

**SPECIAL ISSUE PAPER**

# Informed decision-making: Statistical methodology for surrogacy evaluation and its role in licensing and reimbursement assessments

Christopher J. Weir<sup>1</sup>  | Rod S. Taylor<sup>2</sup>

<sup>1</sup>Edinburgh Clinical Trials Unit,  
University of Edinburgh, Edinburgh, UK

<sup>2</sup>Institute of Health & Wellbeing,  
University of Glasgow, Glasgow, UK

**Correspondence**

Christopher J. Weir, Edinburgh Clinical  
Trials Unit, University of Edinburgh,  
Edinburgh, UK.

Email: [christopher.weir@ed.ac.uk](mailto:christopher.weir@ed.ac.uk)

**Funding information**

MRC Methodology Research Panel,  
Grant/Award Number: MR/V038400/1

**Abstract**

The desire, by patients and society, for faster access to therapies has driven a long tradition of the use of surrogate endpoints in the evaluation of pharmaceuticals and, more recently, biologics and other innovative medical technologies. The consequent need for statistical validation of potential surrogate outcome measures is a prime example on the theme of statistical support for decision-making in health technology assessment (HTA). Following the pioneering methodology based on hypothesis testing that Prentice presented in 1989, a host of further methods, both frequentist and Bayesian, have been developed to enable the value of a putative surrogate outcome to be determined. This rich methodological seam has generated practical methods for surrogate evaluation, the most recent of which are based on the principles of information theory and bring together ideas from the causal effects and causal association paradigms. Following our synopsis of statistical methods, we then consider how regulatory authorities (on licensing) and payer and HTA agencies (on reimbursement) use clinical trial evidence based on surrogate outcomes. We review existing HTA surrogate outcome evaluative frameworks. We conclude with recommendations for further steps: (1) prioritisation by regulators and payers of the application of formal surrogate outcome evaluative frameworks, (2) application of formal Bayesian decision-analytic methods to support reimbursement decisions, and (3) greater utilization of conditional surrogate-based licensing and reimbursement approvals, with subsequent reassessment of treatments in confirmatory trials based on final patient-relevant outcomes.

**KEYWORDS**

licensing, reimbursement, surrogate outcome

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Pharmaceutical Statistics* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Andy Grieve's outstanding work in pharmaceutical statistics has spanned a broad spectrum of drug development activities, from toxicology studies, through later phase trials to manufacturing and pharmacoeconomics.<sup>1</sup> A common thread has run through all of these areas regarding the valuable role of practical applications of Bayesian methodology. One distinctive feature, given the direct probabilistic inference that Bayesian approaches allow on parameters of interest, is to enable informed decision-making during the development and deployment of pharmaceutical interventions.

Co-author Weir first encountered Andy – or “Mr Bayes”, to use the moniker assigned by the clinical colleague who made the introduction – when on secondment with Pfizer Global Research & Development in the early 2000s. During the subsequent year of collaboration, the topic of statistical evaluation of candidate surrogate outcomes was one methodological area of focus. Since this field contains parallels with many of Andy's wider contributions to pharmaceutical statistics—in particular on the theme of statistical support for decision-making in drug development—we select it for further attention in this reflective piece.

Driven by patients' and society's desire for faster access to therapies, and in parallel with the development of statistical methods for surrogate evaluation described below, there has been a long tradition of the use of surrogate endpoints in drug development. In addition to quicker market decisions (and return on R&D investment), surrogate endpoints also offer the healthcare industry the benefit of more efficient (smaller and quicker) and thus less costly clinical trial portfolios or, alternatively, the targeting of a broader population indication for a similar size of clinical trial.<sup>2</sup>

We consider the evolving role played by statistical methods for evaluating surrogate outcomes in the assessment of treatment efficacy by researchers, regulators and more recently in the decision-making around reimbursement for the adoption of licensed medicines by healthcare systems. Themes include the contrasting approaches of causal inference and association-based methods for surrogate evaluation; the potential use of Bayesian methodology; the role of surrogate evaluation frameworks in licensing to synthesize evidence on a surrogate; and the potential for decision-analytic approaches to support reimbursement decision-making. We conclude with recommendations on potentially fruitful areas for further exploration on this topic.

## 2 | STATISTICAL METHODS FOR SURROGATE EVALUATION

### 2.1 | Definitions and notation

To help orientate the reader, we adopt in this article the definitions for biomarker, clinical endpoint and surrogate endpoint established by The National Institutes of Health (NIH) Biomarkers Definitions Working Group<sup>3</sup> (Box 1). We add the term intermediate endpoint which now has increasingly prominent usage in the schema of outcome measures, particularly in the field of health policy.<sup>4</sup> Note that the final patient-relevant outcome of interest is not necessarily clinical, for example in the case of health-related quality of life.

