.</p>	Rizk JG & Lewin JC (2023). FDA’s dilemma with the aducanumab approval: public pressure and hope, surrogate markers and efficacy, and possible next steps. BMJ Evidence-based Medicine, 28(2), 78-82. (Rizk & Lewin, 2023)

Country or region	Agency	Document reviewed	Key information considered or extracted	Source of document reviewed
US	Food and Drug Administration (FDA)	BEST (Biomarkers, EndpointS, and other Tools) Resource, 2020 (FDA-NIH Biomarker Working Group, 2021)	Discussed definitions and categories of surrogates and their use in clinical trials. The document included "Off-target" effect found in the Cardiac Arrhythmia Suppression Trials (CAST 1 and 2).	FDA-NIH Biomarker Working Group
Europe	European Medicines Agency (EMA)	Qualification Opinion for GFR slope as a Validated Surrogate Endpoint for RCT in CKD, 2023 (European Medicines Agency (EMA), 2023)	Key aspects of validation of a surrogate endpoint used in this study included: biological plausibility, individual-level associations and trial-level analyses.	EMA website
Canada	Health Canada	Guidance Document: Notice of Compliance with Conditions (NOC/c), 2016 (Health Canada, 2005)	The decision to authorise or reject a product for market authorisation is complex and based on numerous factors governed by regulatory requirements. The basis for NOC/c decision is not the surrogate marker status but rather the body of evidence supporting the efficacy of the product in the drug submission.	Health Canada website
Australia	Therapeutic Goods Administration (TGA)	Provisional determination eligibility criteria, 2021 (Therapeutic Goods Administration (TGA), 2021)	For a provisional determination and subsequent provisional registration application, the clinical data available may be limited. For example, data on final outcomes such as morbidity and mortality may not be available yet and results may be based on surrogate endpoints that are reasonably likely to predict clinical benefit. The scientific evidence may therefore be less comprehensive than would typically be required but needs to be adequate and convincing evidence based on clinical trials (usually randomised controlled trials). Where a surrogate endpoint is used it should be recognised to be reasonably likely to predict an effect on clinical outcomes that establish direct clinical benefit (for example, morbidity and mortality). A surrogate endpoint does not need to be validated (for	TGA website



Country or region	Agency	Document reviewed	Key information considered or extracted	Source of document reviewed
			<p>example, be an endpoint that is known to predict clinical benefit and could be used for standard approval) but needs empirical evidence to support that it is reasonably likely to predict direct clinical benefit. Whether an endpoint is reasonably likely to predict clinical benefit (such as the suitability of the endpoint) will be determined on a case-by-case basis. Appropriateness of a surrogate endpoint in one condition may not necessarily be appropriate for a different condition.</p>	

### **A.3.1 Definitions**

1. A clinical endpoint is a clinically meaningful measure of how a patient feels, functions, or survives (U.S. Food and Drug Administration, 2018c). Ideally, regulatory agencies require randomised data on a final patient-relevant endpoint such as overall survival or perhaps a morbidity endpoint (for example, stroke or myocardial infarction). Quality of life is sometimes mentioned as a final patient-relevant endpoint but seems to be rarely used as the pre-specified primary endpoint in pivotal regulatory trials for marketing approval due to missing data-points and concerns about the lack of sensitivity and specificity of quality-of-life measurement tools (U.S. Food and Drug Administration, 2018c).
2. A biological marker is a physical sign or laboratory measurement that occurs in association with a pathological process and that has putative diagnostic and or prognostic utility (U.S. Food and Drug Administration, 2018c). Biomarkers may include molecular, histologic, radiographic, or physiological characteristics (U.S. Food and Drug Administration, 2018c).
3. A surrogate endpoint is 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 (U.S. Food and Drug Administration, 2018c). The correlation between the surrogate endpoint and the clinical benefit may be based on the following hierarchy: randomised data, epidemiologic or non-randomised data, mechanistic or pathophysiological reasoning, and other scientific evidence (U.S. Food and Drug Administration, 2018c).

### **A.3.2 Categories of surrogate endpoints**

The FDA (U.S. Food and Drug Administration, 2018c) has defined 3 categories of surrogate endpoints, depending on the level of validation: candidate surrogate endpoint, reasonably likely surrogate endpoint and validated surrogate endpoint.

Candidate surrogate endpoint 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.

Reasonably likely surrogate endpoint is 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. This may be used for accelerated approval for medicines (partway through the clinical development program) 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.

Validated surrogate endpoint 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. They 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.

### **A.3.3 Use of surrogate endpoints in regulatory decision- making**

The FDA (U.S. Food and Drug Administration, 2018c) outlined that surrogate endpoints may be used:

- to address various challenging situations in which conducting clinical endpoints studies would be considered unethical and direct patient-relevant outcomes would require prolonged studies
- when accelerated approval for dire diseases is considered to expedite access to promising treatments and there is an unmet clinical need
- to assess harm (for example, Hy's law as a predictor of hepatic toxicity).

A (US 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 the table. This table represents a good resource for drug developers considering pursuing 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.

## A.4 Summary

As is the case with validation of surrogates 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.

## A.5 References

European Medicines Agency (EMA). (2023). Qualification Opinion for GFR slope as a Validated Surrogate Endpoint for RCT in CKD.

<https://www.ema.europa.eu/en/documents/other/quali>.

FDA-NIH Biomarker Working Group. (2021). BEST (Biomarkers, EndpointS, and other Tools) Resource (Internet). Silver Spring (MD): Food and Drug Administration (US) (p. <https://www.ncbi.nlm.nih.gov/books/NBK338449/>, acc).

Health Canada. (2005). Guidance Document. Notice of Compliance with conditions (NOC/c).

Huang, Y., & Yuan, J. (2024). Improvement of assessment in surrogate endpoint and safety outcome of single-arm trials for anticancer drugs. *Expert Review of Clinical Pharmacology*, 17(5–6), 477–487.

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18, 1–7.

Rizk, J. G., & Lewin, J. C. (2023). FDA's dilemma with the aducanumab approval: public pressure and hope, surrogate markers and efficacy, and possible next steps. *BMJ Evidence-Based Medicine*, 28(2), 78–82.

Therapeutic Goods Administration (TGA). (2021). Provisional determination eligibility criteria. <https://www.tga.gov.au/sites/default/files/provisi>.

US Food and Drug Administration. (2018a). Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff DRAFT GUIDANCE. FDA. FDA-2018-D-4267, <https://www.fda.gov/regulatory-information/search->.

US Food and Drug Administration. (2018b). Clinical trial endpoints for the approval of cancer drugs and biologics: guidance for industry. US Food and Drug Administration, <https://www.fda.gov/regulatory-information/search->.

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

US Food and Drug Administration. (2019). Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstrating-substantial-evidence-effectiveness-human-drug-and-biological-products>

Zhang, J., Pilar, M. R., Wang, X., Liu, J., Pang, H., Brownson, R. C., Colditz, G. A., Liang, W., & He, J. (2020). Endpoint surrogacy in oncology phase 3 randomised controlled trials. *British Journal of Cancer*, 123(3), 333–334.

## A.6 Additional materials

### A.6.1 Search outputs

#### A.6.1.1 Overview

**Databases searched:** Ovid MEDLINE(R)

**Date of search:** February 15, 2024

**Alerts:** None

**Search filters applied:** N/A

**Limits**

- Publication date limit: 2019 to current
- Language limit: English
- Limited to abstracts

#### Ovid MEDLINE(R) ALL 1946 to February 15, 2024

#	Searches	Results
1	(surrogate and (outcome* or endpoint* or biomarker*))	367407
2	limit 1 to (abstracts and english language and yr="2019 -Current")	104179
3	Regulatory guidance or Drug Approval	17314
4	limit 3 to (abstracts and english language and yr="2019 -Current")	2097
5	2 and 4	38

#### A.6.1.2 Articles with regulatory guidance included from search (n=3)

Huang Y & Yuan J (2024). Improvement of assessment in surrogate endpoint and safety outcome of single-arm trials for anticancer drugs. *Expert Review of Clinical Pharmacology*, 17(5-6), 477-487.

<https://dx.doi.org/10.1080/17512433.2024.2344669>

Rizk JG & Lewin JC (2023). FDA's dilemma with the aducanumab approval: public pressure and hope, surrogate markers and efficacy, and possible next steps. *BMJ Evidence-based Medicine*, 28(2), 78-82. <https://dx.doi.org/10.1136/bmjebm-2022-111914>

Zhang J, Pilar MR, Wang X, Liu J, Pang H, Brownson RC, He J (2020). Endpoint surrogacy in oncology phase 3 randomised controlled trials. *British Journal of Cancer*, 123(3), 333-334. <https://dx.doi.org/10.1038/s41416-020-0896-5>

Surrogate endpoints in health technology assessment – Supplementary material

### A.6.1.3 Excluded articles (n=35)

Alabaku O, Laffey TN, Suh K & Li M (2022). Trends in endpoint use in pivotal trials and efficacy for us food and drug administration-approved solid tumor therapies, 1995-2021. *Journal of Managed Care & Specialty Pharmacy*, 28(11), 1219-1223. <https://dx.doi.org/10.18553/jmcp.2022.28.11.1219>

Allen PB, Bonovas S, Danese S & Peyrin-Biroulet L (2020). Evolving primary and secondary endpoints in randomized controlled trials leading to approval of biologics and small molecules in ibd: an historical perspective. *Expert Opinion on Biological Therapy*, 20(2), 151-161. <https://dx.doi.org/10.1080/14712598.2020.1702020>

Bakker E, Starokozhko V, Kraaijvanger JWM, Heerspink HJL & Mol PGM (2023). Precision medicine in regulatory decision making: biomarkers used for patient selection in European public assessment reports from 2018 to 2020. *Clinical and translational science*, 16(11), 2394-2412. <https://dx.doi.org/10.1111/cts.13641>

Ballreich J, Socal M, Bennett CL, Xuan A, Trujillo A & Anderson G (2023). Accelerated approval drug labels often lack information for clinical decision-making. *Pharmacotherapy: The Journal of Human Pharmacology & Drug Therapy*, 43(4), 300-304. <https://dx.doi.org/10.1002/phar.2789>

Chen EY, Joshi SK, Tran A & Prasad V (2019). Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. *JAMA Internal Medicine*, 179(5), 642-647. <https://dx.doi.org/10.1001/jamainternmed.2018.8351>

Chen EY, Raghunathan V & Prasad V (2019). An overview of cancer drugs approved by the us food and drug administration based on the surrogate end point of response rate. *JAMA Internal Medicine*, 179(7), 915-921. <https://dx.doi.org/10.1001/jamainternmed.2019.0583>

Conforti F, Pala L, Bagnardi V, De Pas T, Colleoni M, Buyse M, Gelber RD (2022). Surrogacy of pathologic complete response in trials of neoadjuvant therapy for early breast cancer: critical analysis of strengths, weaknesses, and misinterpretations. *JAMA Oncology*, 8(11), 1668-1675. <https://dx.doi.org/10.1001/jamaoncol.2022.3755>

Surrogate endpoints in health technology assessment – Supplementary material

Florian J, Sun Q, Schrieber SJ, White R, Shubow S, Johnson-Williams BE, Strauss DG (2023). Pharmacodynamic biomarkers for biosimilar development and approval: a workshop summary. *Clinical Pharmacology & Therapeutics*, 113(5), 1030-1035. <https://dx.doi.org/10.1002/cpt.2795>

Gafter-Gvili A, Tibau A, Raanani P & Shepshelovich D (2020). Safety-related postmarketing modifications of drugs for hematological malignancies. *Acta Haematologica*, 143(1), 73-77. <https://dx.doi.org/10.1159/000500229>

Green DJ, Sun H, Burnham J, Liu XI, van den Anker J, Temeck J, Burckart GJ (2019). Surrogate endpoints in pediatric studies submitted to the us FDA. *Clinical Pharmacology & Therapeutics*, 105(3), 555-557. <https://dx.doi.org/10.1002/cpt.1117>

Jiao B (2023). Estimating the potential benefits of confirmatory trials for drugs with accelerated approval: a comprehensive value of information framework. *Pharmacoeconomics*, 41(12), 1617-1627. <https://dx.doi.org/10.1007/s40273-023-01303-5>

Hey SP, Feldman WB, Jung EH, D'Andrea E & Kesselheim AS (2019). Surrogate endpoints and drug regulation: what is needed to clarify the evidence. *Journal of Law, Medicine & Ethics*, 47(3), 381-387. <https://dx.doi.org/10.1177/1073110519876167>

Huang Y, Xiong W, Ma L & Wu H (2022). A cross-sectional study of the FDA approved indications and supporting pivotal trials of small-molecular kinase inhibitors in cancer therapies with the biomarker of cancer driver gene. *International Journal of Cancer*, 151(12), 2107-2114. <https://dx.doi.org/10.1002/ijc.34222>

Huber C, Friede T, Stingl J & Benda N (2022). Classification of companion diagnostics: a new framework for biomarker-driven patient selection. *Therapeutic Innovation & Regulatory Science*, 56(2), 244-254. <https://dx.doi.org/10.1007/s43441-021-00352-2>

Jiang H, Jin Y, Yan H, Xu Z, Yang B, He Q & Luo P (2021). Hepatotoxicity of FDA-approved small molecule kinase inhibitors. *Expert Opinion on Drug Safety*, 20(3), 335-348. <https://dx.doi.org/10.1080/14740338.2021.1867104>

Surrogate endpoints in health technology assessment – Supplementary material



Kraus VB, Simon LS, Katz JN, Neogi T, Hunter D, Guermazi A & Karsdal MA (2019). Proposed study designs for approval based on a surrogate endpoint and a post-marketing confirmatory study under FDA's accelerated approval regulations for disease modifying osteoarthritis drugs. *Osteoarthritis & Cartilage*, 27(4), 571-579. <https://dx.doi.org/10.1016/j.joca.2018.11.002>

Kuller LH & Lopez OL (2021). Engage and emerge: truth and consequences? *Alzheimer's & Dementia*, 17(4), 692-695. <https://dx.doi.org/10.1002/alz.12286>

Kunnumakkara AB, Bordoloi D, Sailo BL, Roy NK, Thakur KK, Banik K, Aggarwal BB (2019). Cancer drug development: the missing links. *Experimental Biology & Medicine*, 244(8), 663-689. <https://dx.doi.org/10.1177/1535370219839163>

Levey AS, Gansevoort RT, Coresh J, Inker LA, Heerspink HL, Grams ME, Willis K (2020). Change in albuminuria and gfr as end points for clinical trials in early stages of ckd: a scientific workshop sponsored by the national kidney foundation in collaboration with the us food and drug administration and European medicines agency. *American Journal of Kidney Diseases*, 75(1), 84-104. <https://dx.doi.org/10.1053/j.ajkd.2019.06.009>

Li J, Wang H, Hua Y, Liu Y, Chen Y, Jiang R, Xie J (2023). Progress and challenges of the new conditional approval process in China: a pooled analysis from 2018 to 2021. *Clinical Therapeutics*, 45(11), 1111-1118. <https://dx.doi.org/10.1016/j.clinthera.2023.09.006>

Lienhardt C, Vernon AA, Cavaleri M, Nambiar S & Nahid P (2019). Development of new tb regimens: harmonizing trial design, product registration requirements, and public health guidance. *PLoS Medicine / Public Library of Science*, 16(9), e1002915. <https://dx.doi.org/10.1371/journal.pmed.1002915>

Lu CC, Beckman RA, Li XN, Zhang W, Jiang Q, Marchenko O, Yung G (2024). Tumor-agnostic approvals: insights and practical considerations. *Clinical Cancer Research*, 30(3), 480-488. <https://dx.doi.org/10.1158/1078-0432.CCR-23-1340>

Mayer-Hamblett N & VanDevanter DR (2020). Accelerated approval or risk reduction? how response biomarkers advance therapeutics through clinical trials in

Surrogate endpoints in health technology assessment – Supplementary material

cystic fibrosis. *Trends in Molecular Medicine*, 26(12), 1068-1077.

