

# Development of a framework and decision tool for the evaluation of health technologies based on surrogate endpoint evidence

Oriana Ciani<sup>1,2</sup> | Bogdan Grigore<sup>3</sup> | Rod S. Taylor<sup>4,5</sup> 

<sup>1</sup>Centre for Research on Health and Social Care Management, SDA Bocconi, Milan, Lombardia, Italy

<sup>2</sup>Evidence Synthesis & Modelling for Health Improvement, College of Medicine and Health, University of Exeter, Exeter, Devon, UK

<sup>3</sup>Exeter Test Group, College of Medicine and Health, University of Exeter, Exeter, Devon, UK

<sup>4</sup>MRC/CSO Social and Public Health Sciences Unit & Robertson Centre for Biostatistics, Institute of Health and Well Being, University of Glasgow, Glasgow, Scotland, UK

<sup>5</sup>College of Medicine and Health, University of Exeter, Exeter, Devon, UK

## Correspondence

Rod S. Taylor, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, Scotland, G3 7HR, UK.  
Email: [rod.taylor@glasgow.ac.uk](mailto:rod.taylor@glasgow.ac.uk)

## Funding information

European Commission, Grant/Award Number: 779306

## Abstract

In the drive toward faster patient access to treatments, health technology assessment (HTA) agencies and payers are increasingly faced with reliance on evidence based on surrogate endpoints, increasing decision uncertainty. Despite the development of a small number of evaluation frameworks, there remains no consensus on the detailed methodology for handling surrogate endpoints in HTA practice. This research overviews the methods and findings of four empirical studies undertaken as part of COMED (Pushing the Boundaries of Cost and Outcome Analysis of Medical Technologies) program work package 2 with the aim of analyzing international HTA practice of the handling and considerations around the use of surrogate endpoint evidence. We have synthesized the findings of these empirical studies, in context of wider contemporary body of methodological and policy-related literature on surrogate endpoints, to develop a web-based decision tool to support HTA agencies and payers when faced with surrogate endpoint evidence. Our decision tool is intended for use by HTA agencies and their decision-making committees together with the wider community of HTA stakeholders (including clinicians, patient groups, and healthcare manufacturers). Having developed this tool, we will monitor its use and we welcome feedback on its utility.

## KEYWORDS

cost-effectiveness, decision tool, health technology assessment, surrogate endpoints, validation

## 1 | INTRODUCTION

A surrogate endpoint is defined as a proxy outcome that can substitute for and predict a relevant final patient-relevant outcome, such as mortality or health-related quality of life (Ciani et al., 2016, 2017; DeMets et al., 2020; Gyawali et al., 2019; Robb et al., 2016). In recent years, regulatory agencies, including the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) in the United States (US), have used various accelerated programmes to approve therapies based on surrogate endpoints (Salcher-Konrad et al., 2020; US Food Drug Administration, 2021).

Whilst surrogate endpoints may help speed up the evaluation and approval of therapies by allowing faster outcome accrual and shorter and smaller clinical trials (Ciani et al., 2016, 2017), reliance on such endpoints introduces decision uncertainty for healthcare policy makers. For regulators, surrogate endpoints may fail to fully capture the complete risk-benefit profile of a new therapy (Fleming & DeMets, 1996). In the health technology assessment (HTA) and payer setting, reliance on a surrogate

-----  
This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. Health Economics published by John Wiley & Sons Ltd.

endpoint may result in an inaccurate assessment of a therapy's value. Surrogate endpoints have been shown to result in larger treatment effects than final outcomes (Ciani, Buyse, et al., 2013; Walter et al., 2012) thus leading to a systematic overestimation of clinical (and underestimation of the costs) of a new or emerging technology.

Therefore, it has been recommended that the use of surrogate endpoints be limited only to those that have been validated appropriately (Ciani et al., 2014, 2016, 2017; Schuster Bruce et al., 2019). Ideally, such validation requires evidence from multiple randomised controlled trials, consistently demonstrating a strong statistical association between the treatment-induced change in the surrogate endpoint and the final patient-relevant endpoint (Ciani et al., 2016, 2017). The US FDA website states: "Clinical trials are needed to show that surrogate endpoints can be relied upon to predict, or correlate with, clinical benefit. Surrogate endpoints that have undergone this testing are called validated surrogate endpoints" (US Food Drug Administration, 2021). They have published an approved listing of surrogate endpoints across a wide range of diseases, including cancer (disease or progression free survival), asthma (forced expiratory volume in 1 s), chronic kidney disease (glomerular filtration rate), and diabetes (glycosylated hemoglobin; US Food Drug Administration, 2021).

In 2017, Ciani et al. proposed a methodological framework for the incorporation and reporting of the use of surrogate endpoints in HTA (Ciani et al., 2016). As shown in Figure 1, this framework recommends a three step approach: (1) to establish the level of evidence available (i.e., whether the relationship between the putative surrogate endpoint and final patient relevant outcomes of interest is supported by clinical plausibility ['level 3' evidence], observational data ['level 2' evidence], and clinical trial data ['level 1 evidence']); (2) to assess the strength of the association between the surrogate and final patient relevant outcomes: observational or treatment level effect association/correlation; and (3) to quantify the expected effect on the final patient relevant outcome(s) given the observed effect on the surrogate endpoint.

Despite the development of this and other evaluation frameworks for surrogates (IQWiG, 2011; Lassere et al., 2012), empirical evidence on their application or uptake by HTA agencies or payers is scarce (Ortiz et al., 2021). Furthermore, the traditional focus of the use and application of surrogate endpoints has been in the licensing and coverage of drugs and biologics with little application to other medical technologies, particularly medical devices (Ciani et al., 2016, 2017). Given this context, within work package 2 (WP2) of the COMED (COMED (Pushing the Boundaries of Cost and Outcome Analysis of Medical Technologies) European Union Horizon 2020 funded program, we sought to research current international HTA and payers' practice on their use of surrogate endpoints and explore whether this practice varied across technologies, such as drugs versus devices (COMED, 2021). The overarching goal of WP2 was to develop a framework to support HTA agencies and payers in their decision-making and policy processes when faced with the evaluation of health technologies based on evidence from surrogate endpoints.

In this paper we first overview the methods and findings of our four empirical studies undertaken as part of COMED WP2: (1) an international survey of HTA agencies on their current methodological guidance for use of surrogate endpoints (Grigore et al., 2020), (2) a review of the current international practice of the application and validation of surrogate endpoints in HTA reports and impact on coverage decisions (Ciani et al., 2021), (3) a qualitative study exploring views of stakeholders on the

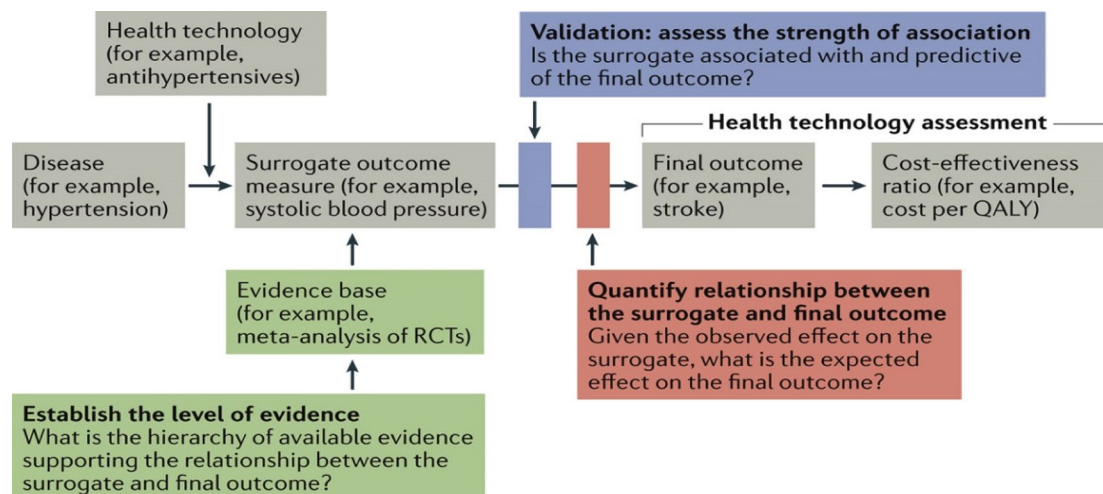


FIGURE 1 Three-step evaluation framework for assessment of surrogate endpoints

issue of surrogacy in HTA decision-making, and (4) a pilot choice experiment to better understand the trade-offs made by HTA stakeholders on their use of surrogate endpoints evidence as a basis for value determinations. Second, we seek to synthesize and apply our research findings, in context of the wider contemporary body of methodological and policy-related literature on surrogate endpoints, to revisit the three-step framework and develop a web-based decision tool to support HTA agencies and payers when faced with surrogate endpoint evidence.

## 2 | METHODS

### 2.1 | International survey of HTA agencies approaches to surrogate endpoints

We updated the listing of European HTA agencies of a previous survey of surrogate endpoints by Velasco-Garrido published in 2009 (Velasco Garrido & Mangiapane, 2009) to include all organisations currently listed as members of major HTA networks (as of March 2018) of Health Technology Assessment International (HTAi), European network for Health Technology Assessment (EUnetHTA), and International Network of Agencies for Health Technology Assessment (INAHTA). We also purposively included Australia and Canada to learn from these jurisdictions with more mature HTA processes.

Following a detailed review of the website of each HTA agency, we considered how their methods guidelines addressed the handling of surrogate outcomes, that is, (1) the level of evidence required, (2) methods of validation, (3) thresholds of acceptability, and (4) whether the guidance was specific to pharmaceuticals or medical devices or both.

### 2.2 | Analysis of HTA international reports practice on validation of surrogate endpoints and impact on coverage decisions

Given their extensive HTA portfolio (technology appraisal guidance, medical technologies guidance, and diagnostics guidance reports), we first screened the health technology guidance published by National Institute of Care Excellence (NICE) in United Kingdom between May 2013 and June 2018 for reports including surrogate endpoint evidence. Second, based on a selected list of NICE reports, we then identified HTA evaluation reports for the same health technology and clinical indication from a further sample of eight international HTA agencies chosen to include different geographical areas, the most prominent HTA organisations, are known (from our survey above) to have HTA methods guidelines, and expertise in report languages within our research team (English, German, and Hungarian). These agencies included Health Improvement Scotland (HIS)/Scottish Medicines Consortium (SMC), Haute Autorité de Santé (HAS) in France, Pharmaceutical Benefits Advisory Committee (PBAC) and Medical Services Advisory Committee (MSAC) in Australia, Canadian Agency for Drugs and Technologies in Health (CADTH) in Canada, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)/Gemeinsame Bundesausschuss (G-BA) in Germany, Zorginstituut Nederland (ZiN) in the Netherlands, and Országos Gyógyszerészeti és Élelmezés-egészségügyi Intézet (NIPN) in Hungary. For each included report from each agency, data was extracted on the methods approach to handling of surrogate endpoint (based on the three-step framework), the judgment on the acceptability of the surrogate endpoint (e.g., “increase in total kidney volume correlates to growth in cyst volume [...] was considered to be an appropriate surrogate for disease progression”), and their coverage recommendation. Mixed effects logistic regression models were used to test the impact of validation approaches on the agency's acceptability of the surrogate endpoint and their coverage recommendation.

### 2.3 | Qualitative study of HTA stakeholders' views on the use of surrogate endpoints

A purposive sampling technique (Palinkas et al., 2015) was used to identify and select the individuals (or organisations) within each of the HTA stakeholder groups across Europe.

1. Clinical/healthcare professionals or representatives of professional organisations.
2. Payers or representatives of HTA agencies
3. Healthcare regulators
4. Researchers in regulatory science, biostatistics, health economics, and HTA.
5. Medicines and medical devices manufacturers.

The study recruitment was drawn from stakeholders known to the COMED consortium partners. Potential participants were sent an email invitation to participate, including background information on the study and expected time commitment. Upon receiving confirmation to participate, they were further contacted by a member of the team to set up the interview date. Up to two additional follow-up messages were sent in case of no reply. Ethical approval was granted by the University of Exeter, College of Medicine and Health, Research Ethics Committee (UEMS REC reference number: 18/09/182) for this qualitative interview-based study.