#### **BOX 1** Definitions of biomarker, intermediate, surrogate endpoint and clinical endpoint

*Biological marker (Biomarker):* A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

*Clinical endpoint:* A characteristic or variable that reflects how a patient feels, functions or survives.

*Intermediate endpoint:* A clinical endpoint such as measure of a function or of a symptom (disease-free survival, angina frequency, exercise tolerance) but not the ultimate endpoint of the disease, such as survival or the rate of irreversible morbid events (stroke, myocardial infarction).<sup>5</sup> Improvement in an intermediate endpoint due to treatment can be of value to the patient even if it does not lead to the improvement of morbidity or mortality.

*Surrogate endpoint:* A biomarker or intermediate endpoint intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (efficacy/effectiveness) or harms.

In the notation that follows, we use  $S$  to denote the surrogate outcome, with  $S_{ij}$  indicating the value of the surrogate on participant  $i$  within trial  $j$ . Similarly  $T_{ij}$  refers to the value of the clinical or final patient-relevant endpoint of interest on participant  $i$  in trial  $j$ .  $Z_{ij}$  will be used as an indicator of randomized treatment allocation for participant  $i$  within trial  $j$ , with a value of 0 for placebo or active comparator, and 1 for the intervention under investigation.

## 2.2 | Statistical methods evaluating surrogate outcome validity: first steps

Since the pioneering step in surrogate evaluation introduced by Prentice in 1989,<sup>6</sup> a plethora of extensions and alternative approaches has emerged. We would direct readers wishing a complete picture of the range of approaches available towards previously published reviews on the topic.<sup>4,7</sup>

A non-exhaustive list of methods (Table 1) indicates the considerable number and diversity of strategies developed, spanning the full range of types of surrogate and clinical outcomes. We now introduce a scheme for classifying methods and discuss in more detail a small number of approaches which have been influential in the overall development of thinking on surrogate evaluation. In particular we identify those which are novel or methodologically notable (such as the Prentice method and an early introduction of a Bayesian approach<sup>8</sup>); those which highlight the complementary roles of causal and associative methods (for example, principal stratification<sup>9</sup> or direct and indirect effects<sup>10,11</sup>); and those which have been widely adopted as tools for surrogate evaluation (such as meta-analytic approaches,<sup>12</sup> especially those grounded in information theory<sup>13</sup>).

## 2.3 | Is a causal interpretation strictly necessary?

An enduring theme which has emerged in the evolution of statistical methodologies for surrogate evaluation is the question of causal inference. In this context, Joffe and Greene<sup>14</sup> provided a helpful typology through which to classify methods for surrogate evaluation. They considered and contrasted two paradigms: first, that of causal effects (CE); and second, the concept of causal association (CA). Fundamentally, the causal association approach uses information on the relationship between the effect of treatment on the surrogate and the treatment effect on the clinical outcome to enable the surrogate effect to be used to predict treatment efficacy on the clinical outcome. In contrast, the causal effects paradigm uses the effect of treatment on the surrogate in combination with knowledge of the effect of the surrogate on the clinical outcome to draw inferences about the effect of treatment on the clinical outcome.

There has been substantial divergence in methodological applications to date, for example with causal effects approaches gaining much uptake in the field of vaccine development<sup>15</sup> and causal association methods being widely implemented in oncology.<sup>16</sup> Thoughts are now turning, however, towards commonalities between the CE and CA paradigms. For normally-distributed  $S$  and  $T$ , Conlon and co-authors provide helpful insights on the interdependencies of CE and CA.<sup>17</sup> Adopting a structural model containing an unobserved confounder, they illustrate the impact of the identifiability assumptions required by CE methods on the quantification of surrogacy using CA approaches. In a reciprocal manner, they also highlight the impact that the assumptions made to support estimation under CA methods would have on CE parameters. We will refer to the causal effects and causal association classification when introducing the key methodological developments below and will subsequently comment on moves towards a more unified approach to surrogacy evaluation.

## 2.4 | Key developments

### 2.4.1 | Prentice criteria

In a first investigation into formal statistical approaches for evaluating surrogate outcomes, Prentice<sup>6</sup> stated that a surrogate should “yield unambiguous information about differential treatment effects on the true endpoint” and defined a confirmed surrogate endpoint as “a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint.” This definition was then operationalized for  $t$ , the time to failure on a given clinical endpoint. In the notation below  $F(t)$  indicates, for clinical endpoint  $T$ , the distribution of censoring history and failure times up to time  $t$ . The key

TABLE 1 Statistical approaches to surrogate outcome evaluation.