<https://dx.doi.org/10.1016/j.molmed.2020.08.002>

Muenzer J, Ho C, Lau H, Dant M, Fuller M, Boulos N, O'Neill C (2024). Community consensus for heparan sulfate as a biomarker to support accelerated approval in neuronopathic mucopolysaccharidoses. *Molecular Genetics & Metabolism*, 142(4), 108535. <https://dx.doi.org/10.1016/j.ymgme.2024.108535>

Naesens M, Loupy A, Hilbrands L, Oberbauer R, Bellini MI, Glotz D, Budde K (2022). Rationale for surrogate endpoints and conditional marketing authorization of new therapies for kidney transplantation. *Transplant International*, 35, 10137.

<https://dx.doi.org/10.3389/ti.2022.10137>

Novack GD (2020). Regulatory and developmental aspects of biomarkers in the treatment of ocular surface disease. *Eye & Contact Lens: Science & Clinical Practice*, 46 Suppl 2, S106-S108. <https://dx.doi.org/10.1097/ICL.0000000000000639>

Rathore AS, Stevenson JG & Chhabra H (2021). Considerations related to comparative clinical studies for biosimilars. *Expert Opinion on Drug Safety*, 20(3), 265-274. <https://dx.doi.org/10.1080/14740338.2021.1876024>

Ribeiro TB, Ribeiro A, Rodrigues LO, Harada G & Nobre MRC (2020). US food and drug administration anticancer drug approval trends from 2016 to 2018 for lung, colorectal, breast, and prostate cancer. *International Journal of Technology Assessment in Health Care*, 36(1), 20-28.

<https://dx.doi.org/10.1017/S0266462319000813>

Scott EC, Baines AC, Gong Y, Moore R Jr, Pamuk GE, Saber H, Beaver JA (2023). Trends in the approval of cancer therapies by the FDA in the twenty-first century. *Nature Reviews. Drug Discovery*, 22(8), 625-640.

<https://dx.doi.org/10.1038/s41573-023-00723-4>

Skydel JJ, Egilman AC, Wallach JD, Ramachandran R, Gupta R & Ross JS (2022). Spending by the centers for medicare & medicaid services before and after confirmation of benefit for drugs granted us food and drug administration accelerated

approval, 2012 to 2017. *JAMA Health Forum*, 3(5), e221158.

<https://dx.doi.org/10.1001/jamahealthforum.2022.1158>

Strauss DG, Wang YM, Florian J & Zineh I (2023). Pharmacodynamic biomarkers evidentiary considerations for biosimilar development and approval. *Clinical Pharmacology & Therapeutics*, 113(1), 55-61. <https://dx.doi.org/10.1002/cpt.2761>

Walia A, Haslam A & Prasad V (2022). FDA validation of surrogate endpoints in oncology: 2005-2022. *Journal of Cancer Policy*, 34, 100364.

<https://dx.doi.org/10.1016/j.jcpo.2022.100364>

Wallach JD, Yoon S, Doernberg H, Glick LR, Ciani O, Taylor RS, Ross JS (2024). Associations between surrogate markers and clinical outcomes for nononcologic chronic disease treatments. *JAMA*, 331(19), 1646-1654.

<https://dx.doi.org/10.1001/jama.2024.4175>

Yokoshiki S & Arato T (2024). Implementation status of pharmacological studies in the development of orphan drugs. *Orphanet Journal Of Rare Diseases*, 19(1), 5.

<https://dx.doi.org/10.1186/s13023-023-03000-0>

Zettler ME (2023). FDA approvals of oncology drugs for tissue-agnostic indications. *Targeted Oncology*, 18(5), 777-792. <https://dx.doi.org/10.1007/s11523-023-00982-6>

## **B. Review of HTA standards for surrogate endpoints**

### **B.1 Objectives**

A targeted review was undertaken to identify and map the standards used by health technology assessment bodies for the validation of surrogate endpoints.

### **B.2 Methods**

#### **B.2.1 Literature review**

To identify existing evidence standards for surrogate endpoints from an HTA perspective, a scoping search was carried out to establish whether there were any published literature reviews on the topic that can inform the project. A recently

published paper by Grigore et al. was identified in which the authors of the paper reviewed methodological guidance from European, Canadian and Australian HTA bodies and summarised how the guidance addressed methods for handling surrogate endpoints (Grigore et al., 2020). The methodology of the paper was considered robust and the scope extensive enough to serve as a starting point to identify HTA guidance on surrogate endpoints. We then proceeded to explore how we could expand on the findings of the paper. Grigore et al. identified 29 HTA bodies across different countries that included considerations for surrogate endpoints in their guidance documents. These included 18 European countries, the European Network for Health Technology Assessment (EUnetHTA) network of bodies, Australia and Canada (Grigore et al., 2020). Guidance related to 1 or more of the following aspects when using surrogate endpoints:

1. acceptability criteria
2. evidence strength assessment
3. validation methods
4. validation values.

We carried out a targeted review of HTA guidance documents from the subset of countries included in (Grigore et al., 2020) that outlined acceptability criteria for surrogate endpoints to check if they have any guidance documents that have been published or updated since the searches were carried out in the paper (March 2018) (Grigore et al., 2020). We also reviewed the guidance from the Institute for Clinical and Economic Review (ICER) in the US, which was not included in Grigore et al. 2020.

### **B.2.2 Data extraction**

Data was extracted from guidance documents in line with the following descriptive categories in the Grigore et al. (2020) review: acceptability criteria; evidence strength assessment; validation approaches. Data were also extracted about whether the guideline included any guidance on the inclusion of surrogate endpoints in economic modelling which was not captured in detail in Grigore et al. (2020) (Grigore et al., 2020). Documents that were not available in English were translated by colleagues of the working group within their bodies. Data were extracted independently by SC. All data were checked and cleaned by ZG.

### **B.2.3 Outline of analysis**

A descriptive analysis was undertaken to summarise standards for use of surrogate endpoints for HTA bodies.

### **B.3 Findings**

A total of 22 additional documents from 18 bodies that were not included in Grigore et al. (2020) were identified and reviewed. Of these, 17 were documents that had been updated or published since the review by Grigore et al. (2020) by bodies in countries covered in the paper; 2 were ICER guidance documents and 2 were from NIHO, a recently established Slovakian HTA body. Of the updated guidance documents additional information about surrogate endpoints had been added to 6 methods guides from the following bodies: the National Authority of Medicines and Health Products (INFARMED) in Portugal, High Health Authority (HAS) in France (Haute Autorité de Santé (HAS), 2020a, 2020b), EUnetHTA (EUnetHTA 21, 2023), the German Federal Joint Committee (G-BA), the Medical Services Advisory Committee (MSAC; (Australian Government Department of Health, 2021); Der Gemeinsamer Bundesausschuss-The Federal Joint Committee (G-BA), 2021) in Australia and the National Institute for Health and Care Excellence (NICE) (National Institute for Health and Care Excellence (NICE), 2023). Table A-1 provides an overview of changes and added information identified since Grigore et al. (2020). In cases where there were no updated guidance documents or details identified since Grigore et al. (2020), the most up-to-date existing guidance was extracted and included in the descriptive analysis. There were no specific details related to surrogate endpoints in the NIHO guidelines and therefore no information was extracted. The full data extraction table is presented in

Table B-2.

**Table B-1. Overview of updated guidance documents and their changes relating to surrogate endpoints (if any) since Grigore et al. (2020)**

Country	Agency (English name)	Agency (original name)	Acronym	References for documents reviewed	Changes relating to surrogate endpoints since Grigore et al. (2020) (Grigore et al., 2020)
Portugal	National Authority of Medicines and Health Products	Autoridade Nacional do Medicamento e Produtos de Saúde	INFARMED	(Perelman et al., 2019; Vinhas et al., 2021)	Guide updated since Grigore et al. (2020) data extraction – further detail on surrogates added aligned to the IQWiG guidance
France	High Health Authority	Haute Autorité de Santé	HAS	(Haute Autorité de Santé (HAS), 2020a, 2020b)	Guide updated since Grigore et al. (2020) data extraction – no additional detail on strength of evidence assessment, validation methods and validation values for surrogate endpoints.
Ireland	Health Information and Quality Authority	N/A	HIQA	(Health Information and Quality Authority (HIQA), 2019, 2020)	Guide updated since Grigore et al. (2020) data extraction – no additional detail on strength of evidence assessment, validation methods and validation values for surrogate endpoints.
Hungary	National Institute of Pharmacy and Nutrition	Országos Gyógyszerészeti és Élelmezés-egészségügyi Intézet	NIPN	(National Institute of Pharmacy and Nutrition, 2022)	Guide updated since Grigore et al. (2020) data extraction – no additional detail on strength of evidence assessment, validation methods and validation values for surrogate endpoints.
Wales	All Wales Therapeutics and Toxicology	N/A	AWTTC	(All Wales Medicines Strategy Group (AWMSG), 2023)	Guide updated since Grigore et al. (2020) data extraction – NICE methods apply in Wales and updates to the NICE methods will be implemented in Wales.
EU	European Network for Health Technology Assessment	N/A	EUnetHTA	(EUnetHTA, 2015; EUnetHTA 21, 2023)	Additional guidance published since Grigore et al (2020) (D4.4) – additional detail on strength of evidence assessment, reporting and transferability. Information from EUnetHTA 2015 surrogate endpoints document on statistical validation.
Slovakia	National Institute for Value and Technologies in Healthcare	Národný inštitút pre hodnotu a technológie v zdravotníctve	NIHO	(National Institute for Value and Technologies in Healthcare (NIHO), 2024b, 2024a)	New agency set up since Grigore et al. (2020). No specific details related to surrogates identified.
Italy	Italian Medicines Agency	Agenzia Italiana del Farmaco	AIFA	(Italian Medicines Agency (AIFA), 2016)	No additional published guidance since Grigore et al. (2020) identified.
Netherlands	National Health Care Institute	Zorginstituut Nederland	ZIN	(National Health Care Institute-Zorginstituut Nederland (ZIN), 2023, 2024)	Assessment guideline and economic guideline updated since Grigore et al. (2020) – no additional detail on strength of evidence assessment, validation methods and validation values for surrogate endpoints.

Country	Agency (English name)	Agency (original name)	Acronym	References for documents reviewed	Changes relating to surrogate endpoints since Grigore et al. (2020) (Grigore et al., 2020)
Germany	German Agency for HTA at the German Institute for Medical Documentation and Information	Deutsche Agentur für Health Technology Assessment	DIMDI	(Mangiapane & Garrido, 2009)	No additional published guidance since Grigore et al. (2020) identified. This is a review of the literature and current methodological guidelines (not the same status in decision making as some of the other manuals).
Germany	The German Federal Health Care Joint Committee	Gemeinsame Bundesausschuss	G-BA	(Der Gemeinsamer Bundesausschuss-The Federal Joint Committee (G-BA), 2021)	Guide updated since Grigore et al. (2020) data extraction - some additional information added about validation methods (consistent with IQWiG guidelines).
Germany	Institute for Quality and Efficiency in Health Care	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen	IQWiG	(Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2011, 2022)	Data extraction methods guide updated since Grigore et al. (2020).  Technical document on surrogate endpoints in oncology not updated
Spain	Health Technology Assessment Agency	Agencia de Evaluación de Tecnologías Sanitarias	AETS	(Imaz Iglesia et al., 1999)	No additional published guidance since Grigore et al. (2020) identified.
Spain	Andalusian Agency for Health Technology Assessment	Agencia de Evaluación de Tecnologías Sanitarias de Andalucía	AETSa	(Martín et al., 2013)	No additional published guidance since Grigore et al. (2020) identified.
US	Institute for Clinical and Economic Review	N/A	ICER	((ICER), 2022; Institute for Clinical and Economic Review (ICER), 2023)	ICER not included in Grigore et al. (2020).
Australia	The Pharmaceutical Benefits Advisory Committee	N/A	PBAC	(Pharmaceutical Benefits Advisory Committee (PBAC), 2016)	No additional published guidance since Grigore et al. (2020) identified.
Australia	Medical Services Advisory Committee	N/A	MSAC	(Australian Government Department of Health, 2021)	Guide updated since Grigore et al. (2020) data extraction – additional information on validation methods identified.
England	National Institute for Health and Care Excellence	N/A	NICE	(Bujkiewicz et al., 2019; Kaltenthaler et al., 2011; National Institute for Health and Care Excellence (NICE), 2023)	NICE health technology evaluation manual updated since Grigore et al. (2020) and additional technical documents also published (NICE TSD 13) — further information on evidence strength, validation methods and transferability provided.

### **B.3.1 Overall acceptability of surrogate endpoints**

All 17 of the HTA bodies included for data extraction mention overall acceptability. We found that there was a broadly consistent messaging around the acceptability of surrogate endpoints from HTA bodies in the manuals considered in our review. There was a preference for final outcomes to be used, and if a final outcome was not available then a surrogate endpoint can be used. If a surrogate was to be used, then it should be a validated surrogate. If there are no hard outcomes and no validated surrogates then some HTA bodies suggest submissions could include a non-validated surrogate, for example, the National Institute of Pharmacy and Nutrition (NIPN) in Hungary and the National Authority of Medicines and Health Products (INFARMED) in Portugal. Other bodies, such as the independent Institute for Quality and Efficiency in Health Care (IQWiG) in Germany, state that non-validated surrogates may be presented in reports but are not accepted as evidence of additional benefit of an intervention regardless of benefits observed. In all cases, the use of the surrogates (validated or non-validated) must be justified, and their validity evaluated.