Interviews were conducted either face-to-face or by telephone. A semi-structured schedule was developed to guide interviews and included questions on: (1) the meaning/definition of surrogate endpoint, (2) methods and guidance available to validate such endpoints, (3) level of surrogate-related uncertainty considered acceptable for regulatory and value determinations, and (4) specific consideration of these issues for drugs versus medical devices. The four questions included a closed component, where interviewees were asked to provide a score on a six-point Likert scale (e.g., on a scale from 0 (not relevant) to 5 (essential), how important is...?). Piloting was conducted with three potential participants, and the draft interview guides were revised accordingly. A copy of the interview schedule is provided in the Appendix 3. Interviews were conducted by a member of the research team either in English or native language of the interviewee, where possible.

A total sample of >20 participants was judged to be achievable and provide sufficient information for qualitative analysis (Vasileiou et al., 2018). Interviews were digitally recorded, transcribed, and (where necessary) translated into English. Transcripts were deidentified (pseudonymized) and analyzed thematically, according to the principles of reflexive thematic analysis (Braun & Clarke, 2006), independently coded the qualitative data, analyzed by two members of the research team (OC/BG). Any discrepancies/differences in interpretation were resolved through discussion and involvement of a third researcher (RST). Quantitative data from closed questions were summarized descriptively.

## 2.4 | Pilot choice experiment with HTA stakeholders on their decision making when faced with surrogate endpoint evidence

We recruited participants from two sources: (1) postgraduate students from the Bocconi University Master Program in International Healthcare Management, Economics and Policy (MIHMPEP) studying health economics; (2) colleagues from partner institutions in the COMED consortium. It was anticipated that all participants had adequate background knowledge to complete the choice tasks. Ethics approval obtained from the Bocconi University Ethics Committee (SA000160/September 28, 2020). The study was conducted online (Qualtrics, Provo, UT, USA) and presented/collected information to participants as follows: an introduction to the study (explaining the rationale of the study) and collection of their demographic details, choice tasks, and debriefing questions asking for feedback on the questionnaire (structure, content and intelligibility and any suggestions for improvement). A copy of the online survey script is provided in Appendix 1.

The choice task attributes (see Appendix 2) were developed iteratively by the project team based on our understanding of the challenges in the use of surrogate endpoints in HTA (Ciani et al., 2021) and our findings of the three empirical WP2 studies outlined above. A two-step vignette was used where participants were asked to: (i) first determine the acceptability of the surrogate endpoint illustrated in the vignette, and (ii) then, make a coverage decision. Two hypothetical—but based on real world examples - vignettes were developed: one based on a pharmaceutical (i.e., a drug therapy indicated for the treatment of neuromuscular symptoms of a metabolic myopathy, ‘Scenario 1’) and the other, a medical device (i.e., a device-based procedure for the treatment of resistant hypertension using renal denervation, ‘Scenario 2’). Each vignette was framed within a detailed case study of the hypothetical surrogate endpoint—Scenario 1: serum levels of the enzyme matrix metalloproteinase 9 (MMP); Scenario 2: systolic blood pressure, each with four attributes of the surrogate endpoint at two levels:

1. Source of evidence for the validation of the surrogate endpoint (stronger vs. weaker evidence);
2. Class of therapies providing evidence for the validation of the surrogate endpoint (same vs. different treatment class);
3. Strength of association between the surrogate and patient-relevant endpoint (weaker vs. stronger association);
4. Surrogate threshold effect (STE) that is, the minimum effect on the surrogate to predict a significant effect on the patient-relevant endpoint (Lower vs. higher STE likely to be observed; Ciani et al., 2016, 2017).

Based on the information provided, the participant was asked to characterize the surrogate measure as either ‘acceptable’ (i.e., valid) or ‘not acceptable’.

Following their judgment on the acceptability of the surrogate measure, the participant was asked to judge as a payer whether the therapy should be approved, given four contextual factors (each with 2 levels):

1. Condition prevalence (lower vs. higher prevalence);
2. Condition baseline (health-related quality of life) utility score (on 0–1 scale; more severe vs. less severe disease);
3. Comparator (no alternative treatment(s) versus existing alternative treatment(s));
4. Effect on the final patient-relevant outcome based on ‘immature data (positive vs. negative trend).

We sought to recruit ~20 participants based on previous choice experiments (Braun & Clarke, 2006) assessing determine feasibility and acceptability. Given the pilot nature of this study, the primary focus of our data analysis and presentation was descriptive. We report a summary of the recruitment process, response rate, questionnaire completion rate, and summary characteristics of the participants. Response counts for the choice tasks were tabulated and individual choice responses summarized through a Sankey flow diagram.

## 3 | RESULTS

### 3.1 | International survey of HTA agencies approaches to surrogate endpoints

Of the 74 HTA agencies included, 44 had methods guidelines, 29 (66%) of which included consideration of the handling of surrogate endpoints in their methods guidance. Although the extent to which guidelines provided specific consideration on the use of surrogate endpoints varied across agencies, the majority were based on the guidance on surrogate endpoint methods of the EUnetHTA collaboration ‘Endpoints used in relative effectiveness analysis of pharmaceuticals: Surrogate Endpoints’ published in November 2015 (EUnetHTA, 2015). Seven agencies had methods guidelines that included detailed methodological consideration of surrogate endpoints—IQWiG (Germany), NICE (UK); AOTMiT (Agency for Health Technology Assessment and Tariff System, Poland); INFARMED (National Authority of Medicines and Health Products, Portugal); PBAC and MSAC (Australia), and CADTH (Canada). These guidelines included recommendations of the methods for the validation of surrogate endpoints and, in two cases, cut-offs for the acceptance of surrogates based on their validation (IQWiG, PBAC). Two countries had separate drug and medical device specific HTA processes (i.e., UK: NICE Technology Appraisal and Medical Technology Evaluation programmes; Australia: PBAC and MSAC programmes). In both case, methods guidance for medical devices appeared less specific and did not include any specific recommendations on the handling of surrogate endpoints.

### 3.2 | Analysis of HTA international reports practice on validation of surrogate endpoints and impact on coverage decisions

We included 23 NICE technology assessments of which: 21 (91%) were pharmaceuticals and 2 (9%) were medical devices; 12 (52%) in oncology indication, 3 (13%) cardiovascular indications, 2 (9%) for either an endocrinology or a nephrology indication, and the remainder spread across a variety of conditions (i.e., chronic hepatitis C, biliary cholangitis, vitreomacular traction, pulmonary fibrosis). We identified a total of 124 reports across all 8 HTA agencies matching these NICE appraisals. There was a median of 5 evaluations per technology: 4 technologies (alirocumab, evolocumab, pirfenidone, ribociclib) were evaluated by all 8 agencies and one technology (Geko device; FirstKind Ltd High Wycombe, UK) was only evaluated by NICE. Of the 124 included reports, 61 (49%) discussed the level of evidence to support the relationship between the surrogate and the final patient relevant outcome, 27 (22%) reported a correlation coefficient/association measure, and 40 (32%) quantified the expected effect on the patient-relevant final outcome. The level of depth and scrutiny applied by different agencies in relation to the validation of surrogate endpoints varied considerably: NICE was the agency most likely to report on the level of evidence, strength of association, and quantification of effect related to the validation of a surrogate endpoint. However, the statistical association (e.g.,  $R^2$ , Spearman's  $r$  correlation coefficient) and quantification of the expected treatment effect on the patient-related final outcome, based on the observed effect on the surrogate endpoint, were rarely reported. Overall, the surrogate endpoint was deemed acceptable in 49 (40%) reports ( $k$ -coefficient 0.10,  $p = 0.004$ ). Any consideration of the level of evidence (level 1 to 3) was associated with acceptance of the surrogate endpoint as valid (odds ratio [OR], 4.60; 95% confidence interval [CI], 1.60 to 13.18,  $p = 0.005$ ). However, we did not find strong evidence of an association between accepting the surrogate endpoint and the coverage recommendation (OR, 0.71; 95% CI, 0.23 to 2.20;  $p = 0.55$ ).

### 3.3 | Qualitative study of HTA stakeholders' views on the use of surrogate endpoints

A total of 27 interviews were conducted between October 2019 and February 2020 from individuals based across Europe (Austria 1, Switzerland 1, Germany 5, France 1, Hungary 4, Italy 4, Netherlands 3, Sweden 1, and UK 7) that included representation from regulators (3), payers & HTA agencies (9), clinicians (6), researchers (4), and healthcare manufacturers (5).

Respondents rated the importance of having a “valid” surrogate endpoint (when evidence on patient-relevant endpoints was absent), as ‘high; with a mean rating of 4.3 (median 4.0; with rating scale of 0 ‘not relevant’ to 5 ‘essential’). Eleven (41%) respondents noted that surrogates were acceptable in HTA with one respondent stating “*while ideally one would have information on the patient-relevant endpoint, if that is not available, then the surrogate is the next best thing*” [E09]. When asked about surrogate endpoint validity, 8 (30%) respondents referred to the concept of surrogacy, that is, (1) a candidate surrogate endpoint must be shown to forecast outcome in the same fashion as a prognostic marker (‘individual-level’ surrogacy and (2) that the effect of treatment on the candidate surrogate endpoint must be closely correlated with the effect of treatment on the patient-relevant endpoint (‘trial-level’ surrogacy). However, illustrated by the following quotation “*there is no generally accepted criterion, which would be sufficient to prove validity*” [E21], 6 (22%) respondents pointed out there they believed there to be agreed criterion for surrogate validity. Most respondents (17, 63%) indicated that they would like to see at least level 2 evidence (i.e., observational studies showing the association of the surrogate endpoint and the final outcome) and 10 (37%) indicated they would prefer Level 1 evidence. Respondents indicated that it would be slightly more difficult to validate surrogate endpoints for a medical device as compared to drug (mean and median ‘2’, where 0 is ‘same difficulty’ and five is ‘much more difficult’). Eight (30%) interviewees stated that they considered drugs and medical devices to be similar and illustrated by the following quotes: “*in principle, the challenge is exactly the same*” for medical devices [E13] (“*insulin pumps in diabetes; there would not be a problem of sample size. I don't see why the reasoning should be different with medtech. We should follow the same requirements*” [E22]) and the issue was more one of the disease area (“*the problem lies less in the type of therapy (e.g., radiotherapy vs. drugs) but in the specific indication (e.g., prostate cancer)*” [E08]). When asked about what methods are available and used to validate potential surrogate endpoints, 14 (52%) interviewees replied they knew none. Four researcher respondents pointed to the meta-analytic approach on stating “*use only aggregate data from RCTs, which are publicly available, so much easier to implement in a HTA setting*” [E02].

The majority (15, 56%) of respondents reported they did not know any specific guidelines for surrogate endpoints. One respondent referred to the approach followed in the IQWiG methods, six (22%) referred to NICE methods guidelines (EUnetHTA, 2015), and four (15%) cited the EUnetHTA guidelines (Braun & Clarke, 2006). Only 2/27 (7%) respondents indicated that were ‘completely satisfied’ with current available guidance for the use of surrogate endpoints with a satisfaction mean score of ‘2’ (median 2; scale: 0 ‘not at all’ to 5 ‘completely satisfied’) and that they thought it was very important to improve guidance for the use of surrogate endpoints in health care decision making (mean 4.2, median 4). The need for harmonization in HTA agencies between jurisdictions were highlighted (“*England and Wales, sometimes Scotland, made a different decision based on the same evidence*” [E09]), “*HTA agencies sometimes have very different approaches to surrogate endpoints*” [E13]) as was the need for harmonization between regulators and payers/HTA agencies (“*I have come across situations where regulators have approved treatments based on very weak evidence*” [E09]).

### 3.4 | Pilot choice experiment with HTA stakeholders on their decision making when faced with surrogate endpoint evidence

A total of 20 participants (13 postgraduate students and 7 COMED partners) completed online choice experiments. The recruitment process is summarized in Figure 2. There were no differences in characteristics between those who did and did not complete the online task.

A summary of the characteristics of survey completers is provided in Table 1.

Excluding 5 participants who did not complete the task in one sitting, the mean time to complete the survey was 14 min (range: 5–24 min). Overall, 16 (80%) respondents strongly or somewhat agreed with the statement “The background information provided clear explanation about the purpose of the study”, with none strongly disagreeing.

Fourteen (70%) and 10 (50%) agreed the two scenarios were plausible and agreed with the statement “The choice tasks were relatively easy to perform” respectively. Specific feedback comments included minor typographical or syntax errors, or suggestions for clearer wording (e.g., for statement “With an innovative technology, it is always acceptable to rely on evidence for validating a surrogate endpoint derived from a previous class of therapies”).