References	Method	Type of surrogate	Type of clinical outcome
6	Prentice criteria (Prentice, 1989)	Time to event	Time to event
57	Proportion of treatment effect explained (Freedman et al, 1992)	Binary	Binary
58	Proportion of treatment effect explained (Lin et al, 1997)	Continuous	Time to event
59	Single trial: relative effect and adjusted association (Buyse and Molenberghs, 1998)	Continuous	Continuous
12	Multi-trial: joint model for S and T (Buyse et al, 2000)	Continuous	Continuous
60	Multi-trial: two stage model (Tibaldi et al, 2003)	Continuous	Continuous
8	Multi-trial: Bayesian joint model (Daniels and Hughes, 1997)	All	All
61–64	Multi-trial: extension to time to event (Burzykowski et al, 2001; Renfro et al, 2012; Ghosh et al, 2012; Tibaldi et al, 2004)	Time to event	Time to event
65–67	Multi-trial: extension to categorical outcomes (Molenberghs et al, 2001; Renard et al, 2002; Alonso et al, 2002)	Binary/ordinal/ continuous	Binary/ordinal/ continuous
21,68,69	Multi-trial: repeated measures (Alonso et al, 2003; Alonso et al, 2004; Alonso et al, 2006)	Continuous	Continuous
70	Multi-trial: repeated measures (Pryseley et al, 2010)	Continuous	Discrete
71	Multi-trial: repeated measures (Renard et al, 2003)	Continuous	Time to event
23	Surrogate threshold effect (Burzykowski and Buyse, 2006)	All	All
72	Likelihood reduction factor (individual level surrogacy) (Alonso et al, 2004)	All	All
73,74	Proportion of information gain (Qu and Case, 2007; Miao et al, 2012)	All	All
13	Multi-trial: information theoretic approach (Alonso and Molenberghs, 2007)	Binary	Binary
75	Multi-trial: information theoretic approach extension (Alonso and Molenberghs, 2008)	Time to event	Time to event
76	Multi-trial: information theoretic approach extension (Pryseley et al, 2007)	Binary	Continuous
77,78	Multi-trial: information theoretic approach extension (Ensor and Weir, 2020; Ensor and Weir, 2021)	Binary/ordinal	Binary/ordinal
9,19	Principal stratification, principal surrogate (Frangakis and Rubin, 2002; Li et al, 2011)	Binary	Binary
18	Principal stratification extension (Wolfson and Gilbert, 2010)	Continuous	Binary
79	Principal stratification extension (Conlon et al, 2014)	Continuous	Continuous
19,25	Principal stratification extension – Bayesian (Li et al, 2010; Li et al, 2011)	Binary	Binary
80	Principal stratification extension – Bayesian (Zigler and Belin, 2012)	Continuous	Binary
81	Causal effect predictiveness (CEP) (Gilbert and Hudgens, 2008)	Discrete or continuous	Binary
82	CEP for time to event (Qin et al, 2008)	Continuous	Time to event
83	Average causal effect (Chen et al, 2007)	Discrete or continuous	Discrete or continuous
84	Distributional causal effect (Ju and Geng, 2010)	All	All
11	Direct and indirect effects via path analysis (Qu and Case, 2006)	All	All
85	Proportion of treatment effect explained – extension to joint modelling (Deslandes and Chevret, 2007)	Repeated measures Time to event (multistate)	Time to event Binary
86,87	Structural equation modelling (Ghosh et al, 2010; Emsley et al, 2010)	All	All
88	Missing data perspective (Chen et al, 2003)	Continuous	Continuous
89	Missing data perspective (Chen et al, 2008)	All	All
90	Missing data perspective (Benda and Gerlinger, 2007)	Continuous	Binary

(Continues)

TABLE 1 (Continued)

References	Method	Type of surrogate	Type of clinical outcome
91	95% prediction model for true outcome (Baker et al, 2012)	Binary/time to event	Binary/time to event
92,93	Joint modelling permitting multiple surrogates (Xu and Zeger, 2001; Lin et al, 2002)	Continuous	Time to event
94	Joint modelling incorporating repeated measures (Taylor and Wang, 2002)	Continuous	Time to event

criterion is that surrogate S should capture all of the dependence of the clinical endpoint hazard rate  $\lambda_T$  on treatment group Z:

$$\lambda_T \{t|S(t), Z\} \equiv \lambda_T \{t|S(t)\}$$

Furthermore, S should predict T,

$$\lambda_T \{t|S(t)\} \neq \lambda_T \{t\}$$

A further condition imposed is that the class of alternatives to the null hypothesis on S must alter the average true endpoint risk. In other words, any treatment effect observed on S must lead to a corresponding effect on the clinical endpoint T:

$$E[\lambda_T \{t|S(t)\} | Z; F(t)] \neq E[\lambda_T \{t|S(t)\} | F(t)]$$

This ground-breaking initial step in surrogacy evaluation prompted numerous attempts to build on this conceptual framework. It was subsequently classified as a CE method by Joffe and Greene.<sup>14</sup>

## 2.4.2 | Principal surrogacy

Frangakis and Rubin (F&R) introduced principal stratification as a method of modelling, for binary outcomes, the effect of a surrogate endpoint on a true endpoint.<sup>9</sup> The key concept is to consider that, for example in a randomized parallel group clinical trial with two treatment groups, each participant would have two potential outcomes, one for each of the treatments in the trial, but only one of these can be observed, depending on the treatment group to which the participant was randomized.