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.

### **B.3.2 Strength of evidence**

Of the 17 HTA bodies, 8 provide detail on strength of evidence and this is usually made by referring to the following levels from (Ciani et al., 2017):

- 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 a preference for effect shown in multiple studies (for example, level 1 evidence). Some bodies like NICE and DIMDI refer to a holistic approach for assessing the strength of evidence supporting surrogacy assumptions, for example, stating that the statistical relationship needs to be accompanied with explanation of biological plausibility.

### **B.3.3 Validation process, methods and values**

Of the 17 HTA bodies, 7 mention validation of surrogate endpoints in their guidance documents. In terms of approach to take, bodies generally describe establishing validity by using a staged process that reflects the levels of evidence outlined in (Ciani et al., 2017). In terms of statistical approaches to validation, these are mentioned by 2 bodies (NICE and IQWiG). NICE Decision Support Unit guidance document 20 refers to the use of bivariate meta-analytic methods for randomised controlled trials (RCTs) while IQWiG guidance states that there is no universally applicable measure or estimation method or threshold that if exceeded would prove validity of a surrogate.

Two HTA bodies, IQWiG and INFARMED, along with the EUnetHTA 2015 guideline on surrogate endpoints, refer to validation values with INFARMED referring values in the IQWiG guidance. Correlation coefficient values between 0.85 to 0.95 are cited in reference to thresholds for validation of surrogates in the documents. However, it is highlighted that there is a lack of consensus on what the values should be. The EUnetHTA 2015 guideline highlights that even if there is no high correlation demonstrated, conclusions might be made if the surrogate threshold effect is taken into consideration in line with Burzykowski et al. (2005) (Burzykowski et al., 2005; EUnetHTA, 2015).

### **B.3.4 Transferability**

References to transferability of surrogate endpoints are made by 5 of the 17 bodies, with validation generally considered to be context specific. When considering

whether a surrogate endpoint that was validated in previous studies is valid or not, HTA bodies take the following into account:

- population
- interventions and mechanism of action
- setting in which data are collected
- disease and disease stage.

If validation studies are available only in different contexts it is recommended to consider the feasibility of developing a new validation model in a more comparable context and provide studies examining heterogeneity.

### **B.3.5 Reporting requirements**

We found that most guidance documents did not outline reporting requirements for cases where a surrogate endpoint is used, and the only guidance documents in which they were clearly described were from G-BA and EUnetHTA (D4.4) which highlighted that the following needs to be reported (Der Gemeinsamer Bundesausschuss-The Federal Joint Committee (G-BA), 2021; EUnetHTA 21, 2023):

1. The final outcome that the surrogate replaces
2. Rationale for use of the surrogate endpoint
3. Biological or medical rationale for link between final and surrogate endpoint
4. Validation of surrogate
  - a. Level of evidence for association
  - b. Strength of association
  - c. Certainty of association
5. Transferability: Alignment of studies used in validation with population, intervention (mechanism of action) and disease concerned in the submission

## 6. Additional uncertainties

### **B.3.6 Considerations in economic modelling**

Only 5 of the 22 guidance documents reviewed mention surrogate endpoints in relation to economic modelling, and 2 of the documents outlined modelling consideration in detail. The PBAC guidance outlines that using a proposed surrogate measure in an economic evaluation requires its transformation to a final clinical outcome and instances where this transformation can be used to calibrate economic models. The NICE manual on health technology evaluations mentions that the usefulness of surrogate endpoints for estimating quality-adjusted life years (QALYs) in cost-utility analyses is greatest when there is strong evidence that it predicts health-related quality of life or survival. It also highlights that the 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.

### **B.4 Summary**

According to our review, there has not been much change since Grigore et al. (2020) in terms of guidance on the use of surrogate endpoints in HTA (Grigore et al., 2020). The level of detail provided by different HTA bodies across the categories we considered varied considerably. There is a lack of information on the preferred statistical approaches for validation and how surrogates should be considered when conceptualising economic models for HTA decision making. This highlights the need for clear, consistent, detailed guidance for those developing economic models to inform HTA decision making.

## B.5 References

- (ICER), I. for C. and E. R. (2022). 2020-2023 Value Assessment Framework. [https://icer.org/wp-content/uploads/2020/11/ICER\\_2020\\_2023\\_VAF\\_02032022.pdf](https://icer.org/wp-content/uploads/2020/11/ICER_2020_2023_VAF_02032022.pdf)
- All Wales Medicines Strategy Group (AWMSG). (2023). Form B guidance notes. <https://awtc.nhs.wales/files/appraisal-process/form-b-guidance-notes-pdf-236kb/>
- Australian Government Department of Health. (2021). Guidelines for preparing assessments for the Medical Services Advisory Committee. Version 1.0. [http://www.msac.gov.au/internet/msac/publishing.nsf/Content/E0D4E4EDDE91EAC8CA2586E0007AFC75/\\$File/MSAC\\_Guidelines-complete-16-FINAL\\_\(18May21\).pdf](http://www.msac.gov.au/internet/msac/publishing.nsf/Content/E0D4E4EDDE91EAC8CA2586E0007AFC75/$File/MSAC_Guidelines-complete-16-FINAL_(18May21).pdf)
- 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.
- 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>
- Der Gemeinsamer Bundesausschuss-The Federal Joint Committee (G-BA). (2021). Anlage II.6: Modul 4 - Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen [Appendix II.6: Module 4 - Medical Benefits and Medical Benefits, Patient Groups with Therapeutically Significant Added. [https://www.g-ba.de/downloads/17-98-4825/2021-12-16\\_An12\\_6\\_Modul4.pdf](https://www.g-ba.de/downloads/17-98-4825/2021-12-16_An12_6_Modul4.pdf)

EUnetHTA. (2015). Endpoints used in Relative Effectiveness Assessment: Surrogate Endpoints.

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

Fleming, T. R., & DeMets, D. L. (1996). Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine*, 125(7), 605–613.

<https://doi.org/10.7326/0003-4819-125-7-199610010-00011>

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>

Haute Autorité de Santé (HAS). (2020a). Choices in Methods for Economic Evaluation. [https://www.has-sante.fr/upload/docs/application/pdf/2020-11/methodological\\_guidance\\_2020\\_-\\_choices\\_in\\_methods\\_for\\_economic\\_evaluation.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2020-11/methodological_guidance_2020_-_choices_in_methods_for_economic_evaluation.pdf)

Haute Autorité de Santé (HAS). (2020b). Transparency Committee doctrine Principles of medicinal product assessments and appraisal for reimbursement purposes.

Health Information and Quality Authority (HIQA). (2019). Guidelines for Evaluating the Clinical Effectiveness of Health Technologies in Ireland.

Health Information and Quality Authority (HIQA). (2020). Guidelines for the Economic Evaluation of Health Technologies in Ireland. <https://www.hiqa.ie/sites/default/files/2020-09/HTA-Economic-Guidelines-2020.pdf>

Imaz Iglesia, I., Gonzalez-Enriquez, J., Alcaide, J., & Olasagasti, J. (1999). Guía para la elaboración de informes de Evaluación de Tecnologías Sanitarias.

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].

Institute for Clinical and Economic Review (ICER). (2023). ICER's Reference Case for Economic Evaluations: Elements and Rationale. [https://icer.org/wp-content/uploads/2023/10/ICER\\_Reference-Case\\_For-Publication\\_Sept2023.pdf](https://icer.org/wp-content/uploads/2023/10/ICER_Reference-Case_For-Publication_Sept2023.pdf)

Italian Medicines Agency (AIFA). (2016). Criteri Per La Valutazione Dell'innovativita' [English translation: Criteria for the evaluation of innovativeness]. [https://www.aifa.gov.it/documents/20142/241044/Allegato\\_1\\_1.pdf](https://www.aifa.gov.it/documents/20142/241044/Allegato_1_1.pdf)

Kaltenthaler, E., Tappenden, P., Paisley, S., & Squires, H. (2011). NICE DSU Technical Support Document 13: Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models.

Mangiapane, S., & Garrido, M. V. (2009). Surrogatendpunkte als Parameter der Nutzenbewertung [English translation: Surrogate endpoints as parameters of the benefit assessment]. [https://portal.dimdi.de/de/hta/hta\\_berichte/hta250\\_bericht\\_de.pdf](https://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf)

Martín, L. A.-K., Calvo, C. B., Castillo Muñoz, M. A., Muñoz, C., Peláez, S. M., López, T. M., Navarro Caballero, J. A., Tabares, A. R., Pérez, R. U., & Vega Coca, D. (2013). Guía para la elaboración de informes de evaluación de medicamentos. <https://www.aetsa.org/como-lo-hacemos/>

National Health Care Institute-Zorginstituut Nederland (ZIN). (2023). Beoordeling stand van de wetenschap en praktijk 2023 [English translation: Assessment of the state of science and practice (2023 version)]. <https://www.zorginstituutnederland.nl/publicaties/publicatie/2023/04/11/beoordeling-swp-2023>

National Health Care Institute-Zorginstituut Nederland (ZIN). (2024). Guideline for economic evaluations in healthcare (2024 version).

<https://english.zorginstituutnederland.nl/about-us/publications/reports/2024/01/16/guideline-for-economic-evaluations-in-healthcare>

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>

National Institute for Value and Technologies in Healthcare (NIHO). (2024a).

Farmako-ekonomický rozbor lieku (FER) na účely kategorizácie liekov [English translation: Pharmacoeconomic analysis of a medicinal product (FER) for the purpose of categorisation of medicinal products].

[https://health.gov.sk/Zdroje?/Sources/kategorizacia/doku\\_kl/Farmako-ekonomicky-rozbor-lieku-FER.docx](https://health.gov.sk/Zdroje?/Sources/kategorizacia/doku_kl/Farmako-ekonomicky-rozbor-lieku-FER.docx)

National Institute for Value and Technologies in Healthcare (NIHO). (2024b).

Metodická príručka ku vyhláške Ministerstva zdravotníctva Slovenskej republiky č. 422/2011 Z. z. o podrobnostiach farmako-ekonomického rozboru lieku. [English translation: Methodological guide to the Decree of the Ministry of Health of the Slovak Republic. [https://health.gov.sk/Zdroje?/Sources/kategorizacia/doku\\_kl/MP-podrobnosti-farmako-ekonomickeho-rozboru-lieku.docx](https://health.gov.sk/Zdroje?/Sources/kategorizacia/doku_kl/MP-podrobnosti-farmako-ekonomickeho-rozboru-lieku.docx)

National Institute of Pharmacy and Nutrition, (NIPN). (2022). Országos

Gyógyszerészeti és Élelmezés-egészségügyi Intézet, Technológia-értékelő Főosztály. Módszertani Ajánlások. Verziószám: 1.0. ISBN-szám: 978-615-01-6954-5. [English translation of title: Technology Assessment Department. Methodological Recommendations. [https://ogyei.gov.hu/dynamic/ogyei\\_tef-modszertani-ajanlasok\\_20221221.pdf](https://ogyei.gov.hu/dynamic/ogyei_tef-modszertani-ajanlasok_20221221.pdf)

Perelman, J., Soares, M., Mateus, C., Duarte, A., Faria, R., Ferreira, L., Saramago, P., Veiga, P., Furtado, C., Caldeira, S., Teixeira, 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>

Vinhas, J., Dias, S., Melo, G., Viegas Dias, C., Sousa, D., Oliveira, J., Perelma, J., Azevedo, L., Marques, N., Saramago, P., Faria, R., Torres, S., Correia, A., & Couto, S. (2021). Pharmacotherapeutic assessment methodology, Version 3.0.

Weinstein, M. C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., & Luce, B. R. (2003). Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value in Health*, 6(1), 9–17.

<https://doi.org/https://doi.org/10.1046/j.1524-4733.2003.00234.x>

Weir, C. J., & Walley, R. J. (2006). Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Statistics in Medicine*, 25(2), 183–203.

<https://doi.org/https://doi.org/10.1002/sim.2319>



## B.6 Additional materials

### Data extracted for the review

**Table B-2 Summary of the data extracted for this review**

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
<p>England NICE</p> <p>(Bujkiewicz et al., 2019; Kaltenthaler et al., 2011; National Institute for Health and Care Excellence (NICE), 2023)</p>	<p>Health technology evaluation manual:</p> <p>For cost-utility analyses:</p> <ul style="list-style-type: none"> <li>Clinical endpoints that reflect how a patient feels, functions, or how long a patient lives are considered more informative than surrogate endpoints</li> <li>When using 'final' clinical endpoints is not possible and data on other outcomes are used to infer the effect of the technology on mortality and health-related quality of life, evidence supporting the outcome relationship must be provided together with an explanation of how the relationship is quantified for use in modelling.</li> </ul>	<p>TSD document 20: Bivariate meta-analytic methods</p> <p>Health technology evaluation manual: 4.6.6 Three levels of evidence for surrogate relationships can be considered in decision making (Ciani et al., 2017):</p> <p>Level 3: biological plausibility of relation between surrogate endpoint and final outcomes.</p> <p>Level 2: consistent association between surrogate endpoint and final outcomes. This would usually be derived from epidemiological or observational studies.</p> <p>Level 1: the technology's effect on the surrogate endpoint corresponds to commensurate effect on the final outcome as shown in RCTs.</p> <p>4.6.7 For a surrogate endpoint to be considered validated, there needs to be good evidence that the relative effect of a technology on the surrogate endpoint is predictive of its relative effect on the final outcome. This evidence preferably comes from a meta-analysis of level 1 evidence (that is, RCTs) that reported both the surrogate and the final outcomes, using the</p>	<p>Health technology evaluation manual: In cost-utility analyses, the usefulness of the surrogate endpoint for estimating QALYs will be greatest when there is strong evidence that it predicts health-related quality of life or survival.</p> <p>In all cases, the uncertainty associated with the relationship between the surrogate endpoints and the final outcomes should be quantified and presented. It should also be included through probabilistic sensitivity analysis and can be further explored in scenario analysis.</p> <p>TSD document 13 also lays out preference for data sources for health economic models</p> <p>For example, 1+: meta-analysis of RCTs with direct comparison and final outcomes, 1: single RCT with direct comparison and final outcome, 2+: meta-analysis of direct comparison measuring surrogate endpoint, 2: single RCT with direct comparison with surrogate endpoint, 3+: meta-analysis with placebo comparison with surrogate endpoint, 3: single trial with placebo control measuring surrogate, 4: case control or cohort, 5: non analytic 6 :expert opinion.</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
		<p>recommended meta-analytic methods outlined in technical support document 20 (bivariate meta-analytic methods).</p> <p>Show biological plausibility for all surrogate endpoints, but committees will reach decisions about the acceptability of the evidence according to the decision context. For example, for certain technologies indicated for rare conditions, and some diagnostic technologies and medical devices, the level of evidence might not be as high.</p> <p>4.6.8 The validation of a surrogate endpoint is specific to the population and technology type under consideration.</p> <p>4.6.9 Thoroughly justify extrapolating a surrogate to final relationship to a different population or technology of a different class or with a different mechanism of action.</p> <p>4.6.10 Extrapolation should be done using the recommended meta-analytic methods that allow borrowing of information from similar enough classes of technologies, populations, and settings, as outlined in technical support document 20. Existing relevant meta-analytical models may be used. However, when historical models are based on data collected in a different setting, then development of a new model using appropriate meta-analytic techniques is recommended. This may include network meta-analysis or hierarchical methods reflecting differences in mechanism of action between classes of technologies or for first-in-class scenarios.</p>	
Portugal INFARMED	Methodology for pharmacotherapeutic assessment of health technologies: When surrogate endpoint measures are used in a	Methodology for pharmacotherapeutic assessment of health technologies:	Methodological guidelines for economic evaluation studies of health technologies:

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
(Perelman et al., 2019; Vinhas et al., 2021)	<p>submission, it should also contain information on which clinical outcome measure the surrogate measure replaces and include demonstration of the validation of the surrogate measures used, using the methodology recommended herein.</p> <p>Thus, in assessing the additional benefit of an intervention, surrogate measures of therapeutic efficacy may be considered as substitutes for measures of clinical efficacy provided they have been previously validated. Non-validated surrogate endpoint measures may be accepted where there is a reasonable likelihood that the marker is capable of predicting clinical benefit, provided that the practical impossibility of validating the surrogate endpoint measure is demonstrated, for example, because the time required to observe the event (clinical outcome measure) is excessively long. For the purposes of this 'reasonability' there must be at least biological plausibility (level three validation), and a correlation must be observed between the surrogate and the clinical outcome measure (level two validation).</p>	<p>Validation goes through 3 stages and the stages from Ciana (2017) are outlined.</p> <p>There is no consensus on the correlation values (thresholds) required for validation of a surrogate, but often correlation coefficient values (Rstudy or Rindividual) between 0.85 and 0.955 are given. If there is not a high correlation, the surrogate threshold effect (STE) can still be used.</p> <p>Additional details on interpretation of the values are provided.</p> <p>In the case of new health technologies, which commonly use surrogate endpoint measures, evidence should be sought from other studies evaluating the same or similar health technologies (including drugs from the same class or, if this evidence is not available, including drugs from different classes).</p>	<p>Where the effectiveness evidence focuses on intermediate (or surrogate) endpoints, it is necessary to justify its association to final outcomes as implemented in the model, in accordance with Health Technology Assessment Commission (Comissão de Avaliação de Tecnologias de Saúde) recommendations.</p>
France HAS (Haute Autorité de Santé (HAS), 2020a, 2020b)	<p>Transparency Committee doctrine Principles of medicinal product assessments and appraisal for reimbursement purposes:</p> <p>The Transparency Committee (TC) considers that the primary outcome measure of a study must be a relevant clinical endpoint wherever it is possible to collect one. If a relevant clinical endpoint is not used in the trials, justification by the company explaining this choice is expected.</p> <p>The use of a surrogate endpoint – in particular a biomarker – is considered to be a relevant clinical endpoint on condition that a link with a clinical morbidity and mortality endpoint has been demonstrated in the disease concerned, in accordance with the definition of a surrogate endpoint.</p>	<p>Association between the surrogate endpoint and the patient-centred outcome. If a HTD submits a surrogate endpoint to replace an outcome requested by a pr if no patient-centred outcome is requested or available, the HTD should demonstrate the strength of the association between the surrogate endpoint and the patient-centred outcome and the treatment effect. This is often done via regression analysis for single studies, or meta-regression in the case of multiple studies. Ideally the association will be demonstrated at both the individual level and the trial level. The HTD can also provide scientific literature which demonstrates the link.</p>	<p>Choices in methods for economic evaluation:</p> <p>If the reference case analysis is a cost-effectiveness analysis, the health outcome criterion to be used should be that of life years. The mortality indicator should be the all-cause mortality rate.</p> <p>If the data required for the measurement of life years are unavailable, the use of a predictive criterion of expected survival time may be acceptable, but only if there is strong, established evidence of the predictive nature of this surrogate endpoint.</p> <p>A cost-effectiveness analysis may be based on other health outcome criteria in a supplemental analysis, with arguments supporting the choice of the criteria.</p> <p>If the data required to measure life years are not available, a survival prediction criterion may be used, but only if there</p>

Met opmerkingen [AK1]: It looks like something is missing here

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>The use of an intermediate outcome measure (without demonstration of a link with a relevant clinical endpoint) may be taken into account in assessment of the clinical added value (CAV).</p> <p>For example, in the field of oncology, the TC may take into account progression-free survival in situations whereby overall survival cannot be documented in the short or medium term (for example, long life expectancy, multiple subsequent therapeutic conditions) or where a link has been demonstrated between these 2 endpoints.</p> <p>For all outcomes requested in the assessment scope, the health technology developer (HTD) should provide, in addition to the previously reported follow up, the latest available data cut, regardless of how immature it is. The presence of surrogate endpoint data, regardless of their validity, does not change this requirement. For example, if an intervention is expected to impact overall survival, the latest data cut on overall survival should always be presented, even if the length of follow-up or the number of events is insufficient.</p> <p>Uncertainty: A surrogate endpoint may lead to greater uncertainty surrounding the benefit of the technology under assessment.</p>		<p>is strong, established evidence of the predictive character of this surrogate endpoint.</p> <p>The correlation factor should be presented and duly justified. The uncertainty generated by the predictive relationship should be explored through a sensitivity analysis.</p> <p>In the absence of this scientific demonstration, the reference case analysis should use the assumption of the lack of difference in survival between the intervention evaluated and its comparators. Where relevant, the results of an evaluation based on an assumption of higher survival should be presented in a scenario analysis.</p>
<p>Ireland HIQA</p> <p>(Health Information and Quality Authority (HIQA), 2019, 2020)</p>	<p>Guidelines for Evaluating the Clinical Effectiveness of Health Technologies in Ireland:</p> <p>A surrogate endpoint, also called an intermediate endpoint, is an objectively measured endpoint that is expected to predict clinical benefit or harm based on epidemiological, pathophysiological, therapeutic and other scientific evidence. They are typically physiological or biochemical markers that can be relatively quickly and easily measured. The effect of the technology on the surrogate endpoint must predict the effect on the clinical endpoint. (Fleming &amp; DeMets, 1996). The effect on the surrogate should be of a similar</p>	<p>No guidance details identified for these categories</p>	<p>No specific guidance beyond acceptability of surrogates and mention of instances where extrapolation may be needed.</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>magnitude to the effect on a final endpoint. If surrogate endpoints are assessed, caution must be exercised in directly extrapolating from these to final endpoints unless underpinned by a clear biological or medical rationale or they have a strong or validated link. Although a surrogate endpoint may have a strong link to an endpoint of interest, it may not itself represent a meaningful endpoint to the patient.</p> <p>Guidelines for the Economic Evaluation of Health Technologies in Ireland 2020:</p> <p>It may be necessary to extrapolate short-term outcome data or surrogate measures to final outcomes using modelling techniques. There are a variety of options to do this including superimposing the efficacy estimates from clinical trials on baseline probability estimates of survival from population-based sources (Weinstein et al., 2003).</p>		
<p>Hungary NIPN (National Institute of Pharmacy and Nutrition, 2022)</p>	<p>Translation of Országos Gyógyszerészeti és Élelmezés-egészségügyi Intézet</p> <p>Technológia-értékelő Főosztály: The use of validated surrogate endpoints – for example, an endpoint where the correlation between the degree of efficacy achieved and the degree of efficacy achieved by hard clinical endpoints is supported by evidence – is acceptable only if hard endpoints are not available. The use of non-validated surrogate endpoints may become necessary if neither hard clinical endpoints nor validated surrogate endpoints are available. A rationale must be provided for the use of non-validated surrogate endpoints.</p> <p>The data on the health benefit must be primarily based on endpoints for the long-term, clinically relevant effect (mortality, morbidity, quality of life assessed by patients in the various stages of a disease) of the health technology, instead of the exclusive use of surrogate endpoints.</p>	<p>No details provided for these categories</p>	<p>No details provided for this category</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
Wales AWMSG (All Wales Medicines Strategy Group (AWMSG), 2023)	Guidance notes for Clinical effectiveness: The applicant company is required to explain the approach used where there may be, for example, a disputed surrogate endpoint. If data are not specific to Wales and, for example, a different comparator has been used, the approach and rationale must be clearly explained. The response to this question should be based on the outcomes studied in clinical trials. It should discuss whether trials have directly measured health outcomes such as mortality, survival, incidence of disease, morbidity, functional performance, quality of life, or whether surrogate markers have been measured such as reduction in blood pressure, increase in FEV1, peak flow, and so on.	No guidance details identified for these categories	No guidance details identified for this category
EU EUnetHTA  (EUnetHTA, 2015; EUnetHTA 21, 2023)	Individual Practical Guideline Document. D4.4- Outcomes (Endpoints):  For diseases with expected long-term survival, it might be impossible to obtain mature mortality data from clinical trials at the time at which the joint clinical assessment (JCA) report is generated. If it is not feasible to measure a final outcome, then intermediate or surrogate endpoints may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome (EUnetHTA, 2015).  Points of attention for the assessment scoping process: a validated surrogate endpoint should only be used to replace a final patient-centred outcome of interest if absolutely necessary: <ul style="list-style-type: none"> <li>• If evidence for a patient-centred outcome such as morbidity, overall mortality and HRQoL is likely to be available, then this should be requested during the scoping process.</li> </ul>	Level of evidence: As detailed in 'Endpoints used in relative effectiveness assessment: surrogate endpoints' (EUnetHTA, 2015), appraisal of the association between the surrogate and the final outcome should take into account the level of evidence: 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; and establishment of correlation between effects on the surrogate endpoint and the patient-centred outcome; 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).	No guidance details identified for this category

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<ul style="list-style-type: none"> <li>Surrogate outcomes can be requested in addition to patient-centred outcomes where relevant. However, only surrogate outcomes for which validity has previously been clearly established should be requested where possible. This may not be possible at the scoping stage in many instances, although in some cases this might have been established by previous JCAs or in other literature on the same indication (EUneHTA, 2015).</li> </ul> <p>Requirements for JCA reporting</p> <p>The assessor should consider the following for the JCA report:</p> <ul style="list-style-type: none"> <li>The level of evidence for the association between the surrogate endpoint and the final patient-centred outcome.</li> <li>Details on whether this association is based on biological plausibility and/or empirical evidence.</li> <li>A description of whether this association has been studied in the disease stage, population and intervention of interest.</li> <li>In cases for which the association between the surrogate endpoint and the final patient-centred outcome has previously been examined but for a different disease stage, population or intervention, the assessment report should consider the implications for the validity of this association in the current population and intervention of interest.</li> <li>The strength of the association between the surrogate endpoint and the patient-centred outcome.</li> <li>The strength of the association between the treatment effect on the surrogate endpoint and the patient-centred outcome.</li> </ul>	<p>Endpoints used in relative effectiveness assessment:</p> <p>Surrogate endpoints:</p> <p>Overviews of statistical methods for the validation of surrogates have been given. The majority of the procedures, even those that have been applied to real data examples, rely on meta-analyses of several RCTs and estimate the correlation of the effects on the surrogate and the effects on the clinical endpoint. There is no clear consensus of which correlation values are sufficient to assume adequate surrogacy, but values of between about 0.85 and 0.95 are often discussed. If there is no high correlation demonstrated, conclusions might still be made if the STE is considered (Burzykowski et al., 2005). Also based on an analysis of several RCTs, the STE defines the minimum absolute value of the effect on the surrogate which has to be observed in a new trial to deduce an effect on the clinical endpoint. Accordingly, the STE can be computed for a certain level of change in a biomarker that will translate into clinical benefit. In both cases, certainty of the conclusions depends on prespecified levels of significance.</p>	

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<ul style="list-style-type: none"> <li>Any uncertainties associated with the evidence and quantified if available.</li> <li>The limitations of the use of a surrogate endpoint should be explicitly explained.</li> <li>An indication of whether a patient-centred outcome is likely to be available at a later date.</li> <li>Clearly outline any remaining areas of uncertainty.</li> </ul>		
Italy AIFA (Italian Medicines Agency (AIFA), 2016)	<p>Translation of Criteri Per La Valutazione Dell'innovativa:</p> <p>For oncology drugs, the gold standard outcome is overall survival (OS). Lack of OS evidence will need to be justified. Progression free survival (PFS), disease-free survival (DFS), duration of complete response or other surrogate endpoints whose value in the prediction of clinical benefit has been established in the context of the mechanism of action of the active substance, the type of cancer and the clinical setting, may be considered.</p>	No guidance details identified for these categories	No guidance details identified for this category
Netherlands ZIN (National Health Care Institute- Zorginstituut Nederland (ZIN), 2023, 2024)	<p>Translation of Beoordeling stand van de wetenschap en praktijk 2023:</p> <p>If no studies are available that measure the effect of the intervention to be assessed on the desired outcome, the effect on a validated surrogate endpoint can be looked at. This is an outcome that is usually easier or can be measured in a shorter time, and whose outcome has been shown to be correlated with the crucial outcome. Validated means that this connection or correlation has been sufficiently demonstrated in the relevant target group through research. For instance, for several types of cancer, the outcome on PFS has been shown to predict the effect on the outcome 'survival' well. And for other types of cancer it</p>	No guidance details identified for these categories	No guidance details identified for this category



Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>does not. Depending on the certainty of this relationship, there will be more or less confidence that the effect found on a surrogate endpoint says something about the crucial outcome. This is worked out using GRADE. Next, we determine for each outcome when the outcome is large enough, or in other words when there is a substantial improvement or worsening for the patient. To this end, as mentioned, it must first be determined how a specific outcome can be validly measured, or in other words with which (surrogate) outcome measures. For each outcome measure, the minimum difference in effect that must be found to be clinically relevant ('minimal important difference', MID) is then determined.</p>		
<p>Germany DIMDI (Mangiapane &amp; Garrido, 2009)</p>	<p>Translation of Surrogatendpunkte als Parameter der Nutzenbewertung:</p> <p>The criteria that a surrogate parameter need to fulfil in order to be recognised an acceptable and valid endpoint can be summarised as follows:</p> <ul style="list-style-type: none"> <li>• Biological plausibility: There is evidence from animal models and epidemiological studies of a causal relationship between the surrogate parameter and the clinical relevant endpoint. The surrogate is part of the pathophysiological causal path leading to the health outcome.</li> <li>• Magnitude of the association between surrogate and relevant endpoint: Epidemiological evidence has shown repeatedly and consistently that changes in the surrogate are qualitative and quantitative associated with changes in the relevant health outcome.</li> <li>• Evidence of effect from RCTs: There is evidence from RCTs showing that the changes induced by an intervention in the surrogate lead to changes in the relevant</li> </ul>	<p>Translation of Surrogatendpunkte als Parameter der Nutzenbewertung:</p> <p>In the full report, we summarise the different statistical methods discussed in the literature for the validation of surrogate endpoints. In summary, we conclude that there is no gold standard for the validation of surrogate endpoints. Since the generalisation of results from single studies is more prone to produce fallacies, approaches summarising results from several studies (such as meta-analysis) are preferred.</p> <p>(p10)</p> <p>In order to be considered valid and acceptable, a surrogate needs to fulfil several criteria. Thus, favourable results from statistical validation approaches are not a sufficient condition to conclude on the validity of a surrogate endpoint. Information on biological and pathophysiological factors is also required. In addition, the validity of a surrogate is to be seen as technology-specific. Whether a surrogate is able to capture the full effect of a technology</p>	<p>No guidance details identified for this category</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>outcome in the same direction. The effect of the intervention is fully captured by the surrogate. Even in the case of very similar active principles, the mechanism of action may differ. Thus, the transferability of conclusions on the validity of a surrogate from one technology to another needs to be carefully assessed.</p> <p>(p 9 from 98) A total of 13 from 23 analysed INAHTA member methodological papers' and 7 of 11 from 'fourth-hurdle bodies' provide information on how to choose outcome parameters for the assessment. All institutions agree that patient-relevant outcome parameters are strongly preferred in the assessment of the benefit of a health technology. All bodies underline that hard outcome parameters are to be preferred to surrogate endpoints. Nevertheless, the majority of bodies describe that under some circumstances surrogate endpoints may exceptionally be accepted – provided the validity of the surrogate is well established. In order to accept a surrogate, HTA bodies require the presentation of evidence which supports the causal relationship between surrogate and clinical relevant endpoint. None of the methodological guidance papers from HTA bodies provided a list of well-established and generally accepted surrogate endpoints. (p10)</p>	<p>depends on the mechanism of action of the technology in question. This is irrespective of whether a strong and consistent association between surrogate and relevant health outcome has been well established. (p11)</p>	
<p>Germany G-BA</p> <p>(Der Gemeinsamer Bundesausschuss-The Federal Joint Committee (G-BA), 2021)</p>	<p>Translation of Anlage II.6: Modul 4 - Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen: The use of surrogate endpoints requires justification. In addition, it should be explained whether and why the surrogate endpoints used are valid surrogate endpoints in the context under</p>	<p>Translation of Anlage II.6: Modul 4 - Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen</p> <p>To validate surrogate endpoints, generally a meta-analysis of studies studying not just the effects of the surrogate endpoint but also the effects of the endpoint most relevant to patients. These studies should be performed</p>	<p>No guidance details identified for this category</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>consideration or allow statements to be made about patient-relevant endpoints.</p> <p>Report in addition to the studies used for validation or to your justification to use surrogate endpoints the following, as a minimum:</p> <ul style="list-style-type: none"> <li>• patient population</li> <li>• intervention</li> <li>• control</li> <li>• data origin</li> <li>• used methodology</li> <li>• corresponding results (related to methodology)</li> <li>• study to the robustness</li> <li>• potentially any studies looking at transferability (p 5).</li> </ul>	<p>patient collectives and interventions, where not only the disease area but also the given medicine, as well as the comparator.</p> <p>The concept of STE is a possibility when no conclusive validation is available. There is also the possibility to consider quantitative correlation measures from surrogate endpoints and patient-relevant endpoints ('individual level') as well as effects on the surrogate endpoint on the interesting patient-level endpoint ('study level'). It should be noted that the lower bounds of the 95% confidence interval for such correlation measures are high enough. Other validation methods (see (Weir &amp; Walley, 2006)) should be justified sufficiently, especially when the basis of the data is a single study."</p>	
<p>Germany IQWiG</p> <p>(Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2011, 2022)</p>	<p>Translation of Allgemeine Methoden Entwurf für Version 7.0:</p> <p>Surrogate endpoints may only be considered when they have been previously based on suitable statistical methods within a sufficiently comparable patient population and have been validated with comparable interventions (for example, medicines with comparable mechanisms of actions). The effect on the patient-relevant endpoint can be replaced by the effect on the surrogate endpoint but this needs to be explained sufficiently to be considered valid. (p 44). Surrogate endpoints that are not valid or for which an adequate validation procedure has not been performed, may still be presented in the institute's reports. However, such endpoints are not suitable as evidence of the (additional) benefit of an intervention, regardless of the observable effects for a proof of the (additional) benefit of an intervention.</p>	<p>Translation of Allgemeine Methoden Entwurf für Version 7.0:</p> <p>In general, there is no standard process for validation, nor a best method for estimating. However, in the literature there often are correlation-based methods for surrogate validation with estimated correlation measures at the study and at the individual level. The institute's benefit assessment prefers such a validation method. Such a procedure normally requires meta-analysis of several randomised studies, where not only the effects in the surrogate endpoint but also the effects on the patient-relevant endpoints of interest are studied. Alternative methods are considered only in exceptional cases. (p45)"</p> <p>Validation of a correlation-based process is done on one hand on the individual level through a high correlation between surrogate</p>	<p>No guidance details identified for this category</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>Translation of Aussagekraft von Surrogatendpunkten in der Onkologie:</p> <p>When considering surrogate endpoints in the benefit assessment, 2 levels must be separated. At the first level, the validity of the surrogate must first be assessed. On the second level, the question arises as to what conclusions can be drawn regarding patient-relevant outcomes from the surrogate effects depending on the degree of validity of the surrogate.</p> <p>If results for a surrogate endpoint are presented for a benefit assessment, it must first be checked whether data are available to validate the surrogate. If this is not the case, there is no evidence (or indication or clue) of an effect regarding a patient-relevant endpoint from the results for the surrogate endpoint. If data is available to validate the surrogate, the first step is to assess its reliability. If the reliability of the validation results is low, there is basically no evidence of an effect on a patient-relevant endpoint from the results for the surrogate endpoint. If the validation results have high, limited or moderate reliability, the conclusion on patient-relevant endpoints depends on the degree of correlation between the effects of the surrogate and the patient-relevant endpoint in the validation studies and, if applicable, on the size of the effect for the surrogate in the benefit assessment studies. For this purpose, the effect on the surrogate resulting from the benefit assessment studies is compared with a surrogate threshold effect. The basic principle is that the more certain an effect on the surrogate also reflects an effect on the patient-relevant endpoint, the higher the certainty of the statement in the benefit assessment (for</p>	<p>and patient-relevant endpoint, and on the other hand on the study level through a high correlation of the effects between surrogate and the patient-relevant endpoints. The assessment of the benefit of the intervention is based on patient groups and therefore the judgement of validity is based on the degree of correlation at the level of treatment effects. The reliability of the results is taken into account for validation purposes and various criteria are used. As an example, connections between surrogate endpoint and the relevant patient-relevant endpoints for a given intervention might not be relevant in the same disease area but for another intervention if the other intervention has a different mechanism of action. Validation studies must be carried out in patient populations and interventions so that can make statements about the usage of surrogate endpoints in the disease area and the intervention to be assessed, as well as the comparator. If validation studies include different disease entities or interventions, to assess transferability, at least suitable studies on heterogeneity should be available.</p> <p>If a surrogate endpoint could not be conclusively validated, the concept of so-called STE can be applied. The surrogate endpoint resulting from the benefit assessment study is compared to the STE. (p45) The institute can conclude on the suitability of a surrogate endpoint based on the validation evidence or the consideration of the STE. The decisive factors for the former are the degree of correlation of the effects on the surrogate and the patient-relevant endpoint and the reliability of the validation in the validation studies. When considering a STE, the size of the effect on the surrogate in the benefit assessment studies compared to the STE is the decisive criterion. Depending</p>	

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>example, evidence of or indication of an effect on a patient-relevant endpoint). (p xv).</p>	<p>on the constellation, if there is a statistically significant effect on the surrogate endpoints, all gradations of the statements on (additional) benefit with regard to the associated patient-relevant endpoint are met.</p> <p>Translation of Aussagekraft von Surrogatendpunkten in der Onkologie:</p> <ol style="list-style-type: none"> <li>1. Proof of a correlation between surrogate and clinical endpoint alone is not sufficient for surrogate validation.</li> <li>2. The transferability of statements about the validity of surrogates between different diseases or manifestations of a disease or between different interventions is not easily possible. The validity of a surrogate can be both disease and intervention specific, for example, a final surrogate validation using statistical methods may only be possible within 1 indication and 1 intervention. The extent to which statements about the validity of surrogates can be transferred between different diseases or interventions must be examined and justified. To assess transferability, validation studies that include different disease entities or interventions should at least examine heterogeneity.</li> <li>3. To validate a surrogate, extensive data is required, preferably a meta-analysis of several randomised studies with sufficient certainty of results. In these studies, both the surrogate and the clinical endpoint must have been recorded.</li> <li>4. For surrogate validation, there is neither a universally applicable measure nor a generally best estimation method nor a generally accepted limit, where when exceeding this limit would mean proof of validity.</li> </ol>	

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
		<p>5. Currently, correlation-based methods are primarily used in practice to validate surrogate endpoints and applied with estimates of study- and individual-level correlation measures. In the literature, on a study-level, a 'high' correlation is required between the surrogate endpoint and the patient-relevant effects; various authors suggest a correlation of 0.9 as a possible limit value.</p> <p>Particularly if, taking into account the associated confidence intervals, there is no high correlation (lower confidence limit of the estimated correlation coefficients below 0.85), and the validity of the surrogate remains unclear, the STE concept according to (Burzykowski et al., 2005; Burzykowski &amp; Buyse, 2006) can be used. If the effects for the surrogate are sufficiently large, statements can still be made regarding patient-relevant endpoints. The STE concept is used to determine threshold values for the decision-making process as to whether an observed effect on the surrogate is also associated with an effect on the endpoint of interest with sufficient certainty. To do this, the lower confidence limit of the treatment effect with regard to the surrogate must be greater than the STE. (p ix and x)</p>	
Spain AETS (Imaz Iglesia et al., 1999)	<p>Translation of Guía para la elaboración de informes de Evaluación de Tecnologías Sanitarias:</p> <p>If measures are used that describe the clinical outcomes of indirectly or intermediately, such as changes produced in a biochemical parameter, doubts may arise about the real effectiveness of the technology evaluated. It is important to draw attention to the frequent use of indirect or surrogate endpoints, which are,</p>	No guidance details identified for these categories	No guidance details identified for this category

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>generally, physiological or biochemical parameters that are part of a causal chain or association with clinical results or effects (see example in table 1). (pg 19) The example in table 1 is about reduction in arrhythmias and influence post-infarction mortality (CAST, Cardiac Arrhythmia Suppression Trial). Originally believed these correlated but the trial found the opposite. The assumption that intermediate measures predict the final result is risky, for 2 fundamental reasons. First, because biology is extremely sensitive to a multitude of factors that are difficult to understand and know in their entirety and the biological associations demonstrated in specific cases cannot be used in a general way to predict complex clinical behaviours. Furthermore, it has often been observed that the existence of association, between an intermediate result A, and final result B, under laboratory or ideal conditions, dilutes or disappears completely in the conditions of usual practice. (pg 19-20). Intermediate or surrogate endpoints are often used to draw conclusions about the consistency of a cause-effect relationship between an intervention and a clinical outcome. Although indirect tests often provide useful knowledge, it is necessary to warn and carefully assess the limitations of its use (pg 42)</p>		
<p>Spain AETSA (Martín et al., 2013)</p>	<p>Translation of Guía para la elaboración de informes de evaluación de medicamentos:</p> <p>Measurements of results (outcomes) The efficacy and safety results will be defined and will be considered in the systematic review. The results will be categorised as primary or secondary where possible. Clinically relevant variables should be included as an ideal outcome. In the event that there are no studies in which these variables are analysed they will consider surrogate and or intermediate</p>	<p>No guidance details identified for these categories</p>	<p>No guidance details identified for this category</p>

Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	<p>variables. Any outcome measure that is not widely known should be explained in an additional annex. (pg 33)</p> <p>Sometimes, there is no information on the effectiveness of interventions health services in terms of final variables, so it is not possible to calculate cost ratios by QALY or AVG and therefore no references to thresholds can be made. In these cases, approximations to cost effectiveness can be made using calculations related to variables intermediate or surrogate based on the results of the systematic review of effectiveness (pg 54)</p>		
<p>US ICER (ICER), 2022; Institute for Clinical and Economic Review (ICER), 2023)</p>	<p>ICER Value Assessment Framework:</p> <p>When evidence on patient-important outcomes is limited or unavailable, we will seek evidence on surrogate endpoints that might be associated with outcomes important to patients and families. Health outcomes, for example, changes in symptoms or conditions that people experience and that affect the quantity or quality of life (such as change in pain, quality of life, length of life) are given greater weight than intermediate outcomes (such as change in cholesterol).</p>		<p>ICER's Reference Case for Economic Evaluations: Elements and Rationale:</p> <p>When there are challenges in translating the outcome measures used in clinical trials or available patient-reported data into QALYs, analysts should conduct a search for 'mapping' studies that allow translation of surrogate endpoints into quality-of-life measures. If used, the report should discuss the validity of the mapping studies and their translation into QALYs, as well as the rationale for choosing specific mapping algorithms. If an analysis is using the URD framework, the model report should acknowledge and highlight additional uncertainty in translating patient outcomes into QALY measures, if relevant.</p>
<p>Australia PBAC (Pharmaceutical Benefits Advisory Committee (PBAC), 2016)</p>	<p>The Pharmaceutical Benefits Advisory Committee Guidelines. Appendix 5: Translating comparative treatment effects of proposed surrogate measures to target clinical outcomes.</p> <p>General guidance is that where possible, present evidence from direct randomised trials of the treatment effect of the proposed medicine on clinically relevant outcomes. Where no such evidence is available, establish the likely comparative treatment effect on clinically relevant outcomes by transforming the</p>	<p>The Pharmaceutical Benefits Advisory Committee Guidelines. Appendix 5: Translating comparative treatment effects of proposed surrogate measures to target clinical outcomes:</p> <p>Use the following types of evidence to analyse a PSM-TCO relationship (listed from strongest to weakest): 1. Multi trial meta-regression; 2. Single trial or small number of randomised trials where individual patient data are available (including multicentre</p>	<p>The Pharmaceutical Benefits Advisory Committee Guidelines. Appendix 5: Translating comparative treatment effects of proposed surrogate measures to target clinical outcomes:</p> <p>In general, once a comparative treatment effect on a TCO and some estimate of its uncertainty has been generated based on transforming a PSM, incorporation into the economic evaluation would proceed as usual with the transformed TCO taking the place of a directly measured TCO as an input variable for the economic evaluation. For example, the separate transformation of a TCO to a QALY</p>