FIGURE 2 Recruitment flow diagram

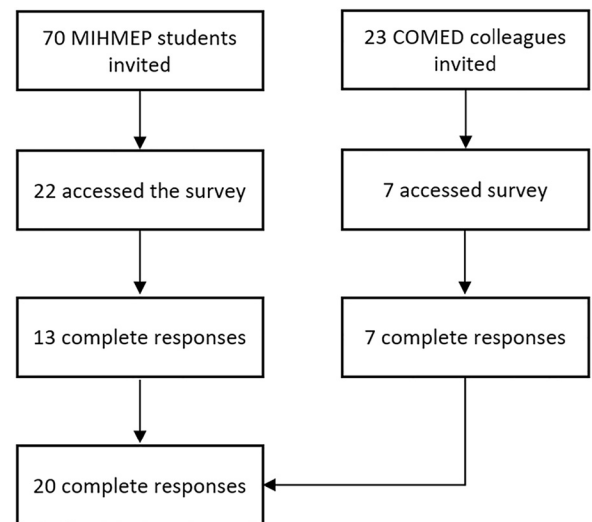


TABLE 1 Participant characteristics of survey completers

	Number (percentage, <i>N</i> = 20)
<b>Population</b>	
MIHMEP students	13
COMED partners	7
<b>Gender</b>	
Male	7 (35%)
Female	13 (65%)
Prefer not to say	–
<b>Age group</b>	
18–24	–
25–34	14 (70%)
35–44	5 (25%)
45–54	1 (5%)
55–64	–
65–74	–
75 or older	–
<b>Background</b>	
Economics	9 (45%)
Engineering	2 (10%)
Humanities/Law	1 (5%)
Medicine	1 (5%)
Nursing/healthcare profession	2 (10%)
Pharmacy/Biomedical sciences	5 (25%)
<b>Current occupation</b>	
Academia	10 (50%)
Public agency/competent authority/government	2 (10%)
Industry	4 (20%)
Healthcare organization	–
Consulting firm	4 (20%)

In Scenario 1, seven (35%) participants assessed the surrogate endpoint as unacceptable; of these seven, only four chose not to reimburse the technology. In total, six (30%) chose to reimburse the technology described in the scenario. In Scenario 2, nine (45%) respondents chose not to accept the surrogate as valid; the majority choosing not to reimburse the technology. In total, four (20%) of the respondents chose to reimburse/approve the technology (see Table 2).

The individual choices by each participant in the two scenarios are presented in Figures 3 and 4 (and details is provided Appendices 4 and 5). To illustrate the process, we describe the choices made by two of the participants. In Scenario 1 (see Figure 3), participant ID13 was presented with the permutation IC10 of surrogate endpoint evidence (i.e., validation evidence based on observational data, derived from the same class of therapies, with a stronger association between surrogate and final endpoints, and a lower surrogate threshold effect) and concluded that the surrogate endpoint was likely an acceptable outcome measurement. They were then presented with permutations IC7T of the technology characteristics (i.e., higher prevalence, more severe disease, where therapeutic options already exist, and a positive effectiveness suggested by the immature final endpoint data), based on which they concluded the technology should likely be reimbursed. In Scenario 2 (see Figure 4), participant ID16 considered the surrogate as acceptable based on permutations IIC9 (i.e., evidence from a meta-analysis of randomised controlled trials on the same therapeutic class, with a weaker association and a lower surrogate threshold effect). When presented with the technology characteristics permutations IIC11 T (lower prevalence, less severe disease, with existing therapies and immature final point evidence favoring the control), the participant chose not to reimburse the technology.

## 4 | DISCUSSION

Whilst the use of surrogate endpoints as primary endpoints in clinical trials can accelerate regulatory approval and market access for selected healthcare technologies, they increase decision uncertainty for HTA agencies and payers faced with making coverage and funding decisions for health technologies. Work package 2 of the EU H2020 funded COMED program undertook four linked empirical research studies with the overarching goal of improving the knowledge base to support HTA agencies and payers in their handling of surrogate endpoints.

Our research findings have important implications for current HTA and policy practice. First, our updated survey of international HTA agencies (Grigore et al., 2020) demonstrates there has been an increase in the development of methodological guidance for the use of surrogate endpoints over the last decade, largely driven by the adoption of EUnetHTA guidance on surrogates published in 2015 (Braun & Clarke, 2006). Nevertheless, only a small number of HTA settings (Australia, Canada, Germany, and UK) have developed what we deemed to be sufficiently detailed advice on the statistical methods of surrogate validation or clear transparency on their criterion for acceptance (or rejection) of surrogates (COMED, 2021). This was further evidenced by our analyses of HTA reports (Ciani et al., 2021) that showed considerable variability across HTA agencies in their application of these validation methods or criteria for surrogate endpoint acceptance. Our interview-based study highlighted variability across stakeholder groups in their confidence, familiarity, and understanding of these issues. Research (including statisticians and health economists based in academia, the healthcare industry, and regulatory agencies) had detailed knowledge and understanding of surrogate validation methods. Whilst often critical of the value surrogate endpoints, we found patient representative groups to be much less familiar with technical approaches to assessing surrogate validity. Across the payers/*Payers or representatives of HTA agencies* interviewed there appeared to a range of expertise and confidence in handling of surrogate endpoints. Such heterogeneity in institutional expertise and methods may explain the variability in acceptance and coverage decisions that we observed when we compared HTA agencies across a common basket of health technologies decisions.

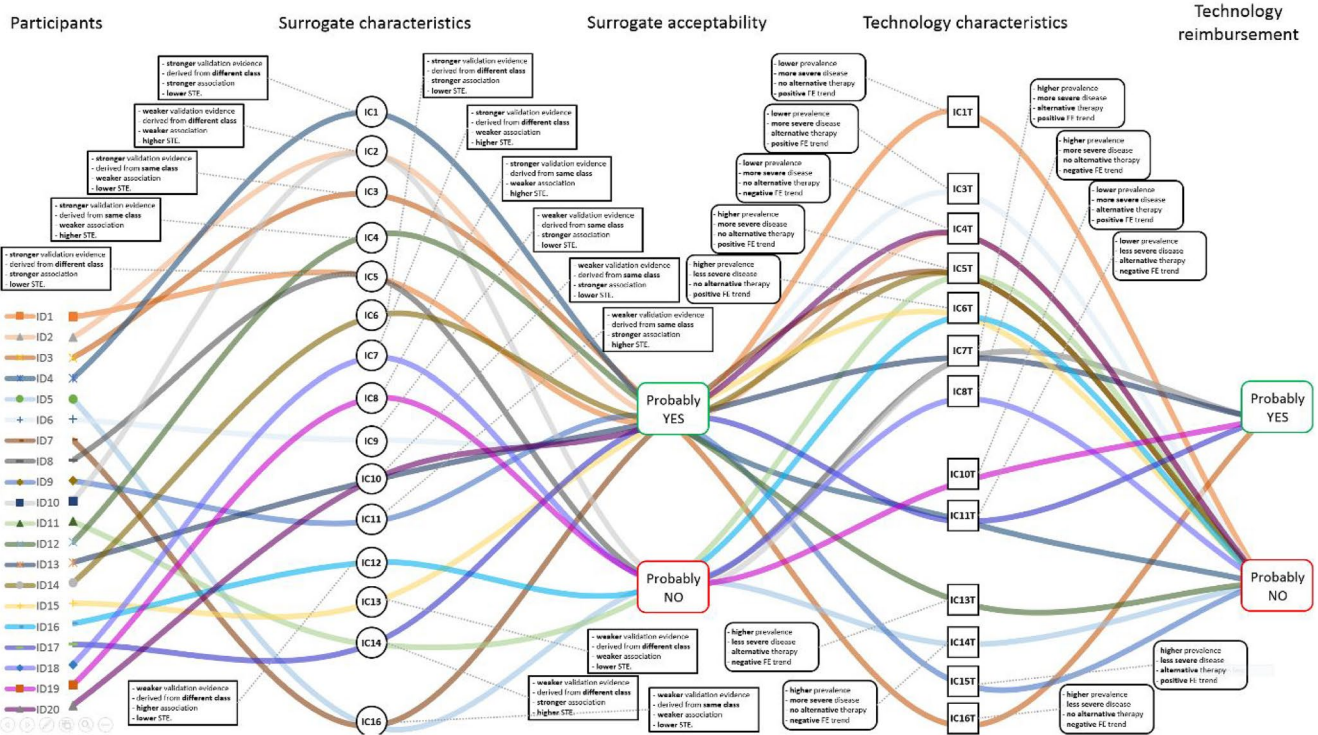
Second, our choice experiment study and analysis of HTA reports (Grigore et al., 2020) showed an intriguing ‘disconnect’ by HTA agencies and payers in their judgment of surrogate acceptance and their coverage decision. Both these studies indicated that whilst stronger validation evidence (e.g., regression analysis of randomised controlled trials demonstrating the association between the intervention effect on the surrogate endpoint and final patient-relevant outcome) would lead to increased likelihood of an agency's ‘acceptance’ of the surrogate, our analyses showed that this did not necessarily appear to translate to a positive cover-

**TABLE 2** Summary of responses to choice tasks

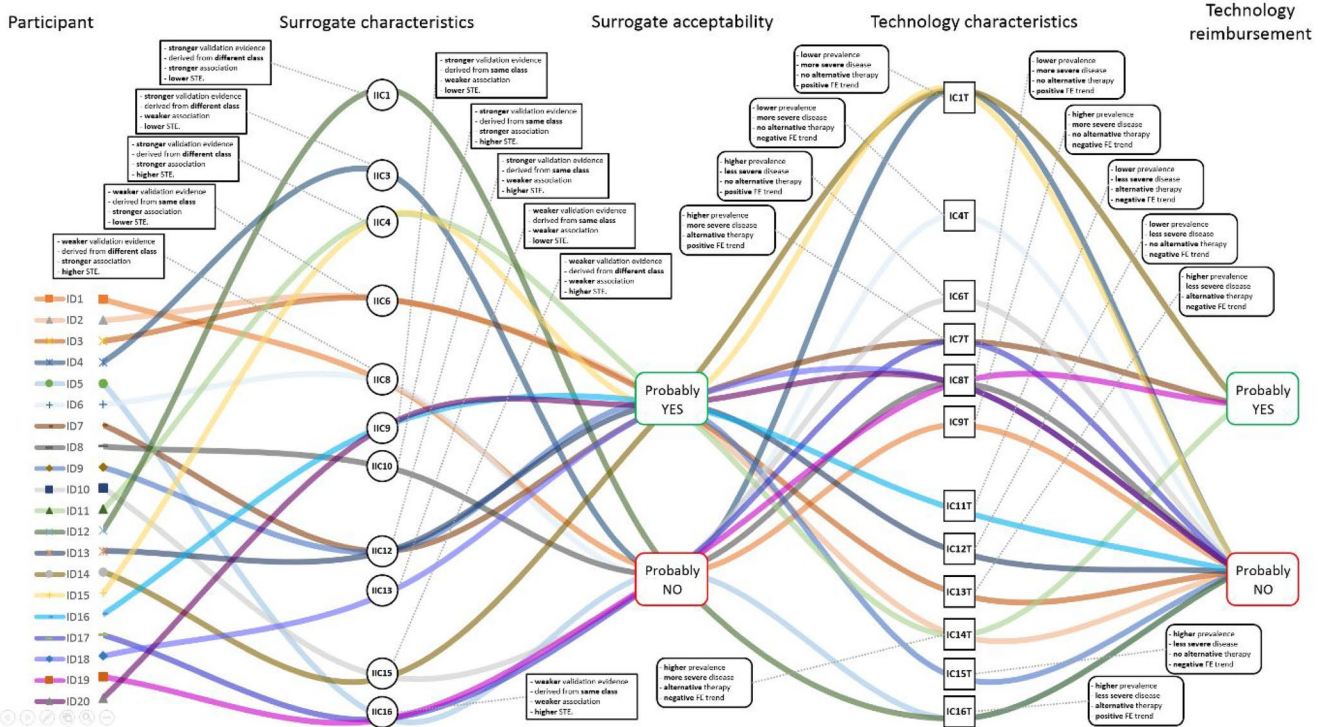
	Valid surrogate – YES	Valid surrogate – NO		Valid surrogate – YES	Valid surrogate – NO
<b>Scenario 1</b>			<b>Scenario 2</b>		
Full coverage – YES	3 (15%)	3 (15%)	Full coverage – YES	3 (15%)	1 (5%)
Full coverage – NO	10 (50%)	4 (20%)	Full coverage – NO	8 (40%)	8 (40%)

Note: Values represent number of responses, percentages are in brackets (out of a total of 20).





**FIGURE 3** Choices in Scenario 1: pharmaceutical evaluation (Section C: experimental scenario I). IC1-IC16 indicate variants displayed for surrogate evidence (contents of these variants presented in the connected rectangle boxes); IC1T-IC16 T indicate variants displayed for technology evaluation (contents of these variants presented in the connected rounded corner boxes); STE – surrogate threshold effect; FE – final endpoint



**FIGURE 4** Choices in Scenario 2: medical device evaluation (Section C: experimental scenario II). IIC1-IIC16 indicate variants displayed for surrogate evidence (contents of these variants presented in the connected rectangle boxes); IIC1T-IIC16 T indicate variants displayed for technology evaluation (contents of these variants presented in the connected rounded corner boxes); STE – surrogate threshold effect; FE – final endpoint

age decision. Whilst this ‘disconnect’ may simply reflect the variation in depth of methodological approach applied by agencies, it may also indicate that consideration of the treatment effect (based on a surrogate endpoint in this case) is only one factor that contributes toward a coverage decision in the appraisal of a health technology (National Institute for Health and Care Excellence (NICE), 2013; Rawlins & Culyer, 2004). For example, if an indication is rare, has a severe disease impact, or has a high unmet treatment burden, HTA agencies and payers may be willing to trade off their uncertainty in treatment effect and make a positive technology funding decision. Given that regulators have traditionally supported surrogate endpoints as part of their accelerated programmes (Gyawali et al., 2019; Salcher-Konrad et al., 2020), such as for orphan conditions or innovative treatments with high unmet treatment need, this is perhaps one specific area of application of surrogate endpoints where regulators and payers can develop a consensus and compile a joint list of approved (validated) surrogate endpoints (US Food Drug Administration, 2021).

## 4.1 | Development of a decision tool for HTA agencies/payers

Our research underscores the urgent need for technical support for the HTA and payer community in their assessment of the clinical and cost-effectiveness of health technologies when based on surrogate endpoint evidence. Such technical support includes: (1) the use of appropriate statistical validation methods, (2) clarity around the criteria for surrogate acceptance (or rejection), (3) transparency in the quantification of surrogate endpoint treatment effects in economic models (where relevant), and (4) incorporation of appropriate uncertainty into the decision-making process. We have identified a small number of HTA agencies, together with the European Network of HTA agencies, that have developed methods guidance for the surrogate endpoints that address the majority of these issues - UK NICE, German GBA and IQWiG, Australian PBAC and Canadian CADTH. However, despite the public availability of this detailed technical guidance, there remain potential barriers within individual HTA agency or payer settings to the systematic implementation of these approaches, such as technical capacity and difference in HTA processes. Initiatives such as the EU proposal of joint HTA clinical assessment (Kanavos et al., 2019) may provide the opportunity for implementation of a harmonized approach to the validation of the handling of surrogate endpoints across European agencies.

To facilitate such implementation and based on the three-step framework for surrogate endpoints introduced above, we have developed a web-based decision tool to support HTA agencies and payers when faced with the assessment of health technologies based on surrogate endpoint evidence. The tool was empirically developed by two of the authors (OC & RST) based on the findings of our research work undertaken in WP2 and the wider literature on the use of surrogate endpoints in HTA. We acknowledge that its development has not involved the wider community of researcher and policy makers. We therefore much welcome feedback on the utility of the decision tool. The tool is based on a process flowchart that provides the user with a step-by-step guide through the decision-making process (see Figure 5) and is available online (<https://www.sphsu.gla.ac.uk/comed/index.php>). It is important to emphasise that this tool does not replace the existing HTA assessment, appraisal, and policy making processes but rather provides a guide to support these existing processes in those situations where the clinical evidence base relies primarily on a surrogate endpoint.

The decision support tool directs the user through five questions (diamond shaped boxes), the responses to which determine the path through the algorithm arriving at a final recommended decision (blue circles).

### 4.1.1 | Is the primary endpoint a surrogate?

A surrogate endpoint definition widely applied in the regulatory setting is ‘a laboratory measurement or physical sign used in therapeutic trials as a substitute for a clinically meaningful endpoint, that is a direct measure of how a patient feels, functions, or survives, and is expected to predict the effect of the therapy’ (De Gruttola et al., 2001). This definition is usually limited to biomarkers (e.g., blood pressure, bone density, tumor size), clinical measures that can be objectively quantified but may not be perceived by patients. However, within the HTA context, a broader surrogate endpoint definition is needed that also includes the concept of an intermediate outcome, that is, an outcome of value to the patient that is thought to capture the causal pathway through which the disease process affects the final patient-relevant outcomes (e.g., for a therapy for heart failure, exercise capacity may serve as valid surrogate (intermediate) endpoint for cardiovascular mortality (Ciani et al., 2018)). HTA agencies and payers may also seek to broaden the definition of final patient-relevant outcomes to include not only clinical events (such as disease-related hospitalization or death) but also health-related quality of life (NICE, 2013). A key element of the HTA process is for stakeholders to clarify at the outset of a technology appraisal what might be an acceptable surrogate end point and final patient-patient outcome, and how they are measured (Drummond et al., 2008).

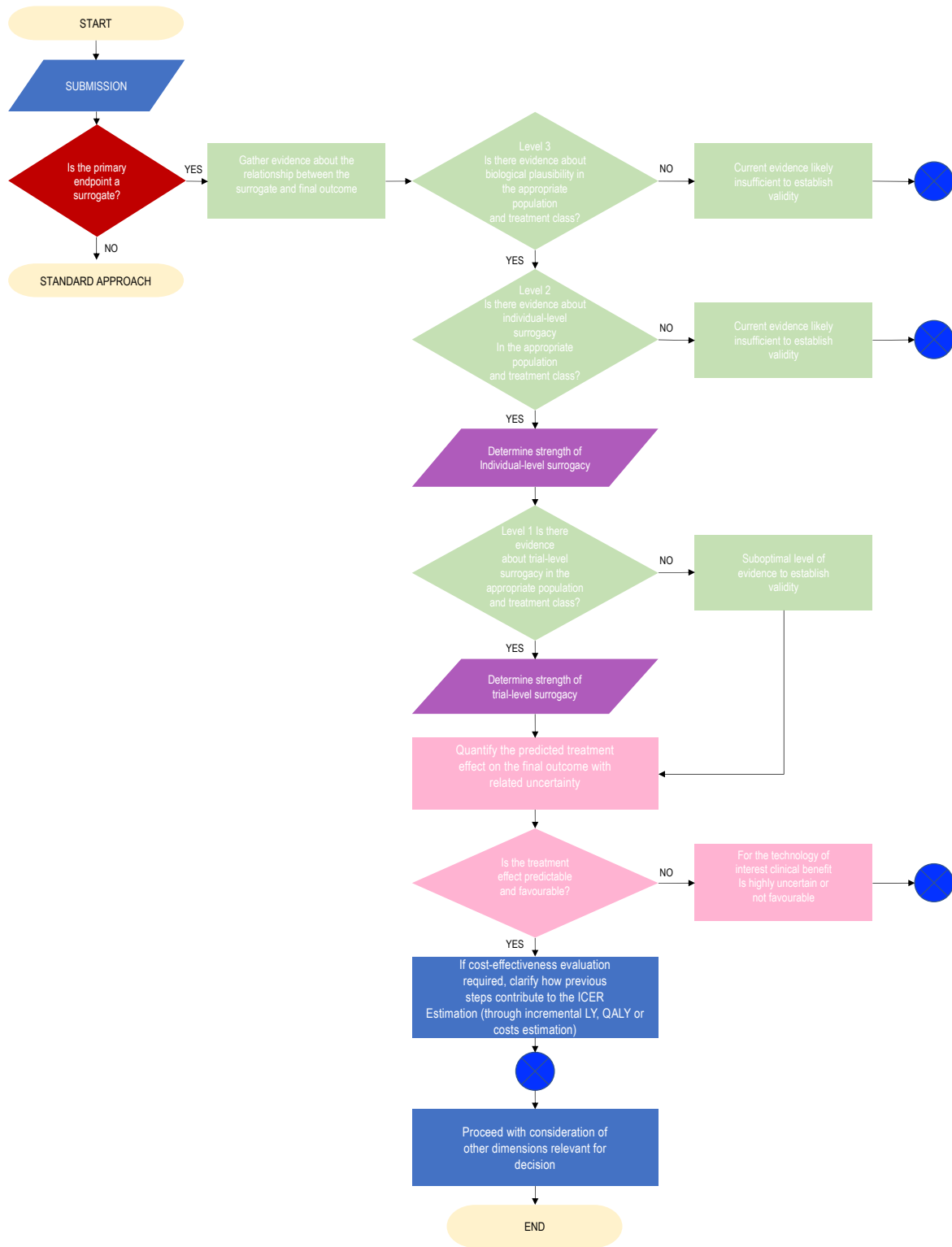


FIGURE 5 Surrogate outcome decision support tool. LY, life years; QALY, quality-adjusted life years

If the answer to this question (*Is the primary endpoint a surrogate?*) is ‘no’ (e.g., trial-based evidence is available with mature overall survival data), the decision tool will direct the user to a traditional approach for the assessment of clinical and cost-effectiveness assessment. However, if the evidence is primarily based on a surrogate endpoint, the tool will progress the user to a set of operations (rectangles) that deal with gathering the evidence about the relationship between the surrogate endpoint and final patient-relevant outcomes.

#### 4.1.2 | Gather the evidence about the surrogate-to-final outcome relationship

Establishing the validity of a surrogate endpoint requires a comprehensive review of the evidence for relationship between the surrogate and final patient-relevant outcome of interest and be based on a systematic review approach (Ciani et al., 2016, 2017). This review should include the following types of evidence: (1) randomised controlled trials, where both the surrogate and final outcome have been measured, (2) observational (or large single-arm interventional) studies, where both the surrogate and final outcome have been measured, and (3) mechanistic clinical studies designed to understand a biological process, the pathophysiology of a disease, or the mechanism of action of an intervention. This evidence should be interpreted using the Bradford Hill criteria for epidemiological causality (Hill, 1965) including whether the surrogate-final outcome relationship is consistent (i.e., do different studies show a consistent relationship?), strong (e.g., is there a high correlation coefficient?), plausible (i.e., does the relationship fit with existing biological/clinical thinking), and specific (i.e., is the relationship demonstrated in the relevant specific disease population/indication and class of treatment). Based on the different sources of evidence that might be available, the decision tool directs the user through the following three specific questions.

#### 4.1.3 | Is there evidence of biological plausibility?

According to the hierarchy of evidence, whilst necessary and important, biological plausibility of the relationship between the surrogate and final outcome (based on mechanistic studies and understanding of the pathophysiology of the disease and mechanism of action of the intervention) is necessary but not sufficient and corresponds to 'Level 3 evidence'. Recent developments of in silico disease modeling may contribute at this level (Musuamba et al., 2021). Such evidence alone is insufficient to establish the validity of the putative surrogate endpoint and the decision tool therefore directs the user to a question on the next level of evidence.

#### 4.1.4 | Is there observational evidence of an association between surrogate and final outcome?

This 'individual (or patient) level' evidence is typically sourced from observational design (e.g., cohort studies) or analysis of interventional studies the surrogate and final outcome association is assessed irrespective of treatment. For a surrogate to be considered valid, there should be a strong association quantified by statistical approaches such as a correlation coefficient (Pearson's, Spearman's, Kendall's) or coefficients of determination ( $R^2$ ). Prognostic model research or alternative measures derived from information theory could be used. In general, the stronger the correlation, the more likely the causal link between the surrogate and final outcomes, provided adjustment for confounders has been performed (Bucher et al., 1999). Whilst necessary, level 2 evidence is not generally considered sufficient for acceptance of a surrogate endpoint (Ciani et al., 2016, 2017). However, our COMED research has shown this level of evidence can be deemed sufficient by HTA agencies in particular circumstances including rare disease and conditions with high unmet treatment need, where flexibility is expected in the decision-making process. If Level 2 evidence is available, the decision tool directs the user to assess the strength of the association between the surrogate and final patient-relevant endpoint (individual-level surrogacy; see case study 2).