Country Agency References	Acceptability of surrogate	Strength of evidence, validation approach or validation values	Guidance on managing surrogate endpoints in development of economic models
	comparative treatment effect of a surrogate measure.	<p>analysis where participants were randomised by centre); 3. One randomised trial – no individual patient data or not randomised by centre; 4. No randomised trial data.</p> <p>The following steps are recommended in establishing the likely comparative treatment effect on clinically relevant outcomes by transforming the comparative treatment effect of a surrogate measure. 1. Define the proposed surrogate measures (PSM) and the final clinical outcomes (TCOs).</p> <p>2. Establish the biological reasoning for the link between the PSM and the TCO, including how pivotal the PSM is to the causation pathway of the TCO, and present epidemiological evidence to support this.</p> <p>3. Present randomised trial evidence to support the nature of the PSM-TCO comparative treatment effect relationship.</p> <p>4. Translate the comparative treatment effect on the PSM from the studies to an estimate of the comparative treatment effect for the TCO.</p>	<p>via the use of QALY weights would need to be done irrespective of the basis for estimating the TCO. There are however, some particular implications for the sensitivity analysis and consequential resource use. Several modelling approaches are used in economic evaluations to estimate final outcomes and some rely on assumptions that are independent of a PSM to TCO transformation. In such instances, it might be possible to use the transformation to generate an estimate of the TCO which could be used to calibrate the model. Such a calibration exercise would not only be limited by the extent of uncertainty around the transformed point estimate, but also because, as discussed above, the TCO might need to be more tightly defined to a single point in time in order to be subjected to analysis by meta-regression.</p>

## **C. Review of statistical methods for validation of surrogate endpoints**

### **C.1 Objectives**

We performed a scoping review to identify, categorise, and describe statistical methods that have been proposed to validate surrogate endpoints.

### **C.2 Methods**

#### **C.2.1 Literature review**

We followed a pre-specified project plan informed by accepted methods for scoping reviews (Peters et al., 2020), with adaptations to complete the review in a short time frame.

#### **C.2.2 Selection criteria**

The selection criteria are in Table C-1. We included review articles and guidance documents published in English within the past 5 years (January 2019 to December 2023) that had a focus on (for example, as an objective) providing guidance on or descriptions of statistical approaches used to validate surrogate endpoints related to any population, intervention or condition. The decision to focus on review articles and guidance documents published in the past 5 years was made to expedite the scoping review process, because it would have been infeasible to review the many primary methods papers (such as proposing new methods) that are available.

**Table C-1 Selection criteria for review of statistical methods for validation of surrogate endpoints**

<b>Criterion</b>	<b>Include</b>	<b>Exclude</b>
<b>Population</b>	Any population (including non-human or simulated data), intervention, and condition	Not applicable
<b>Concept</b>	A focus of the paper is to provide guidance on or descriptions of statistical approaches used to validate surrogate endpoints	Papers providing information about statistical approaches but not as a main focus or objective; guidance on statistical approaches that have been applied for reasons other than validating a surrogate; papers presenting evidence for the validity of a surrogate
<b>Context</b>	Any	Not applicable
<b>Sources of evidence</b>	Review articles (we included any type of review article, as defined by the authors of the papers, because systematic reviews were unlikely to be available) Guidance documents	Primary methods papers, opinion pieces, commentaries, abstracts
<b>Date of publication</b>	January 2019 to December 18, 2023 (we considered that review papers more than 5 years old were likely to present information that would be redundant with more recent reviews)	Prior to 2019
<b>Language</b>	English	Any language other than English

### **C.2.3 Search strategy**

An information specialist conducted a literature search on key resources including MEDLINE and Scopus. The search approach was customised to retrieve a limited set of results, balancing comprehensiveness with relevancy and resource constraints. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine’s MeSH (Medical Subject Headings), and keywords. Search concepts were developed based on the elements of the research questions and selection criteria. The main search concept was surrogate endpoints. Limiting concepts included methodological, validation and statistical terminology. The search was completed on December 18, 2023. No publication date or language limits were applied. The literature search strategy is presented in the supplementary materials. Duplicates were removed manually in EndNote.

To supplement the database search, we scanned the reference lists of pertinent review articles to identify additional papers that met the selection criteria.

#### **C.2.4 Selection of documents**

We exported the results of the literature search to a Microsoft Excel database for screening. We selected documents for inclusion in a 2-stage approach, first by title and abstract, and then by full text. A pilot round preceded each stage, where 2 reviewers screened a sample (for example, 50 abstracts, 3 full texts) of records independently in duplicate to ensure a mutual understanding of the selection criteria. The 2 reviewers discussed disagreements and clarified elements of the selection criteria, after which a single reviewer screened the remaining records. When needed, the single reviewer clarified uncertainties with a second reviewer. A single reviewer screened records identified in reference lists.

#### **C.2.5 Data extraction, analysis and presentation**

We extracted relevant data directly into a table in Microsoft Word for presentation. To ensure relevance to the intended users, we iteratively adapted the table contents in the early phases of data extraction. Relevant data items were:

- name of the method
- published review sources
- early and related reference article(s); this is not intended to be an exhaustive list
- description of how the method can be used to evaluate surrogacy, and proposed guidance for assessment (for example, threshold value), if available
- important considerations, including key assumptions of the statistical approach.

A single reviewer with statistical expertise extracted data into the table, starting with the most recently published review article. Subsequent review articles were used to fill in missing information and to expand on or clarify concepts from other review articles. Where relevant, related methods (for example, extensions or adaptations) were grouped within the table when review articles characterized a method as an addition to another approach. Following extraction of data from all review articles, we consolidated all available information to streamline the presentation. Methods were grouped based on common themes: foundational methods (such as initial methods

developed for evaluating surrogates), methods for multiple trials, causal inference methods, and other relevant methods. Within each grouping, methods were sorted by first publication of the method.

### **C.3 Findings**

Of 1,286 unique records identified by the database searches, we screened 347 that were published after 2018. Of these, we screened 29 full texts and included 3 review articles (Elliott, 2023; Weir & Taylor, 2022; Zhuang & Chen, 2020). We were aware of an additional review article (Ensor et al., 2016) that did not meet our eligibility criteria (published pre-2019), but included it due to its high relevance. We did not identify additional relevant reviews in the reference list scan. Papers excluded following full text review are in section C.7.2. Statistical methods identified in the review articles are presented in Table C-2.

**Table C-2 Validation methods for surrogate endpoints**

Method name(s) and review source	Related reference article(s)	Description	Important considerations
<b>Foundational or first methods to discuss validation of surrogate endpoints</b>	–	–	–
Prentice criteria  (Elliott, 2023; Ensor et al., 2016; Weir & Taylor, 2022; Zhuang & Chen, 2020)	(Prentice, 1989)	<b>Surrogacy evaluation</b> 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.  <b>Key assumptions</b> <ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome.</li> </ul>
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)	<b>Surrogacy evaluation</b> 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.  <b>Proposed criteria for assessment</b> Freedman et al. (Freedman et al., 1992) suggest that the lower limit of the 95% confidence interval for PTE should exceed a critical value, such as 0.5 or 0.75, before declaring a surrogate as adequate.	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.  <b>Key assumptions</b>

Method name(s) and review source	Related reference article(s)	Description	Important considerations
			<ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome.</li> <li>Treatment effect is primarily mediated by the surrogate endpoint.</li> <li>No interaction between treatment and surrogate.</li> <li>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.</li> </ul>
Proportion of treatment effect explained (a generalized version)  (Elliott, 2023; Ensor et al., 2016; Zhuang & Chen, 2020)	(Deslandes & Chevret, 2007; X. Wang et al., 2020; Y. Wang & Taylor, 2002)	<b>Surrogacy evaluation</b> 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.	Generalized 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.  <b>Key assumptions</b> <ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome.</li> </ul>
<b>Methods for multiple trials</b>	–	–	–

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)	<b>Surrogacy evaluation</b> 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.  <b>Key assumptions</b> <ul style="list-style-type: none"> <li>• The relationship between the treatment, surrogate endpoint, and the clinical outcome are the same across multiple trials.</li> <li>• Participants across trials are exchangeable.</li> <li>• No unmeasured confounding between the surrogate and the clinical outcome.</li> </ul>
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)	<b>Surrogacy evaluation</b> 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. Other measures have also been proposed such as the PIG.	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.  Evaluation based on trial level $R^2$ has been criticised for lacking a clear interpretation outside of the setting in which



Method name(s) and review source	Related reference article(s)	Description	Important considerations
		<p><b>Proposed criteria for assessment</b></p> <p>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.</p>	<p>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.</p> <p><b>Key assumptions</b></p> <ul style="list-style-type: none"> <li>• The relationship between the treatment, surrogate endpoint, and the final outcome are the same across multiple trials.</li> <li>• Participants across trials are exchangeable.</li> <li>• 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.</li> <li>• No interaction between treatment and surrogate.</li> </ul>
<p>Information theoretic approach</p> <p>(Ensor et al., 2016; Weir &amp; Taylor, 2022)</p>	<p>(Alonso &amp; Molenberghs, 2007)</p>	<p><b>Surrogacy evaluation</b></p> <p>Quantifies the relationship through an R<sup>2</sup> 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).</p>	<p>Requires multiple trials on the same treatment with the same surrogate and final outcome.</p> <p>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.</p> <p><b>Key assumptions</b></p> <ul style="list-style-type: none"> <li>• The relationship between the treatment, surrogate endpoint, and the final outcome are the same across multiple trials.</li> <li>• Participants across trials are exchangeable.</li> </ul>

Method name(s) and review source	Related reference article(s)	Description	Important considerations
			<ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome.</li> </ul>
<b>Causal inference methods</b>	–	–	–
Natural indirect effects  (Elliott, 2023; Ensor et al., 2016; Zhuang & Chen, 2020)	(Robins & Greenland, 1992)	<p><b>Surrogacy evaluation</b></p> <p>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 generalized PTE in certain situations.</p>	<p>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.</p> <p><b>Key assumptions</b></p> <ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome.</li> <li>No interaction between treatment and surrogate endpoint (extensions to avoid this assumption have been developed).</li> </ul>

Method name(s) and review source	Related reference article(s)	Description	Important considerations
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)	<p><b>Surrogacy evaluation</b> Expresses the strength of a surrogate as a function of the values of the surrogate under different treatments referred to as the CEP surface.</p> <p><b>Proposed criteria for assessment</b> 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 &amp; Hudgens, 2008) as well as graphical approaches (Conlon et al., 2014). One metric, the PAE (Gilbert &amp; 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.</p>	<p>Initially formulated under the setting where both the final outcome and surrogate endpoint are binary. Extensions to other select settings have also been developed.</p> <p>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 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.</p> <p>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.</p> <p><b>Key assumptions</b></p> <ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome. Approaches have been developed to resolve this assumption but are currently not practically implementable.</li> </ul>
<b>Other relevant methods</b>	–	–	–
Single trial: relative effect and adjusted association	(Buyse & Molenberghs, 1998)	<p><b>Surrogacy evaluation</b> 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.</p>	<p>Formulated in the linear regression setting (for example, continuous measures of the surrogate and final outcome).</p> <p><b>Key assumptions</b></p>

Method name(s) and review source	Related reference article(s)	Description	Important considerations
(Ensor et al., 2016; Weir & Taylor, 2022; Zhuang & Chen, 2020)			<ul style="list-style-type: none"> <li>No unmeasured confounding between the surrogate and the final outcome.</li> <li>Regression model is accurate based on 1 trial only, which is an untestable assumption.</li> </ul>
Non-parametric approach  (Elliott, 2023)	(Parast et al., 2016; Price et al., 2018; X. Wang et al., 2020)	<p><b>Surrogacy evaluation</b></p> <p>Multiple approaches have been described, each of which generally focus on evaluation of a surrogate based on the optimization 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.</p>	Unlike other methods described in this table, this approach generally avoids model-based (parametric) assumptions. The approach can be extended to combine multiple surrogate 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

## C.4 Discussion

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 HTA bodies who need to make decisions when evidence for only a surrogate endpoint is available, or to inform future work on best practices for validation of surrogate endpoints.

Best practices about the statistical approach to validation of surrogate endpoints have not yet been established. HTA bodies and others have described an approach to validation that includes 3 levels of evidence for surrogate endpoints (Ciani et al., 2022; Elston & Taylor, 2009; EUnetHTA, 2015; National Institute for Health and Care Excellence (NICE), 2023; Thorlund et al., 2024). Level 1 evidence is obtained from 1 or more RCTs demonstrating a treatment effect on the final outcome and a proportionate change in the surrogate endpoint; level 2 evidence demonstrates a consistent association between the surrogate and final outcomes and may be obtained from interventional, epidemiological or observational studies; level 3 evidence demonstrates biological plausibility of an association between the surrogate and final outcome and may be obtained from pathophysiological studies and understanding of the disease process (Ciani et al., 2017). The statistical methods described in table C-2 were largely conceptualized for application to the assessment of level 1 evidence, using RCT data with drug intervention, surrogate endpoint and final outcome available in a data set. For this reason, the important considerations highlighted in Table C-2 focus on this setting. Application of any of these methods to real-world or other data sets would require careful consideration of additional assumptions relevant to non-randomized data, most notable of which are concerns of confounding between treatment and the final clinical outcome and between treatment and surrogate endpoint.

In our review, we found few criteria or thresholds for determining a 'good surrogate' based on statistical measures (for example, trial-level  $R^2$ ) used in evaluation. 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 clinical outcome and the uncertainty around the predicted treatment effect on the clinical outcome (Christensen et al., 2024; Ciani et al., 2017, 2022; Dawoud et al., 2021).