#### 4.1.5 | Is there evidence from randomised controlled trials showing that treatment changes in the surrogate and final outcome are associated?

For unequivocal surrogate endpoint acceptance, level 1 evidence is required, that is, a number of randomised controlled trials showing a strong association between the treatment change in surrogate and outcome undertaken in the appropriate patient population and therapy class. The more the RCTs the better, however the number of trials needed is hard to quantify as it may depend on the surrogacy pattern. This surrogate-final outcome treatment association is expressed at the trial-level typically based on a meta-analysis of a number of randomised trials and expressed as coefficient of determination from a linear regression of treatment effects on endpoints (see case study 1; Buyse et al., 2010), although linear regression is known to underestimate the uncertainty around the parameters describing the surrogate relationship (Bujkiewicz et al., 2017; Bujkiewicz, Achana, et al., 2019; Welton et al., 2020). Bivariate meta-analytic methods that involve appropriate inclusion of all relevant uncertainty are preferred (Daniels & Hughes, 1997).

There has been debate around what might constitute an ‘acceptable’ level of surrogate-final outcome association, with a correlation of  $\geq 0.6$  is often cited as ‘acceptable’ (Ciani et al., 2014). However, advocating a predefined threshold for the strength of association may not be the best approach needed. Instead, it may be better to reflect the quality of the surrogate-final outcome association in the uncertainty around the predicted treatment effect on the final outcome (Ciani et al., 2016, 2017).

Meta-analysis of individual participant data (IPD) from single or multiple randomised controlled trials provides an optimal approach to surrogate validation as it enables the standardization of statistical methods across trial data sets and analysis of the association at both the individual and trial levels (Xie et al., 2019). However, meta-analyses based on aggregate (trial level) data remains the most reported approach to surrogate validation (Ciani et al., 2014). Reporting guidelines of surrogate endpoint evaluation using meta-analyses have been developed to promote greater methodological consistency and facilitate the interpretation and reproducibility of meta-analytic surrogacy evaluation (Xie et al., 2019). More computationally intensive statistical methods of validation have been proposed including Bayesian multivariate meta-analytic methods for modeling surrogate relationships in each treatment pairwise comparison whilst borrowing information from other treatment contrasts, or for borrowing information about surrogate relationships across treatment classes within a hierarchical model (Bujkiewicz, Jackson, et al., 2019; Papanikos et al., 2020). Exploiting full or partial exchangeability of evidence across indications or treatment classes may be needed in those situations where flexibility is warranted by important contextual considerations.

#### 4.1.6 | Is the treatment effect predictable and favorable?

Following assessment of the validity, the next step is the prediction and quantification of the treatment effect on the final outcome with related uncertainty (see Figure 1), given the treatment effect observed effect on the surrogate in the pivotal trial of the technology under evaluation.

For level 1 evidence, predictions can be derived directly from linear regression, or preferably bivariate meta-analytic methods based on intercept, slope, and conditional variance of the linear model of the relationship between the treatment effects on the surrogate and the treatment effects on the final outcome (Bujkiewicz, Achana, et al., 2019). This prediction makes use of all data from the previous relevant studies reporting the treatment effects on both outcomes with corresponding standard errors, together with the data on the treatment effect on the surrogate endpoint and the corresponding standard error for the new study. The unreported treatment effect on the final outcome in the new study, and the corresponding standard error, are considered as missing and predicted from the model by the Markov chain Monte Carlo simulation, by assuming that the two treatment effects and their corresponding population level variances follow the same model as the data from all other studies reporting the treatment effects on both outcomes.

The 95% confidence (or credible) interval (95% CI) provides an estimate of uncertainty of the predicted treatment effect on the final outcome. Surrogate threshold effect (STE) has been defined as the minimum treatment effect on the surrogate necessary to predict an effect on the final patient-relevant outcome (Burzykowski & Buyse, 2006). This metric could support the quantification of the predicted treatment effect on the final outcome, since treatments that are able to induce effects larger than the STE on the surrogate would be expected to also induce a proportionally greater effect on the final outcome.

When only level 2 evidence is available, the treatment effect on the final-reported outcome needs to be estimated indirectly from the observed treatment effect on the surrogate. Proceeding with this step even in the absence of Level 1 evidence may be appropriate in the context of rare diseases or in general interventions for which randomized evidence is rarely available. The approaches vary depending on the type of surrogate endpoint and available observational data. For example, overall survival curves estimated separately for responders and non-responders (with or without a landmark time) may be combined with the observed proportions of patients in the trial who did and did not achieve a response after treatment (Ciani, Hoyle, Pavey, et al., 2013). Consequently, a treatment effect on the final patient-relevant outcome can be estimated. Given the indirect nature of these estimates, additional decision uncertainty should be factored into to the acceptance of clinical effectiveness of the technology in question.

In addition to assessment of clinical effectiveness, many HTA agencies/payers will also require a formal assessment of cost-effectiveness to make a coverage decision for a health technology. When a cost-effectiveness evaluation is required, a common challenge is that the outcome used in the economic evaluation is not the primary endpoint used in the clinical trial. In this respect, a decision support tool such as the one discussed in this study may be of considerable help, given that cost-effectiveness models cannot ignore the clinical effectiveness of the alternative options under evaluation. Ensuring that the issue of the validity of the surrogate endpoint is accounted for in a cost-effectiveness model implies transparency on how the surrogate treatment effect contributed to a model-based assessment of the incremental cost-effectiveness, whether through estimation of life years gained (NICE, 2016), quality-adjusted life years gained (e.g., determining different levels of utility value to use for different health states; SMC, 2015), or incremental costs (Hawkins et al., 2012). Avail-

able studies show consideration of validity of the surrogate endpoint is usually disregarded at this stage, and modeling approaches often rely on extrapolation of immature overall survival data or, when overall survival data are not available, on assumptions based on anecdotal evidence, evidence from different treatment settings, or expert opinion (Beauchemin et al., 2016; Ciani et al., 2021).

The practical implementation of the decision support tool developed is illustrated using two HTA case studies available as an online appendix, one on a medical device technology and the other on oncology drugs: (1) renal denervation for resistant hypertension where level 1 evidence was available (Boer et al., 2021) and (2) dasatinib, nilotinib, and imatinib for newly diagnosed Philadelphia chromosome positive chronic myeloid leukemia where only level 2 evidence was available (Ciani, Hoyle, Cooper, et al., 2013; Pavey et al., 2012).

## 5 | CONCLUSION

This research program sought to analyze contemporary international HTA practice and views and preferences of stakeholders with the overarching objective of improving the decision-making process for the use of surrogate endpoints in the assessment of the clinical effectiveness and cost-effectiveness of new or existing healthcare technologies. Based on our research findings, we adapted the three-step evaluation framework as a web-based decision support tool. This decision tool is intended for use by HTA agencies and their decision-making committees together with the wider community of HTA stakeholders (including clinicians, patient groups, and healthcare manufacturers). Having developed this tool, we plan to monitor its use and we welcome feedback on its utility.

### ACKNOWLEDGEMENTS

The authors wish to thank the various participants and researchers who took part in the COMED empirical studies and Hitarth Narvala (MRC/CSO Social and Public Health Sciences Unit & Robertson Center for Biostatistics, Institute of Health and Well Being, University of Glasgow) for assistance in developing the web-based framework. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement #779306. This result only reflects the author's views, and the EU is not responsible for any use that may be made of the information it contains.

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Zenodo at <https://zenodo.org/>, reference numbers 4293370, 4665387 and 4560288. The Surrogate Outcome Decision Support Tool is openly accessible at <https://www.sphsu.gla.ac.uk/comed/index.php>.

### ORCID

Rod S. Taylor  <https://orcid.org/0000-0002-3043-6011>

### REFERENCES

- Beauchemin, C., Lapierre, M. È., Letarte, N., Yelle, L., & Lachaine, J. (2016). Use of intermediate endpoints in the economic evaluation of new treatments for advanced cancer and methods adopted when suitable overall survival data are not available. *Pharmacoeconomics*, 34(9), 889–890. <https://doi.org/10.1007/s40273-016-0401-4>
- Boer, J., Schmieder, R. E., Lobo, M., Kirtane, A. J., Taylor, R. S., Clarke, C., Murphy, K., van Keep, M., Thu, T. A., Barman, N. C., Schwab, G., & Ron Akehurst, R. (2021). Cost-effectiveness of treating resistant hypertension with ultrasound renal denervation: A UK perspective. *Journal of the American College of Cardiology*, 2021 (in preparation).
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bucher, H. C., Guyatt, G. H., Cook, D. J., Holbrook, A., & McAlister, F. A. (1999). Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-based medicine working group. *JAMA*, 282(8), 771–778. <https://doi.org/10.1001/jama.282.8.771>
- Bujkiewicz, S., Achana, F. A., Papanikos, T., Riley, R. D., & Abrams, K. R. (2019). *NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints*. <http://nicedsu.org.uk/wp-content/uploads/2020/10/TSD-20-mvmeta-final.pdf>

- Bujkiewicz, S., Jackson, D., Thompson, J. R., Turner, R. M., Stadler, N., Abrams, K. R., & White, I. R. (2019). Bivariate network meta-analysis for surrogate endpoint evaluation. *Stats Med*, 38(18), 3322–3341. <https://doi.org/10.1002/sim.8187>
- Bujkiewicz, S., Thompson, J. R., Spata, E., & Abrams, K. R. (2017). Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints. *Statistical Methods in Medical Research*, 26(5), 2287–2318. <https://doi.org/10.1177/0962280215597260>
- Burzykowski, T., & Buyse, M. (2006). Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics*, 5(3), 173–186. <https://doi.org/10.1002/pst.207>
- Buyse, M., Sargent, D. J., Grothey, A., Matheson, A., & de Gramont, A. (2010). Biomarkers and surrogate end points—the challenge of statistical validation. *Nature Reviews Clinical Oncology*, 7(6), 309–317. <https://doi.org/10.1038/nrclinonc.2010.43>
- Ciani, O., Buyse, M., Drummond, M., Rasi, G., Saad, E. D., & Taylor, R. S. (2016). Use of surrogate end points in healthcare policy: A proposal for adoption of a validation framework. *Nature Reviews Drug Discovery*, 15(7), 516. <https://doi.org/10.1038/nrd.2016.81>
- 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., Buyse, M., Garside, R., Pavey, T., Stein, K., Sterne, J. A. C., & Taylor, R. S. (2013). Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ*, 346, f457. <https://doi.org/10.1136/bmj.f457>
- Ciani, O., Davis, S., Tappenden, P., Garside, R., Stein, K., Cantrell, A., Saad, E. D., Buyse, M., & Taylor, R. S. (2014). Validation of surrogate endpoints in advanced solid tumors: Systematic review of statistical methods, results, and implications for policy makers. *International Journal of Technology Assessment in Health Care*, 30(3), 312–324. <https://doi.org/10.1017/s0266462314000300>
- Ciani, O., Grigore, B., Blommestein, H., de Groot, S., Möllenkamp, 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(4), 439–452. <https://doi.org/10.1177/0272989x21994553>
- Ciani, O., Hoyle, M., Pavey, T., Cooper, C., Garside, R., Rodin, C., & Taylor, R. S. (2013). Complete cytogenetic response and major molecular response as surrogate outcomes for overall survival in first-line treatment of chronic myelogenous leukemia: A case study for technology appraisal on the basis of surrogate outcomes evidence. *Value in Health*, 16(6), 1081–1090. <https://doi.org/10.1016/j.jval.2013.07.004>
- Ciani, O., Piepoli, M., Smart, N., Uddin, J., Walker, S., Warren, F. C., Zwisler, A. D., Davos, C. H., & Taylor, R. S. (2018). Validation of exercise capacity as a surrogate endpoint in exercise-based rehabilitation for heart failure: A meta-analysis of randomized controlled trials. *JACC Heart Failure*, 6(7), 596–604. <https://doi.org/10.1016/j.jchf.2018.03.017>
- COMED. (2021). Work Package 2. <https://www.comedh2020.eu/wps/wcm/connect/site/comed/home/project/work+packages/wp2>
- Daniels, M. J., & Hughes, M. D. (1997). Meta-analysis for the evaluation of potential surrogate markers. *Stats Medicine*, 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)
- De Gruttola, V. G., Clax, P., DeMets, D. L., Downing, G. J., Ellenberg, S. S., Friedman, L., Gail, M. H., Prentice, R., Wittes, J., & Zeger, S. L. (2001). Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. *Controlled Clinical Trials*, 22(5), 485–502. [https://doi.org/10.1016/s0197-2456\(01\)00153-2](https://doi.org/10.1016/s0197-2456(01)00153-2)
- DeMets, D. L., Psaty, B. M., & Fleming, T. R. (2020). When can intermediate outcomes be used as surrogate outcomes? *JAMA*, 323(12), 1184–1185. <https://doi.org/10.1001/jama.2020.1176>
- Drummond, M. F., Schwartz, J. S., Jönsson, B., Luce, B. R., Neumann, P. J., Siebert, U., & Sullivan, S. D. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International Journal of Technology Assessment in Health Care*, 24(03), 244–258. <https://doi.org/10.1017/s0266462308080343>
- EUnetHTA. (2015). *Endpoints used for relative effectiveness assessment clinical endpoints amended JA1 guideline final Nov 2015*. <https://www.eunetha.eu/endpoints-used-for-relative-effectiveness-assessment-clinical-endpoints-amended-ja1-guideline-final-nov-2015/>
- 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., Zempenyi, A., & Taylor, R. S. (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>
- Gyawali, B., Hey, S. P., & Kesselheim, A. S. (2019). Assessment of the clinical benefit of cancer drugs receiving accelerated approval. *JAMA Internal Medicine*, 179(7), 906–913. <https://doi.org/10.1001/jamainternmed.2019.0462>
- Hawkins, N., Richardson, G., Sutton, A. J., Cooper, N. J., Griffiths, C., Rogers, A., & Bower, P. (2012). Surrogates, meta-analysis and cost-effectiveness modelling: A combined analytic approach. *Health Economics*, 21(6), 742–756. <https://doi.org/10.1002/hec.1741>
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295–300. <https://doi.org/10.1177/003591576505800503>
- IQWiG. (2011). *Validity of surrogate endpoints in oncology. Executive Summary*. [https://www.iqwig.de/download/a10-05\\_executive\\_summary\\_v1-1\\_surrogate\\_endpoints\\_in\\_oncology.pdf](https://www.iqwig.de/download/a10-05_executive_summary_v1-1_surrogate_endpoints_in_oncology.pdf)
- Kanavos, P., Angelis, A., & Drummond, M. (2019). An EU-wide approach to HTA: An irrelevant development or an opportunity not to be missed? *The European Journal of Health Economics*, 20(3), 329–332. <https://doi.org/10.1007/s10198-019-01037-2>
- Lassere, M. N., Johnson, K. R., Schiff, M., & Rees, D. (2012). Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (BioSurrogate) evaluation Schema (BSES). *BMC Medical Research Methodology*, 12(1), 27. <https://doi.org/10.1186/1471-2288-12-27>

- Musuamba, F. T., Skottheim Rusten, I., Lesage, R., Russo, G., Bursi, R., Emili, L., Wangorsch, G., Manolis, E., Karlsson, K. E., Kulesza, A., Courcelles, E., Boissel, J., Rousseau, C. F., Voisin, E. M., Alessandrello, R., Curado, N., Dall'ara, E., Rodriguez, B., Pappalardo, F., & Geris, L. (2021). Scientific and regulatory evaluation of mechanistic in silico drug and disease models in drug development: Building model credibility. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(8), 804–825. <https://doi.org/10.1002/psp4.12669>
- National Institute for Health and Care Excellence (NICE). (2013). *Guide to the methods of technology appraisal*. <https://www.nice.org.uk/process/pmg9/chapter/foreword>
- NICE. (2016). *Degarelix for treating advanced hormone-dependent prostate cancer* Technology appraisal guidance [TA404]. Retrieved August 24, 2016, from <https://www.nice.org.uk/guidance/ta404>
- Ortiz, Y., Fareli, C. J., Gallegos, V., & Hernández, E. (2021). Surrogate endpoints in oncology: Overview of systematic reviews and their use for health decision making in Mexico. *Value in Health Regional Issues*, *26*, 75–88. <https://doi.org/10.1016/j.vhri.2021.04.002>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*, *42*(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Papanikos, T., Thompson, J. R., Abrams, K. R., Stadler, N., Ciani, O., Taylor, R., & Bujkiewicz, S. (2020). Bayesian hierarchical meta-analytic methods for modeling surrogate relationships that vary across treatment classes using aggregate data. *Stats Medicine*, *39*(8), 1103–1124. <https://doi.org/10.1002/sim.8465>
- Pavey, T., Hoyle, M., Ciani, O., Crathorne, L., Jones-Hughes, T., Cooper, C., Osipenko, L., Venkatachalam, M., Rudin, C., Ukoumunne, O., Garside, R., & Anderson, R. (2012). Dasatinib, nilotinib and standard-dose imatinib for the first-line treatment of chronic myeloid leukaemia: Systematic reviews and economic analyses. *Health Technology Assessment*, *16*(42), iii–iv, 1–277.
- Rawlins, M. D., & Culyer, A. J. (2004). National Institute for clinical excellence and its value judgments. *BMJ*, *329*(7459), 224–227. <https://doi.org/10.1136/bmj.329.7459.224>
- Robb, M. A., McInnes, P. M., & Califf, R. M. (2016). Biomarkers and surrogate endpoints: Developing common terminology and definitions. *JAMA*, *315*(11), 1107–1108. <https://doi.org/10.1001/jama.2016.2240>
- Salcher-Konrad, M., Naci, H., & Davis, C. (2020). Approval of cancer drugs with uncertain therapeutic value: A comparison of regulatory decisions in Europe and the United States. *The Milbank Quarterly*, *98*(4), 1219–1256. <https://doi.org/10.1111/1468-0009.12476>
- Schuster Bruce, C., Brhlikova, P., Heath, J., & McGettigan, P. (2019). The use of validated and nonvalidated surrogate endpoints in two European medicines agency expedited approval pathways: A cross-sectional study of products authorised 2011–2018. *PLoS Medicine*, *16*(9), e1002873. <https://doi.org/10.1371/journal.pmed.1002873>
- SMC. (2015). *Ledipasvir/sofosbuvir, 90mg/400mg, film-coated tablet (Harvoni)*. [https://www.scottishmedicines.org.uk/media/1905/ledipasvir\\_sofosbuvir\\_harvoni\\_final\\_february\\_2015\\_for\\_website.pdf](https://www.scottishmedicines.org.uk/media/1905/ledipasvir_sofosbuvir_harvoni_final_february_2015_for_website.pdf)
- US Food Drug Administration. (2021). Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure, Retrieved March 31, 2021, from <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure#:~:text=FDA%E2%80%99s%20surrogate%20endpoint%20table%20provides%20valuable%20information%20for,and%20discussed%20with%20FDA%20for%20individual%20development%20programs>
- Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, *18*(1), 148. <https://doi.org/10.1186/s12874-018-0594-7>
- Velasco Garrido, M., & Mangiapane, S. (2009). Surrogate outcomes in health technology assessment: An international comparison. *International Journal of Technology Assessment in Health Care*, *25*(03), 315–322. <https://doi.org/10.1017/s0266462309990213>
- Walter, S. D., Sun, X., Heels-Ansdell, D., & Guyatt, G. (2012). Treatment effects on patient-important outcomes can be small, even with large effects on surrogate markers. *Journal of Clinical Epidemiology*, *65*(9), 940–945. <https://doi.org/10.1016/j.jclinepi.2012.02.012>
- Welton, N. J., Phillippo, D. M., Owen, R., Jones, H. E., Dias, S., Bujkiewicz, S., Ades, A. E., & Abrams, K. R. (2020). *CHTE2020 sources and synthesis of evidence; update to evidence synthesis methods*. DSU report. [http://nicedsu.org.uk/wp-content/uploads/2020/11/CHTE-2020\\_final\\_20April2020\\_final.pdf](http://nicedsu.org.uk/wp-content/uploads/2020/11/CHTE-2020_final_20April2020_final.pdf)
- Xie, W., Halabi, S., Tierney, J. F., Sydes, M. R., Collette, L., Dignam, J. J., Buyse, M., Sweeney, C. J., & Regan, M. M. (2019). A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. *JNCI Cancer Spectrum*, *3*(1), pkz002. <https://doi.org/10.1093/jncics/pkz002>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** 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*, 1–29. <https://doi.org/10.1002/hec.4524>



## APPENDIX 1: CHOICE EXPERIMENT QUESTIONNAIRE

### Investigating attitudes to surrogate evidence in health technology assessment: a binary choice experiment pilot

#### Informed consent

The purpose of the following text is to supply you with the information you need in order for you to provide your informed consent for your participation in this research project.

#### Statement of the research being undertaken

The study is part of an European Union funded project called COMED (Pushing the boundaries of Cost and Outcome analysis of MEDical Technologies) that seeks to develop scientifically rigorous recommendations on methods for the assessment of health technologies. Prof Aleksandra Torbica at the Center for Research on Health and Social Care Management (CERGAS) is the principal investigator. This specific study is led by Dr Oriana Ciani ([oriana.ciani@unibocconi.it](mailto:oriana.ciani@unibocconi.it)) from Bocconi University, Prof Rod Taylor ([Rod.Taylor@glasgow.ac.uk](mailto:Rod.Taylor@glasgow.ac.uk)) from Glasgow University and Dr Bogdan Grigore ([b.grigore@exeter.ac.uk](mailto:b.grigore@exeter.ac.uk)) from Exeter University. As former student on the MIHMEP course between 2016 and 2019, SDA Bocconi has given us permission to directly contact you to seek your participation in this research exercise on surrogate endpoints.

#### Procedures and duration

The exercise involves you reading two hypothetical scenarios each of which requires you to give your view on the strength of evidence for the surrogate endpoint and whether you would fund a technology based on this evidence. Before completing these two scenarios, we will ask you some brief demographic questions about yourself, attitudinal questions about the use of health care evidence, and then finally, at the end of this exercise, we would like your views on how you found undertaking this exercise. We anticipate that the completion of the scenarios and questions will take you about 15 min in total.

#### Expected benefits and foreseeable risks

We do not anticipate any further burden to you, beside the time commitment. Furthermore, we expect that participation in this study will re-acquaint you with the current concerns about the use of surrogate endpoints in health technology assessment.

#### Voluntary participation

Participation in this study is voluntary. You will be able to withdraw from the study at any point, and your withdrawal involves no penalty or loss of benefits you would otherwise receive. You do not have to answer questions you don't want to answer. If you have given your consent and you wish to withdraw it, please contact Dr Oriana Ciani ([oriana.ciani@unibocconi.it](mailto:oriana.ciani@unibocconi.it)) or Dr Bogdan Grigore ([b.grigore@exeter.ac.uk](mailto:b.grigore@exeter.ac.uk)). Please note that where our processing of your personal data relies on your consent and where you then withdraw that consent, its withdrawal shall not cause any effect in the lawfulness of the previously processed data.

#### Compensation

There will be no material compensation for participating in the study.

#### Deception

This study does not involve deception.

#### Research participant declaration

Please click the box to confirm you agree to continue this survey.

I confirm that I read the patient documentation, and I declare having read and understood its content. I confirm that I am 18 years of age or older, and volunteer to take part in this research. Taking note that my Data are processed in full compliance with the Law, I freely consent to my Data to be used in the manner and uses described. I also declare having understood my rights and limitations, as well as how to exercise them.

**Thank you for agreeing to participate – please read the text below that provides some background on surrogate endpoints in health care and then complete sections A, B, C and D.**

### What are surrogate endpoints and what are the issues with their use?

For the purpose of this study, we define a *surrogate endpoint* as a biomarker, or an intermediate outcome, that can substitute for a final patient-relevant outcome such as morbidity, mortality and health-related quality of life (De Gruttola et al., 2001). For example, *bone mineral density* has been considered as a surrogate for *bone fracture*.

Surrogate endpoints enable faster and smaller clinical trials that in turn lead to faster healthcare decisions and faster access to treatment for patients. However, surrogate endpoint evidence can sometimes be unreliable, leading to undesirable consequences, such as more harm than benefit to patients or non-cost effective use of healthcare resources.

For this reason, it is important to explore the robustness of surrogate endpoint information through a process generally referred to as *validation*. Our recent work reviewing health technology assessments that relied on surrogate endpoint evidence suggests both a lack of consensus on what constitute “acceptable” surrogate measures, and a lack of consistency in weighting the uncertainty induced by such surrogate measures.