The context of the validation also needs to be considered. A surrogate endpoint that is considered acceptable in one context may not be readily applicable to another where the population, condition, line of therapy, intervention and or comparator drugs differ (EUnetHTA, 2015). These contextual considerations are of additional concern for 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 are often inaccessible. Aside from the 'methods for multiple trials' in Table C-2 (applied in an aggregate data setting), all methods were described in the context of individual patient data. It is possible that other methods in the table could be applied to aggregate data either directly or through the requirement of additional assumptions, but we did not explore this in detail. A NICE technical support document (Bujkiewicz et al., 2019) discusses bivariate meta-analysis for aggregate level data and provides an example for the purpose of surrogate endpoint evaluation.

## **C.5 Summary**

In summary, this work describes published statistical methods that have been proposed to validate surrogate endpoints, extracted from review articles published in the past 5 years. We followed typical scoping review methodology to systematically identify review papers and ensure that the most common statistical methods are described. New or novel methods, and methods that are less frequently used or discussed, may not have been reported. Table C-2 provides a high-level overview of reported statistical methods, limited to the information reported in the review papers. Readers seeking detailed information may consult the primary literature, including the key early and related reference articles that have been highlighted.

## C.6 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>
- Buyse, M., & Molenberghs, G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*, 54(3), 1014–1029.
- 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., & 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>

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](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.

Elston, J., & Taylor, R. S. (2009). Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports. *International Journal of Technology Assessment in Health Care*, 25(1), 6–13. <https://doi.org/10.1017/S0266462309090023>

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. (2015). Endpoints used in Relative Effectiveness Assessment: Surrogate 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.

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

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](https://doi.org/10.1002/(sici)1097-0258(19970715)16:13<1515::aid-sim572>3.0.co;2-1)

Manyara, A. M., Davies, P., Stewart, D., Weir, C. J., Young, A. E., Wells, V., Blazeby, J., Butcher, N. J., Bujkiewicz, S., Chan, A. W., Collins, G. S., Dawoud, D., Offringa, M., Ouwens, M., Ross, J. S., Taylor, R. S., & Ciani, O.

(2023). Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. *Journal of Clinical Epidemiology*, 160, 83–99.

<https://doi.org/https://dx.doi.org/10.1016/j.jclinepi.2023.06.013>

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>

Peters, M. D. J., Godfrey, C., Mcinerney, P., Munn, Z., C., T. A., & Khalil, H. (2020). Scoping reviews. In E. Aromataris, C. Lockwood, K. Porritt, B. Pilla, & Z. Jordan (Eds.), *JBI manual for evidence synthesis*.

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](https://doi.org/10.1111/j.1541-0420.2007.00852_1.x)

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>

Thorlund, K., Shephard, C., Machado, L., Bourgooin, 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>

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>

## C.7 Supplementary materials

### C.7.1 Search strategy

#### C.7.1.1 Overview

**Databases searched:** Ovid Medline All, Scopus

**Date of search:** December 18, 2023

**Alerts:** None

**Search filters applied:** None

**Limits**

- Publication date limit: None
- Language limit: None

## Ovid MEDLINE ALL 1946 to December 15, 2023

#	Searches	Results
1	(surrogate and (outcome* or endpoint* or end point* or marker* or biomarker*)).ti,kf.	3639
2	(surrogate adj5 (outcome* or endpoint* or end point* or marker* or biomarker*)).ab. /freq=2	3229
3	surrogacy.ti. and (end point* or endpoint* or marker* or biomarker*).ti,ab,kf.	62
4	(surrogacy and (surrogate adj5 (outcome* or endpoint* or end point* or marker* or biomarker*))).ab.	233
5	exp biomarkers/ and (surrogate or surrogacy).ti,kf.	2467
6	((intermediate or intermediary) adj3 (outcome* or endpoint* or end point* or marker* or biomarker*)).ti,kf.	817
7	((intermediate or intermediary) adj3 (outcome* or endpoint* or end point* or marker* or biomarker*)).ab. /freq=2	740
8	or/1-7	7578
9	methodolog*.ti,ab,kf.	517335
10	(framework* or policy or policies or guideline*).ti,kf.	285689
11	validation study.pt.	109276
12	(valid* or reproducib* or reliab*).ti,hw,kf.	654394
13	*validation studies as topic/ or **predictive value of tests"/ or ***reproducibility of results"/	5768
14	validity.ab. /freq=2	67221
15	or/9-14	1422358
16	statistical.ti.	44379
17	(statistician* or biostat*).ti,ab,kf.	7661
18	statistic*.jw.	72884
19	bayesian.ti,ab,kf.	69726
20	(biometric* or biometrik*).jw.	8194
21	(theory or theoretic*).ti,kf.	165877
22	(theory or theoretic*).ab. /freq=2	155742
23	or/16-22	466584
24	15 or 23	1839951
25	8 and 24	1135

### C.7.1.2 Scopus

( TITLE-ABS-KEY ( "surrogate endpoint\*" OR "surrogate end point\*" OR "surrogate marker\*" OR "surrogate outcome\*" OR "surrogate biomarker\*" ) AND SRCTITLE ( biometric\* OR biometrik\* OR statistic\* ) AND NOT INDEX ( medline ) ) OR ( TITLE ( "surrogate endpoint\*" OR "surrogate end point\*" OR "surrogate marker\*" OR "surrogate outcome\*" OR "surrogate biomarker\*" ) AND TITLE ( methodolog\* OR framework\* OR policy OR policies OR guideline\* OR valid\* OR reproducib\* OR reliab\* OR evaluat\* OR statistical OR statistician\* OR biostat\* OR bayesian OR theory OR theoretic\* ) AND NOT INDEX ( medline )

197 documents retrieved.

### **C.7.2 Excluded studies list**

#### **Not a review article, n=11**

AGYEMAN, A. S., SIEGEL, J. N. & LEPTAK, C. 2022. Establishing a public resource for acceptable surrogate endpoints to support FDA marketing applications. *Frontiers in Medicine*, 9, 820990.

ALONSO, A., VAN DER ELST, W., MOLENBERGHS, G. & FLOREZ, A. J. 2019. A reflection on the causal interpretation of individual-level surrogacy. *Journal of Biopharmaceutical Statistics*, 29, 529-540.

BURZYKOWSKI, T., BUYSE, M., MOLENBERGHS, G., ALONSO, A., VAN DER ELST, W. & SHKEDY, Z. 2020. Meta-analytic approach to evaluation of surrogate endpoints. *Handbook of Meta-Analysis*. CRC Press.

BUYSE, M., SAAD, E. D., BURZYKOWSKI, T. & PÉRON, J. 2020. Assessing treatment benefit in immuno-oncology. *Statistics in Biosciences*, 12, 83-103.

CIANI, O., GRIGORE, B. & TAYLOR, R. S. 2022a. 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.

CIANI, O., MANYARA, A. M. & TAYLOR, R. S. 2022b. Surrogate end points in cardio-thoracic trials: a call for better reporting and improved interpretation of trial findings. *European Journal of Cardio Thoracic Surgery*, 62, 02.

LUO, P., CAI, Z. & GENG, Z. 2019. Criteria for multiple surrogates. *Statistica Sinica*, 29, 1343-1366.

KNOTTNERUS, J. A. & KNOTTNERUS, B. J. 2022. Decision-making given surrogate outcomes. *Journal of Clinical Epidemiology*, 145, 174-178.

SHYR, Y. & SHYR, D. 2020. What constitutes a valid surrogate end point in cancer clinical trials? *JAMA Oncology*, 6, 1334-1335.

UDELSON, J. E. & WESSLER, B. S. 2022. Straining for surrogacy. *Jacc: Cardiovascular Imaging*, 15, 1542-1544.

ZHANG, A. D. & ROSS, J. S. 2019. Biomarkers as surrogate endpoints: ongoing opportunities for validation. *Journal of Law, Medicine & Ethics*, 47, 393-395.

**Focus was not to review statistical methods for surrogate validation, n=13**

ASSOULINE, S., WIESINGER, A., SPOONER, C., JOVANOVIC, J. & SCHLUETER, M. 2022. Validity of event-free survival as a surrogate endpoint in haematological malignancy: Review of the literature and health technology assessments. *Critical Reviews in Oncology-Hematology*, 175, 103711.

BAKER, S. G. & KRAMER, B. S. 2020. Simple methods for evaluating 4 types of biomarkers: surrogate endpoint, prognostic, predictive, and cancer screening. *Biomark Insights*, 15, 1177271920946715.

BELIN, L., TAN, A., DE RYCKE, Y. & DECHARTRES, A. 2020. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. *British Journal of Cancer*, 122, 1707-1714.

CIANI, O., GRIGORE, B., BLOMMESTEIN, H., DE GROOT, S., MOLLENKAMP, M., RABBE, S., DAUBNER-BENDES, R. & TAYLOR, R. S. 2021. Validity of surrogate endpoints and their impact on coverage recommendations: a retrospective analysis across international health technology assessment agencies. *Medical Decision Making*, 41, 439-452.

CIANI, O., MANYARA, A. M., DAVIES, P., STEWART, D., WEIR, C. J., YOUNG, A. E., BLAZEBY, J., BUTCHER, N. J., BUJKIEWICZ, S., CHAN, A. W., DAWOUD, D., OFFRINGA, M., OUWENS, M., HROBJARTSSON, A., AMSTUTZ, A., BERTOLACCINI, L., BRUNO, V. D., DEVANE, D., FARIA, C., GILBERT, P. B., HARRIS, R., LASSERE, M., MARINELLI, L., MARKHAM, S., POWERS, J. H., REZAEI, Y., RICHERT, L., SCHWENDICKE, F., TERESHCHENKO, L. G., THOMA, A., TURAN, A., WORRALL, A., CHRISTENSEN, R., COLLINS, G. S., ROSS, J. S. & TAYLOR, R. S. 2023. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine*, 65, 102283.

GENG, Z., LIU, Y., LIU, C. & MIAO, W. 2019. Evaluation of causal effects and local structure learning of causal networks. *Annual Review of Statistics and Its Application*, 6, 103-124.

GRIGORE, B., CIANI, O., DAMS, F., FEDERICI, C., DE GROOT, S., MOLLENKAMP, M., RABBE, S., SHATROV, K., ZEMPLENYI, A. & TAYLOR, R. S. 2020. Surrogate endpoints in health technology assessment: an international review of methodological guidelines. *Pharmacoeconomics*, 38, 1055-1070.

GYAWALI, B., HEY, S. P. & KESSELHEIM, A. S. 2020. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. *EClinicalMedicine*, 21.

MANYARA, A. M., DAVIES, P., STEWART, D., WEIR, C. J., YOUNG, A. E., WELLS, V., BLAZEBY, J., BUTCHER, N. J., BUJKIEWICZ, S., CHAN, A. W., COLLINS, G. S., DAWOUD, D., OFFRINGA, M., OUWENS, M., ROSS, J. S.,



TAYLOR, R. S. & CIANI, O. 2023. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. *Journal of Clinical Epidemiology*, 160, 83-99.

MURPHY, P., GLYNN, D., DIAS, S., HODGSON, R., CLAXTON, L., BERESFORD, L., COOPER, K., TAPPENDEN, P., ENNIS, K., GROSSO, A., WRIGHT, K., CANTRELL, A., STEVENSON, M. & PALMER, S. 2021. Modelling approaches for histology-independent cancer drugs to inform NICE appraisals: a systematic review and decision-framework. *Health Technology Assessment (Winchester, England)*, 25, 1-228.

PATIL, S., AGARWAL, V. & DRUPAD, H. S. 2022. Significance of emerging clinical oncology endpoints in support of overall survival. *Indian Journal of Cancer*, 59, S106-S118.

SAAD, E. D. & BUYSE, M. 2020. Statistical considerations for trials in adjuvant treatment of colorectal cancer. *Cancers*, 12, 19.

WALIA, A., HASLAM, A. & PRASAD, V. 2022. FDA validation of surrogate endpoints in oncology: 2005-2022. *Journal of Cancer Policy*, 34, 100364.

#### **Language not English, n=2**

CHENG, Y., LONG, Y., YANG, L., WANG, X., GUO, Q., LIN, Y., DU, L. & HUANG, J. 2022. Surrogate endpoint: definition evolution, evaluation and validation. *Chinese Journal of Evidence-Based Medicine*, 22, 964-970.

CIANI, O. 2020. Surrogate endpoints and health policy: why validation methods matter. *Epidemiologia e Prevenzione*, 44, 96-97.

## **D. Surrogate endpoints in cost-effectiveness analysis for use in health technology assessment – a qualitative analysis**

### **D.1 Objectives**

How have surrogate endpoints been dealt with in previous HTA evaluations by different HTA bodies and what lessons can be learnt from these assessments?

### **D.2 Methods**

Data was collected for this study using focus groups because this allowed for interaction between participants and open discussions. We conducted 3 focus groups globally, facilitated by 3 members of the working group and supported by NICE.

A purposive sample was used to identify members of staff from HTA bodies or their associated organisations. Bodies were identified that were known to use, develop or perform health economics assessments in their evaluation frameworks and decision making. Participants were invited through existing contacts of the working group. They were approached via email and asked to forward to relevant staff. All participants were sent an information sheet and asked to complete a consent form before joining the focus groups.

A document outlining 4 case studies involving the direct use of surrogate endpoints and the aims of the focus group was sent to each participant before the meeting to help facilitate the discussions. This included anticipated questions to allow participants time to think of responses or research previous actions within their body. The case studies included non-alcoholic steatohepatitis (NASH), obesity, Duchenne muscular dystrophy (DMD) and multiple myeloma. These topics were selected by the working group to reflect a range of surrogates and topics that different bodies may have some experience of or an interest in. For example, obesity was chosen as a topic that has validated and well-established surrogates that was likely to have

been seen by all bodies participating. Multiple myeloma was selected as a topic that has a newer, less well-established surrogate that fewer bodies have experience with but is likely to be of interest to many bodies in coming years.

The focus groups were divided into 3 parts: 1) open discussion asking participants to reflect on how their body handles surrogates and issues they have encountered, 2) discussion of specific case studies, and 3) recommendations or advice they would give to health technology developers. Discussions were conducted online via Microsoft Teams, video recorded and transcribed, using Teams, and then checked and edited by a member of the team.

Ethics approval was obtained from Newcastle University (reference number: NICE 42449/2023).

### **D.2.1 Outline of analysis**

A framework analysis approach was adopted for this study (Gale et al., 2013). An initial thematic framework was developed based on the research aims and discussions with the working group. This consisted of 4 key themes:

1. Guidance on the use of surrogate endpoints in economic modelling for HTA
2. Issues when using surrogate endpoints in economic modelling for HTA
3. Consequences of surrogate endpoints in economic modelling in HTA decision making
4. Suggested recommendations for using surrogate endpoints in economic modelling for HTA

These themes were used to guide the focus group discussions. Members of the NICE team applied the framework to one of the focus group transcripts independently and noted any additional themes with no obvious place in the initial framework. The framework was then updated and LF applied it to the 2 remaining focus groups.