In this study, we would like to explore *how surrogate endpoints are valued in the decision-making process* by proposing a choice scenario.

#### A: General questions

First, we would like to know a few things about you.

These questions are not intended to identify you, however they will help assess the adequacy of the survey.

- A1. What is your gender? (male/female/prefer not to answer)
- A2. What is your age group? (18–24/25–34/35–44/45–54/55–64/65–74/75 or older)
- A3. What is your main background?
- Economics
  - Engineering
  - Humanities/Law
  - Medicine
  - Nursing/healthcare profession
  - Pharmacy/Biomedical sciences
  - Other – please specify.
- A4. What is your current occupation?
- Academia;
  - Public agency/competent authority/government;
  - Industry;
  - Healthcare organization;
  - Consulting firm;
  - other – please specify.

#### B: Your views on the role of surrogate endpoints in health care decision making

Before completing the two choice scenarios below, we would like to explore your views on using surrogate endpoint evidence in health technology assessment.

Please indicate whether you agree with the following statements about the use of surrogate endpoints in the evaluation of health technologies.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
In HTA, it is of paramount importance to only consider surrogates that have been previously validated					
The earlier a technology is in its development, the more uncertainty should be allowed in the surrogate measure					
With an innovative technology, it is always acceptable to rely on evidence for validating a surrogate endpoint derived from a previous class of therapies					
The quality of evidence for validating a surrogate endpoint can be overlooked when there are unmet needs					

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
-------------------	-------------------	----------------------------	----------------	----------------

Evidence on the surrogate endpoint is always complemented by evidence on the final point, however immature it may be

Medical devices should be evaluated using the same quality of evidence as pharmaceuticals

Evidence based on non-validated surrogate endpoints is acceptable for the evaluation of medical devices

### Section C: experimental scenario I

Please consider the following hypothetical scenario.

Imagine that you are a member of the board that evaluates new health technologies for reimbursement.

Currently, you are evaluating a new **pharmaceutical therapy**. This therapy is indicated for the treatment of neuromuscular symptoms of a **metabolic myopathy**. The condition is characterized by inefficient synthesis of a mitochondrial enzyme from its precursor. Severity of the condition varies, with many patients completely asymptomatic.

Symptomatic patients usually manifest either *neuromuscular symptoms* (generalised muscle weakness and fatigability or recurrent exercise-induced myalgia) or *cardiac symptoms* (heart failure due to left ventricular non-compaction cardiomyopathy, LVNC). Associated heart failure has been treated with conventional treatment (e.g., angiotensin-converting enzyme inhibitors, beta-blocking agents etc.) with some success; however the prognosis seems to be more dependent on the severity of the disease, quantified by the serum levels of the enzyme, *matrix metalloproteinase 9* (MMP-9).

The therapy was evaluated in a randomised placebo-controlled trial where MMP-9 was the primary outcome. The trial also collected quality of life data using a validated disease specific measure. The analysis of the clinical trial showed there was a statistically significant improvement in MMP-9 levels (mean change  $-0.80$  ng/ml) compared to the placebo group (mean change  $-0.08$  ng/ml) at 18 weeks ( $p < 0.0001$ ).

The evidence linking MMP-9 levels to HR-QoL was **based on** [a meta-analysis of several RCTs/a large observational study].

This data came from a **population of patients with** [the same neuromuscular symptoms/the cardiac symptoms] originated by metabolic myopathy.

The calculated **strength of association between** MMP-9 levels and HR-QoL was [0.30 (95% confidence interval [0.20, 0.40])/0.85 (95% confidence interval [0.77, 0.93])].

The calculated **surrogate threshold effect** (STE), defined as the minimum treatment effect on the surrogate necessary to predict a significant non-zero effect on the true endpoint, **was** [ $-0.10$  ng/ml observed in about 20% of the studies in the indication/ $-0.90$  ng/ml observed in about 70% of the studies in the indication].

**CIa. Based only on the information provided so far in this scenario, would you consider that this is an acceptable (i.e., valid) surrogate for predicting changes in patient health-related quality of life?** [Yes/No]

Now, consider other characteristics of the submission.

The **prevalence of this disease is** [one in 100,000/one in 1000].

**Mean utility score** (where a score of 0 is equivalent to death and score of 1.0 perfect health) of typical untreated patients was [0.30/0.60].

**Current best treatment** for this condition is [best supportive care (i.e., there is no specific therapy)/off-label treatment with a pharmaceutical indicated for heart failure].

**Secondary endpoint** in the trial (HR-QoL at 18 weeks) showed [improvement/deterioration], although with no statistical significance ( $p = 0.10$ ).

**CIb. Based solely on the data presented so far, would you support the full coverage of this therapy?** (Yes/No)

### Section C: experimental scenario II

Please consider a second hypothetical scenario, as member of the reimbursement board.

You have been tasked to evaluate a new **medical device** for the treatment of resistant hypertension through renal denervation. This is achieved through catheterization with a probe able to thermally ablate and disrupt the renal sympathetic nerves while sparing the renal arterial wall.

The medical device was evaluated in a single-blind, randomised, **sham-controlled trial** with 300 participants. Primary endpoint was mean reduction in average daytime ambulatory **systolic BP from baseline** to 2 months post procedure. Analysis

showed greater change with renal denervation ( $-8.5$  mm Hg, standard deviation = 9.3) than with the sham procedure ( $-2.2$  mm Hg, standard deviation = 10.0; baseline-adjusted difference between groups:  $-6.5$  mm Hg, 95% CI  $-9.4$  to  $-3.6$ ,  $p = 0.0001$ ).

The primary endpoint in the trial (change in **systolic blood pressure**) was a surrogate for major cardiovascular events.

The **evidence** linking change in systolic blood pressure to major cardiovascular events was based on [a meta-analysis of several RCT/a single RCT]. This **data came** from a population of patients under treatment with [antihypertensive medication/a non-pharmaceutical technology in the same indication].

The calculated **strength of association** between systolic blood pressure and major cardiovascular events was [0.85 (95% confidence interval [0.77, 0.93])/0.30 (95% confidence interval [0.20, 0.40])].

Calculated surrogate threshold effect (**STE**), defined as the minimal difference between intervention and control needed to ensure a stroke reduction benefit, **was** [ $-4$  mm Hg (observed in about 70% of the studies in the indication)/ $-10$  mm Hg (observed in about 20% of the studies in the indication)].

**CIa. Based only on the information provided so far in this scenario, would you consider that this endpoint is an acceptable (i.e., valid) surrogate for predicting changes in the risk of stroke?** [Yes/No]

Now, consider other characteristics of the submission.

The **prevalence** of this disease is [one in 11/one in 1500] hypertensive patients.

**Mean utility score** (where a score of 0 is equivalent to death and score of 1.0 perfect health) of typical untreated patients was [0.57/0.79].

Currently, there is [another/no] **treatment reimbursed** for resistant hypertension.

**Secondary endpoint** in the trial (cardiovascular events) **was** immature, appearing to favor [the intervention/the control].

**CIb. Based solely on the data presented so far, would you support the full coverage of this medical device?** (Yes/No)

#### Section D. Your views of this exercise

We would welcome your views on the practicalities of completing this exercise.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
The background information provided clear explanation about the purpose of the study.					
The two scenarios are plausible in real life appraisals					
The choice tasks were relatively easy to perform					

Please provide any specific comments and/or suggestions in the box below:

**Thank you for your participation.**

#### APPENDIX 2: ATTRIBUTES AND LEVELS USED IN THE BINARY CHOICE EXPERIMENT

Attribute	Level	Interpretation
<b>Scenario 1: new drug therapy for management of neuromuscular symptoms of a metabolic myopathy</b>		
<b>Question 1. Based only on the information provided so far in this scenario, would you consider that this is an acceptable (i.e. valid) surrogate for predicting changes in patient health-related quality of life?</b>		
Source of evidence for the validation of the surrogate endpoint	A meta-analysis of several RCTs	Stronger evidence
	A large observational study	Weaker evidence

Attribute	Level	Interpretation
Class of therapies providing evidence for the validation of the surrogate endpoint	The same neuromuscular symptoms originated by metabolic myopathy	The same class
	The cardiac symptoms originated by metabolic myopathy	A different class
Strength of association between the surrogate and patient-relevant endpoint	$R^2 = 0.30$ (95% confidence interval [0.20, 0.40])	Weaker association
	$R^2 = 0.85$ (95% confidence interval [0.77, 0.93])	Stronger association
Surrogate threshold effect (i.e. the minimum effect on the surrogate to predict a significant effect on the patient-relevant endpoint)	-0.10 ng/ml (observed in about 70% of the studies in the indication)	Lower STE
	-0.90 ng/ml (observed in about 20% of the studies in the indication)	Higher STE
<b>Scenario 1 Question 2. Based solely on the data presented so far, would you support the full coverage of this therapy?</b>		
Disease prevalence	One in 100,000	Lower prevalence
	One in 1000	Higher prevalence
Baseline utility score (on 0–1 scale)	0.30	More severe disease
	0.60	Less severe disease
Comparator (i.e. therapeutic alternatives)	Best supportive care (i.e. there is no alternative)	No alternative
	Off-label treatment with a pharmaceutical indicated for heart failure	Existing alternative therapy
Effect on the final outcome at 18 weeks based on immature data	Improvement, although not statistical significance ( $p = 0.10$ )	Positive trend
	Deterioration, although not statistical significance ( $p = 0.10$ )	Negative trend
<b>Scenario 2: new medical device for the treatment of resistant hypertension</b>		
<b>Question 1: Based only on the information provided so far in this scenario, would you consider that this endpoint is an acceptable (i.e. valid) surrogate for predicting changes in the risk of stroke?</b>		
Source of evidence for the validation of the surrogate endpoint	A meta-analysis of several RCT	Stronger evidence
	A single RCT	Weaker evidence
Class of therapies providing evidence for the validation of the surrogate endpoint	Antihypertensive medication	The same class
	A non-pharmaceutical technology in the same indication	A different class
Strength of association between the surrogate and patient-relevant endpoint	0.30 (95% confidence interval [0.20, 0.40])	Weaker association
	0.85 (95% confidence interval [0.77, 0.93])	Stronger association
Surrogate threshold effect	-4 mm Hg (observed in about 70% of the studies in the indication)	Lower STE
	-10 mm Hg (observed in about 20% of the studies in the indication)	Higher STE
<b>Scenario 2 Question 2: Based solely on the data presented so far, would you support the full coverage of this medical device?</b>		
Disease prevalence	One in 11 hypertensive patients	Higher prevalence
	One in 1500 hypertensive patients	Lower prevalence
Baseline utility score (on 0–1 scale)	0.57	Less severe disease
	0.79	More severe disease
Comparator (i.e. therapeutic alternatives)	No treatment reimbursed for resistant hypertension	No alternative
	Another treatment reimbursed for resistant hypertension	Existing alternative therapy
Effect on the incidence of cardiovascular events based on immature data	Appearing to favor the intervention	Positive trend
	Appearing to favor the control	Negative trend

Note:  $R^2$  = coefficient of determination, the proportion of the variance in the final endpoint that is predictable from the surrogate endpoint; RCT = randomised controlled trial; STE = surrogate threshold effect = the minimum treatment effect on the surrogate endpoint necessary to predict a non-zero effect on the final endpoint.

## APPENDIX 3: QUALITATIVE STUDY OF HTA STAKEHOLDERS' VIEWS ON THE USE OF SURROGATE ENDPOINTS

### INTERVIEW SCHEDULE

#### SECTION A: INTRODUCTORY REMARKS

Name of interviewee: \_\_\_\_\_

Position: \_\_\_\_\_

Organization: \_\_\_\_\_

NOTE: only summary data will be reported and individual respondents will not be identified

Introduction suggestions (interviewers may use their own words to explain the following points):

-“thank you for your time; I am [...]”;

-“The aim of this project is [...]. We anticipate that the interview will take no longer than one hour”.

-“Throughout the interview process, we will make written notes on your responses to the various questions. Please bear with us if there are short pauses between questions while we note your replies.”

- “We would like to audio record this interview as an additional quality assurance. We may quote some of your words in our report, but they will not be associated with your name. With your permission, may we audio-record the interview?”

#### SECTION B: OVERVIEW

Interviewer: We would like to begin, by learning more about your role and views.