The results were written into a draft report and circulated to focus group participants and participants that were unable to attend the focus group. They were given the opportunity to provide feedback and where appropriate this was incorporated into the analysis.

Once all analyses were complete and feedback was collated from participants and the working group a final framework was developed and used to structure the findings reported below. The 4 final themes were:

1. Experience of surrogate endpoints in economic modelling for HTA
2. Challenges of using surrogate endpoints in economic modelling for HTA
3. Guidance on the use of surrogate endpoints in economic modelling in HTA decision making
4. Recommendations for future use of surrogate endpoints in economic modelling for HTA

### **D.3 Results**

A total of 29 participants took part in the study between May and June 2024. There were 20 HTA bodies or associated organisations representing countries from Europe, the Americas and Asia-Pacific regions.

#### **D.3.1 Experience of using surrogate endpoints in economic modelling for HTA**

##### **D.3.1.1 Defining surrogate endpoints**

The discussions among participants suggested that there was a lack of clarity about the definition of a surrogate. Some participants viewed biomarkers and intermediate outcomes differently and, in some cases, did not consider them always to be a surrogate, despite both being considered to be surrogates by others. For example, some participants did not initially talk about progression-free survival (PFS) as a surrogate because this outcome is well-accepted link to survival as a surrogate in certain cancers.

### **D.3.1.2 Frequency of encountering surrogate endpoints in evaluations**

Across the discussions, regardless of their definition of a surrogate endpoint, participants agreed that their use in HTA submissions was increasing, with upwards of 50% of assessments using a surrogate endpoint. This was thought to be a consequence of trials targeting patients with earlier stage disease and commonly seen in more innovative technologies or rare diseases where evidence is less mature.

“It makes sense that we're seeing more submissions that are using either what we call a true surrogate or the intermediate surrogates. Because in oncology anyway, you've got all the treatments moving earlier and earlier. So [in the] adjuvant neoadjuvant space, you're not going to have overall survival benefit demonstrated in the trial.” European focus group participant.

### **D.3.1.3 Awareness of surrogate endpoints in the pipeline through early engagement with industry**

Participants from bodies that have early scientific advice services for technology developers highlighted the value of knowing about potential new surrogate endpoints in the pipeline through the advice requests. This gave them experience and knowledge of a surrogate before it is presented in a submission.

“We do have a scientific advice programme where we support a sponsor in their very, very early stages of planning a trial. And it's sometimes in that space we see what might happen in the pipeline 5 or 6 years down the line and in that space, we have seen sponsors who are at least planning for the primary or sometimes a composite co-primary clinical endpoint to be [...] NASH CRN scoring system.” Americas focus group participant.

## **D.3.2 Challenges of using surrogate endpoints in economic modelling for HTA**

### **D.3.2.1 Lack of guidance for technology developers around use of surrogate endpoints to inform decision making**

In general, most of the HTA bodies have limited guidance on how to handle surrogates in their methods manuals. Some bodies described additional internal documents to inform their evaluations that are not available to technology developers making submissions. Where internal guidance is available it was said to be centred around providing justifications for the inclusion of the surrogate in the economic model and exploring different scenarios to better understand the uncertainty due to the surrogate. There was consensus among participants that it would be useful to have more detailed guidance for the evaluation of clinical effectiveness based on surrogate endpoints and for its translation to the assessment of cost effectiveness.

### **D.3.2.2 Challenges when evaluating economic models that use surrogate endpoints**

Participants reported that in their experience, economic models are designed around the surrogate, rather than the disease pathway, which makes model outputs difficult to validate and apply to other contexts. It was also highlighted that, in some instances, the surrogate is embedded within the model structure in ways that can add complexity to their evaluation and make it challenging to separate the impact of assumptions around the surrogate relationship. An example of this is when a surrogate endpoint informs multiple transition probabilities that affect different outcomes.

Participants highlighted that in HTA, intermediate outcomes, that include a measure of function of symptoms being used as a substitute for a target outcome in addition to biomarkers, are considered to be surrogates. This can include outcomes such as quality of life outcomes or adverse events. They stressed that it is important that these outcomes are also validated if they are to be used in the economic modelling. As can occur in trials without

surrogates, participants also identified that in some cases there was disconnect between how the surrogate was measured in the clinical trial and what was being modelled, making it challenging to interpret and incorporate the evidence into the modelling.

#### **D.3.2.3 Challenges when assessing the uncertainty arising due to the use of surrogate endpoints**

Due to the opportunity cost of decision making based on surrogates and the potential losses if an incorrect decision is made, uncertainty needs to be mitigated. Some of the HTA bodies use temporary reimbursement decisions or managed access when the evidence to support the use of a surrogate is lacking. For example, a body might give a temporary reimbursement or time-limited recommendation along with the condition that further evidence is provided in 3 years' time. This flexibility alleviates some of the issues that arise when evaluations are a "single decision at a single time point" (European focus group participant). Other participants cited using scenario analyses to explore the uncertainty in their economic models such as considering the best and worst values for the surrogate, removing the surrogate entirely from the model and even using value of information analysis to identify and prioritise further research in areas where the further reduction in uncertainty associated with the use of the surrogate is worth the extra resources required to collect the data.

#### **D.3.2.4 Challenges when validating a surrogate endpoint**

The study 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. This was especially true for those HTA bodies where a decision to recommend a technology cannot be reversed and so sets a precedent for future decision making. However currently many participants felt like the requirements of some validation processes are complex, time consuming and costly for technology developers and the HTA bodies evaluating the evidence.

“The process involved in trying to validate a surrogate is complex and time consuming and it's almost like a submission in itself just to do that.” Asia-Pacific focus group participant.

It was also felt that it is important to consider the validity of a surrogate endpoint in the context of how and why it is collected. Often HTA bodies receive evidence on the validity of a surrogate after it has been evaluated by a regulator. This means it is collected and presented to meet the regulator’s needs and the HTA body is unlikely to contradict any conclusions drawn about a surrogate’s validity. However, the HTA body will still need to see and understand this evidence for their own decision making.

#### **D.3.2.5 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. This is exacerbated by the lack of guidance on how to transparently report the evidence relating to surrogates. This makes it difficult to conduct thorough evaluations for a technology and make robust recommendations. It was also identified as a stumbling block for many of the participants because it made it difficult to properly understand the evidence around the surrogate relationships.

“It's almost like getting information like with little teaspoons” Americas focus group participant.

#### **D.3.3 Guidance on the use of surrogate endpoints in economic modelling in HTA decision making**



### **D.3.3.1 Approach to providing additional guidance to technology developers around use of surrogate endpoints to inform decision making**

Participants highlighted the following aspects that should be considered when developing additional guidance around surrogates:

- Variation in capacity and resource of different HTA bodies: if technology developers are requested to provide extensive volumes of additional information, less-resourced HTA bodies may struggle to assess this information within strict timeframes.
- Flexibility of guidance to be applicable in different situations where surrogates may arise: if guidance is too stringent it may not accommodate situations where specific evidence is lacking, so it is important to present a tiered approach based on aspects like availability of evidence. For example, in NICE's health technology evaluations manual, the EQ-5D is set out as the preferred measure of quality of life in adults. However, multiple options are provided that would be acceptable where this evidence is not available from a clinical trial, for example, using data from the literature or mapping from other quality-of-life measures (NICE, 2023).
- Consistency within an organisation: there is a need to have an approach that can be consistently applied to all submissions within a body.  
Consistency between organisations: greater consistency between HTA bodies would support them to be able to use each other's technical work and analysis. However, participants receiving HTA submissions on a delayed timeline relative to other jurisdictions did identify instances where mature data was available, but technology developers had submitted earlier data based on earlier submissions to HTA bodies in other jurisdictions. Guidance needs to consider launch patterns and changes to data availability over time.

“There's a consistency aspect if you have that flexibility to do something different every time, then what you get is an inconsistent approach every

single time so that is not helpful to us and it's really not helpful to those who submit their evidence to us." Americas focus group participant.

### **D.3.3.2 Importance of the context and relevance of a surrogate endpoint to patients**

When making decisions and incorporating surrogates into economic modelling the HTA bodies suggested that it is always important to keep the focus on what matters to patients and to consider the broader context to the decision making. For example, is there an unmet need in this population, is this a rare disease with few treatment options, is this a life-threatening condition, is it a refractory population and resistant to previous treatments? One body suggested they have developed a multi-criteria decision analysis (MCDA) framework that helps them to consider criteria including effectiveness and severity of disease, especially in relation to rare diseases. This can form part of the processes that many bodies said already includes discussions with patients and clinical experts. Another body suggested they consider relevant regulator decisions that involved the surrogate.

The participants suggested that the use of surrogates in rare and ultra-rare disease can be challenging, and additional guidance would be helpful. The use of surrogates in this setting is increasingly common due to the novelty of treatments in this area, however, evidence to support a surrogate can be minimal due to the small population or the immature data collection in this setting and due to the absence of other treatments.

"The intermediate outcomes, we haven't really assessed those outcomes before and until they've been assessed and [...] had that clinical involvement, we don't know whether or not they're [...] a surrogate for what we should be looking for." European focus group participant.

### **D.3.3.3 Guidance on how to validate a surrogate endpoint for use in an economic model**

Participants were keen to have more guidance on how surrogates should be validated, including appropriate methods and how this should be presented. They wanted to see a set of standards that outlined how the relationship between the surrogate and final outcome should be calculated at an individual and a trial level.

The use of real-world evidence (RWE) and later data cuts from clinical trials were suggested by participants as a potential way to collect more information to validate a surrogate endpoint. For example, a temporary or conditional recommendation could be made on the basis that further evidence is collected to support the surrogate relationship.

“If we have to be making a decision early that there is the opportunity in the future to be re looking at those, but based on real world evidence, based on observational data. And we should be putting in the infrastructure to look at that.” European focus group participant.

While available evidence may be helpful in certain situations, there were concerns that in many situations data may be collected in a way that is not useful for developing the model and for the assessments of clinical effectiveness, especially if different in nature (heterogenous populations and comparators) to the clinical trial evidence that may also be employed to develop the model.

For some bodies, a literature review of clinical studies, quality of life studies, previous HTAs and economic models in the disease area is required to inform the evidence synthesis and economic model, but this was not the case for all. The suggestion was that guidance could outline how a literature review of evidence should be carried out and presented as part of the submission to support the use of a surrogate. This should include the clinical evidence underlying or justifying the use of a surrogate. In the experience of the

participants, on the limited occasions when there is a systematic review of evidence of a surrogate, HTA bodies said they still rely on discussions with their experts (clinical and public experts) to interpret this evidence and provide their views on the surrogate.

Some participants were keen to stress that past economic modelling approaches and HTA decisions should not dictate future decisions but should help to inform them and so the evidence presented should be interpreted with this in mind and decisions reflect more up-to-date knowledge.

“Really you should build on what's been done before. Not carry on doing it.”  
European focus group participant.

#### **D.3.3.4 Guidance to facilitate transparent reporting of evidence (setting expectations with technology developers)**

The participants wanted to see recommendations that would facilitate the better reporting of evidence from technology developers that would help them to understand the surrogate relationships, how the economic models had been built and help them to explore the uncertainties themselves. 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, 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, their limitations and their strengths. It was felt that guidance would help to align the expectations of both bodies and developers during the submission and assessment process.

“I would want to see a transparent model that would allow me to make changes to the model and to test various scenarios”, Asia-Pacific focus group participant.

#### **D.3.3.5 Database of surrogates**

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. However, it was noted by other participants that this could be challenging due to the different contexts that surrogates can be applied even for the same active ingredient or indication.

"I know it's very ideal, a living library of [...] model relationships between surrogates and final outcomes would be lovely." Asia-Pacific focus group participant.

#### **D.3.4 Suggested recommendations for future use of surrogate endpoints in economic modelling for HTA**

The following section outlines specific recommendations identified during the focus groups.

##### **D.3.4.1 Recommendations for developing guidance on surrogate endpoints**

The focus group participants highlighted the following suggestions as important to consider when developing the guidance on surrogates.

Participants wanted guidance that:

1. 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 things factors such as availability of evidence
2. reflects the varying structures and resources of HTA bodies globally
3. considers the different launch patterns of HTA bodies that mean some bodies receive submissions later than others.
4. is mindful of areas that the HTA bodies identified as particularly challenging, for example, for new technologies aimed at rare diseases.

These points will be considered in the development of the final recommendations, alongside the evidence from the other work packages of the wider project, to ensure they reflect the wide range of HTA bodies, the technologies they evaluate and the different organisational structures they have.

#### **D.3.4.2 Recommendations for developing an economic model using a surrogate endpoint**

The focus group participants suggested the following recommendations as being important to consider when developing an economic model that uses a surrogate in HTA. These points will be considered in the context of the other work packages of the wider project and where appropriate included in the final set of recommendations.

1. Model conceptualisation, design and structure:
  - a. The model conceptualisation should be based around the disease area rather than the surrogate.
  - b. 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.
  - c. The use of a surrogate endpoint in an economic model should be justified and supported by the data including how the context of the clinical assessment supports the validity of the surrogacy relationship.
  - d. 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.
  - e. 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 via the surrogate endpoint and the duration of

that effect should be parameters which can be explored within the model.

2. Validation of the surrogate endpoint should include:
  - a. Clinical expertise to understand the clinical relevance of the surrogate endpoint.
  - b. A comprehensive review of the evidence on the relationship between the final outcome and the surrogate endpoint.
  - c. Real-world evidence collection could be used as an option to validate the surrogacy relationship.
3. Assessing uncertainty in the economic model
  - a. Test assumptions around relationship of surrogate endpoint to final outcome in scenario analyses, for example, best and worst-case scenarios
  - b. 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.

#### **D.3.4.3 Recommendations for reporting an economic model using a surrogate endpoint**

Focus group participants recommended the transparent reporting of an economic model using a surrogate endpoint would be helpful to understand the evidence and uncertainties associated with the use of a surrogate endpoint in economic modelling.

#### **D.3.5 Recommendations for further work**

Many of the focus group 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. The database could build on the [FDA's table of surrogate endpoints that were the basis of drug approval or licensure](#) (US Food and Drug Administration, 2022). However, the development of such a resource is beyond the scope of the current research project and it is recommended this is explored in further work.

## **D.4 Summary**

In focus groups with members of staff from HTA bodies and their associated organisations, further guidance on the use of surrogate endpoints in economic modelling of health technology evaluations was strongly emphasised.

Recommendations from participants have been summarised that aim to promote consistency within and between bodies and clarify expectations between technology developers and bodies.

## **D.5 References**

Gale, N., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*, 13.

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>