**B1. Could you briefly describe your role in relation to the regulatory and HTA process in your country? (does this include medical devices?)**

**B2. Are you familiar with surrogate endpoints in the evaluation of health technologies? Have you participated in regulatory and HTA evaluations that relied on surrogate endpoints? If so, please give us details about this process (the process that you recall most clearly/that had the most debate around the surrogate endpoint).**

(Unless you prefer a different example) I propose we focus on the following example for our discussion:

- *Minimal residual disease* as relatively new surrogate endpoint for overall survival in chronic lymphocytic leukaemia – the validation evidence gathered so far is limited.

#### SECTION C: SURROGATE ENDPOINTS

**C1. A common definition of surrogate endpoints is “laboratory measurements or physical signs (e.g. blood pressure, cholesterol level) used as substitutes for a more clinically meaningful outcomes (a direct measure of how a patient feels, functions or survives, such as death, cardiovascular events etc.) when a new drug or health technology is assessed”. Would you agree with this definition? Is there anything not captured in this definition that you would include?**

**C2. On a scale from 0 (not relevant) to 5 (essential), how important is having a valid surrogate endpoint in the decision making process on health technologies when evidence on patient-relevant endpoints is not available?**

0                      1                      2                      3                      4                      5

Please provide a rationale for your score:

**C3. Ensuring the reliability of surrogate endpoints to predict final endpoints is often a challenge.**

In your opinion what makes a “valid” surrogate endpoint? What type of evidence (i.e. number, study design, quality) would be required for validation?

**C4. Are you aware of any methods proposed and used to validate a potential surrogate endpoint?**

**C5. The strength of the evidence supporting the relationship between surrogate endpoint and final outcome has been classified in three discrete levels as listed below.**

- Level 3 - Pathophysiological studies and understanding of the disease process and involvement of the surrogate endpoint (biological plausibility)
- Level 2 – Epidemiological /observational studies showing the association of the surrogate endpoint and the final outcome (observational evidence)
- Level 1 - Randomised controlled trials showing that changes in the surrogate endpoint are associated with commensurate changes in the final outcome (randomised trial evidence)

What level of evidence would you normally consider acceptable for decision-making on health technologies relying on surrogate endpoints?

Are there situations where a lower level of evidence might be acceptable (e.g. innovativeness of the technology, highly unmet needs, lack of therapeutic alternatives)?

**C6. Do you think medical devices pose specific challenges regarding the validation of surrogate endpoints (as compared to medicines)?**

On a scale from 0 (same difficulty) to 5 (much more difficult), how difficult is it to validate surrogate endpoints used in the evaluation of medical devices (compared to medicines)?

0                      1                      2                      3                      4                      5

Please provide a rationale for your score:

**C7. Do you think current regulatory and HTA processes and guidelines allow you to address and evaluate the uncertainty of surrogate-based submissions?**

C8. Do you know of any guidance for validating surrogate endpoints? (Is any of it specific to medical devices?) Which methodological reference do you use/would you use to perform a surrogate validation study?

C9. If you consider current guidance for the use of surrogate endpoints, on a scale from 0 (not at all) to 5 (completely), how satisfied are you that it addresses the acceptability of surrogate endpoints?

0	1	2	3	4	5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please provide a rationale for your score:

C10. The knowledge base and practice might be improved by having common guidance, shared across agencies and countries. On a scale from 0 (not important at all) to 5 (critical), how important do you think it would be to improve guidance for the use of surrogate endpoints in health care decision making?

0	1	2	3	4	5
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please provide a rationale for your score:

C11. According to your opinion, are there particular aspects in the regulatory or payer/HTA process regarding surrogate endpoints that require more methodological research? (Could you describe these aspects?) [if no answer, probe about validation of surrogate endpoints and availability of evidence to support surrogate endpoints]

#### SECTION D: CLOSING REMARKS

Interviewer: We have reached the conclusion of the interview. Do you have any questions or observations that you would wish to make?

If you wish to know about the results of this project, we will notify you once the report and related publication will be available



## APPENDIX 4: CHOICE EXPERIMENT: SCENARIO 1 - PHARMACEUTICAL EVALUATION (SECTION C: EXPERIMENTAL SCENARIO I)

Responder	S1 Q1 (Surrogate acceptability)	YES/ NO	S1 Q2 (Technology coverage)	YES/ NO
ID1	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a different class of therapies</li> <li>- strength of association: was 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 100,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed improvement, although no statistically significant (P=0.10)</li> </ul>	NO
ID2	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a large observational study</li> <li>- validation evidence derived from a different class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 100,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	NO
ID3	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCT</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,000</li> <li>- mean utility score: 0.60</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	YES
ID4	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 100,000</li> <li>- mean utility score: 0.60</li> <li>- current best treatment: off-label treatment with a pharmaceutical indicated for heart failure</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	NO
ID5	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a large observational study</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	NO	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	NO
ID6	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a large observational study</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 100,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: off-label treatment with a pharmaceutical indicated for heart failure</li> <li>- immature data on final endpoint showed improvement, although no statistically significant (P=0.10)</li> </ul>	NO
ID7	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a large observational study</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed improvement, although no statistically significant (P=0.10)</li> </ul>	NO
ID8	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a different class of therapies</li> <li>- strength of association: was 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>	NO	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: off-label treatment with a pharmaceutical indicated for heart failure</li> <li>- immature data on final endpoint showed improvement, although no statistically significant (P=0.10)</li> </ul>	YES
ID9	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a large observational study</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.85 (95% confidence interval [0.77, 0.93])</li> </ul>	YES	<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,000</li> <li>- mean utility score: 0.60</li> </ul>	NO

	- surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.		- current best treatment: off-label treatment with a pharmaceutical indicated for heart failure - immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID10	- source of evidence for surrogate validation: a large observational study - validation evidence derived from a different class of therapies - strength of association: was 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.30 - current best treatment: off-label treatment with a pharmaceutical indicated for heart failure - immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID11	- source of evidence for surrogate validation: a large observational study - validation evidence derived from a different class of therapies - strength of association: was 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.30 - current best treatment: best supportive care (i.e. there is no specific therapy) - immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID12	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from the same class of therapies - strength of association: was 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.60 - current best treatment: off-label treatment with a pharmaceutical indicated for heart failure - immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)	
ID13	- source of evidence for surrogate validation: a large observational study - validation evidence derived from the same class of therapies - strength of association: was 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.30 - current best treatment: off-label treatment with a pharmaceutical indicated for heart failure	
			- immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID14	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a different class of therapies - strength of association: was 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.30 - current best treatment: best supportive care (i.e. there is no specific therapy) - immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID15	- source of evidence for surrogate validation: a large observational study - validation evidence derived from a different class of therapies - strength of association: was 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.60 - current best treatment: best supportive care (i.e. there is no specific therapy) - immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID16	- source of evidence for surrogate validation: a large observational study - validation evidence derived from a different class of therapies - strength of association: was 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.		- disease prevalence: 1 in 1,000 - mean utility score: 0.60 - current best treatment: best supportive care (i.e. there is no specific therapy) - immature data on final endpoint showed improvement, although no statistically significant (P=0.10)	
ID17	- source of evidence for surrogate validation: a large observational study - validation evidence derived from a different class of therapies - strength of association: was 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.		- disease prevalence: 1 in 100,000 - mean utility score: 0.60 - current best treatment: off-label treatment with a pharmaceutical indicated for heart failure - immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)	

ID18	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a different class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	
ID19	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.90 ng/ml observed in about 20% of the studies in the indication.</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 100,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: off-label treatment with a pharmaceutical indicated for heart failure</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	
ID20	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a large observational study</li> <li>- validation evidence derived from the same class of therapies</li> <li>- strength of association: was 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -0.10 ng/ml observed in about 70% of the studies in the indication.</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 100,000</li> <li>- mean utility score: 0.30</li> <li>- current best treatment: best supportive care (i.e. there is no specific therapy)</li> <li>- immature data on final endpoint showed deterioration, although no statistically significant (P=0.10)</li> </ul>	

**APPENDIX 5: CHOICE EXPERIMENT SCENARIO 2: MEDICAL DEVICE (SECTION C: EXPERIMENTAL SCENARIO I)**

Responder	S2 Q1 (Surrogate acceptability)	YES/ NO	S2 Q2 (Technology coverage)	YES/ NO
ID1	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCTs</li> <li>- validation evidence derived from a pharmaceutical</li> <li>- strength of association: 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,500</li> <li>- mean utility score: 0.79</li> <li>- alternative therapy available: yes</li> <li>- immature data on final endpoint appearing to favour the intervention</li> </ul>	
ID2	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCT</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: yes</li> <li>- immature data on final endpoint appearing to favour the control</li> </ul>	
ID3	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCT</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.79</li> <li>- alternative therapy available: yes</li> <li>- immature data on final endpoint appearing to favour the control</li> </ul>	
ID4	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a pharmaceutical</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,500</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: no</li> <li>- immature data on final endpoint appearing to favour the intervention</li> </ul>	
ID5	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCT</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.79</li> <li>- alternative therapy available: yes</li> </ul>	

	- surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- immature data on final endpoint appearing to favour the intervention	
ID6	- source of evidence for surrogate validation: a single RCTs - validation evidence derived from a pharmaceutical - strength of association: 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 1,500 - mean utility score: 0.57 - alternative therapy available: no - immature data on final endpoint appearing to favour the control	
ID7	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a similar medical device - strength of association: 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 11 - mean utility score: 0.57 - alternative therapy available: yes - immature data on final endpoint appearing to favour the intervention	
ID8	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a similar medical device - strength of association: 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 11 - mean utility score: 0.57 - alternative therapy available: no - immature data on final endpoint appearing to favour the control	
ID9	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a similar medical device - strength of association: 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 11 - mean utility score: 0.79 - alternative therapy available: no - immature data on final endpoint appearing to favour the control	
ID10	- source of evidence for surrogate validation: a single RCT - validation evidence derived from a pharmaceutical		- disease prevalence: 1 in 11 - mean utility score: 0.79	
	- strength of association: 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- alternative therapy available: no - immature data on final endpoint appearing to favour the intervention	
ID11	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a pharmaceutical - strength of association: 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 11 - mean utility score: 0.57 - alternative therapy available: yes - immature data on final endpoint appearing to favour the control	
ID12	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a pharmaceutical - strength of association: 0.85 (95% confidence interval [0.77, 0.93]) - surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)		- disease prevalence: 1 in 11 - mean utility score: 0.79 - alternative therapy available: yes - immature data on final endpoint appearing to favour the intervention	
ID13	- source of evidence for surrogate validation: a meta-analysis of several RCTs - validation evidence derived from a similar medical device - strength of association: 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 1,500 - mean utility score: 0.79 - alternative therapy available: no - immature data on final endpoint appearing to favour the control	
ID14	- source of evidence for surrogate validation: a single RCT - validation evidence derived from a pharmaceutical - strength of association: 0.30 (95% confidence interval [0.20, 0.40]) - surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)		- disease prevalence: 1 in 1,500 - mean utility score: 0.57 - alternative therapy available: no - immature data on final endpoint appearing to favour the intervention	

ID15	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a pharmaceutical</li> <li>- strength of association: 0.85 (95% confidence interval [0.77, 0.93])</li> <li>- surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,500</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: no</li> <li>- immature data on final endpoint appearing to favour the intervention</li> </ul>	
ID16	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 1,500</li> <li>- mean utility score: 0.79</li> <li>- alternative therapy available: yes</li> <li>- immature data on final endpoint appearing to favour the control</li> </ul>	
ID17	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCT</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: yes</li> <li>- immature data on final endpoint appearing to favour the intervention</li> </ul>	
ID18	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCT</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: no</li> <li>- immature data on final endpoint appearing to favour the control</li> </ul>	
ID19	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a single RCT</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -10 mm Hg (observed in about 20% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: no</li> <li>- immature data on final endpoint appearing to favour the control</li> </ul>	
ID20	<ul style="list-style-type: none"> <li>- source of evidence for surrogate validation: a meta-analysis of several RCTs</li> <li>- validation evidence derived from a similar medical device</li> <li>- strength of association: 0.30 (95% confidence interval [0.20, 0.40])</li> <li>- surrogate threshold effect (STE): -4 mm Hg (observed in about 70% of the studies in the indication)</li> </ul>		<ul style="list-style-type: none"> <li>- disease prevalence: 1 in 11</li> <li>- mean utility score: 0.57</li> <li>- alternative therapy available: no</li> <li>- immature data on final endpoint appearing to favour the control</li> </ul>	