While space limitations prevent a full exposition of the methodology, it will be useful to introduce the key terminology here. F&R maintained that causal effects should be the basis for surrogate endpoint evaluations, where the causal effect is a comparison between treatment groups of the potential outcomes on the same set of individuals.

The basic principal stratification for a surrogate partitions individuals so that within each stratum, all individuals have the same vector of potential outcomes for the two treatment groups. Building on this, the principal stratification partitions individuals by forming unions of the sets in the basic principal stratification. A principal effect compares outcomes across treatments within a principal stratum, and hence any principal effect is also a causal effect.

F&R then posited two requirements for surrogate validity. The first of these was causal necessity, which requires that an effect of treatment on T can only exist if treatment has also affected S. The second condition was statistical generalizability, which requires good predictive performance of S for T in a future study in which only S is observed.

On this basis F&R then established a new definition for surrogate validity: S is a principal surrogate when comparing the effect of treatments 0 and 1 on T if, for all fixed values s of the surrogate, the comparison of the principal strata of the ordered sets of potential outcomes  $\{T_i(0): S_i(0) = S_i(1) = s\}$  and  $\{T_i(1): S_i(0) = S_i(1) = s\}$  results in equality. A principal surrogate always meets the requirement of causal necessity.

Identifiability of counterfactual estimands under this methodology poses some challenges.<sup>18</sup> Li et al (2011) addressed some practical aspects of implementation<sup>19</sup> and further solutions are outlined in the *Bayesian* approaches

section below. An extension from binary to time-to-event outcomes has been derived.<sup>20</sup> Joffe and Greene<sup>14</sup> consider principal surrogacy to belong within the CA class of methods.

From the contrasting Prentice and principal surrogacy approaches grounded in hypothesis testing and conceptual frameworks, a natural progression is then to consider methods which focus on the estimation of the strength of surrogacy.

### 2.4.3 | Meta-analysis and information theory

It was rapidly recognized that a large quantity of data is required to estimate the value of a putative surrogate outcome. This led to several approaches making use of meta-analysis techniques to allow all available data to be utilized. This in turn resulted in two levels of surrogacy performance being proposed, based on the coefficient of determination between the effects of treatment on the surrogate and clinical endpoint at the individual patient level ( $R^2_{\text{indiv}}$ ) and overall in a clinical trial ( $R^2_{\text{trial}}$ ).<sup>12</sup>

Initially, a joint mixed model, including random intercepts and random treatment effects across trials, was proposed for evaluating the effects of treatment on the surrogate and clinical outcomes.<sup>12</sup> This proved to be computationally burdensome, and so the two-stage fixed effects modelling approach presented by the same authors was a preferable alternative.

The original formulation for continuous outcomes was subsequently extended to a wide range of outcome types (Table 1). Here, the challenge was that in doing so the formulation of  $R^2_{\text{indiv}}$  differed across contexts and for discrete outcomes in some cases had to be assessed at the latent variable level. The likelihood reduction factor (LRF)<sup>21</sup> was introduced as a measure which could be applied consistently across different types of outcome.

Further theoretical support for a consistent framework across outcome types was sought using the information theoretic approach to quantifying uncertainty. Entropy quantifies the uncertainty in a discrete random variable, and entropy power extends this concept to continuous random variables.<sup>22</sup> Using entropy power, Alonso and Molenberghs (2007)<sup>13</sup> presented information theoretic surrogacy evaluation metrics for both the individual and trial levels.  $R^2_{\text{h}}$  is based on entropy power (EP) and can be thought of as the proportion of uncertainty in T at the individual level removed by adjusting for S.

$$R^2_{\text{h}} = \frac{EP(T) - EP(T|S)}{EP(T)}$$

An equivalent measure,  $R^2_{\text{ht}}$ , was derived for trial level surrogacy, this time based on the entropy power of the relationship between estimated treatment effects,  $\alpha_i$  and  $\beta_i$ , on each of S and T in trial  $i$ .

$$R^2_{\text{ht}} = 1 - \frac{EP(\beta_i|\alpha_i)}{EP(\beta_i)}$$

Finally, a meta-analytic  $R^2_{\text{h}}$  was proposed to take account of between-trial heterogeneity, given that  $k$  trials produce  $k_q$  possible  $R^2_{\text{hi}}$ :

$$R^2_{\text{h}} = \sum_{i=1}^{k_q} \tau_i R^2_{\text{hi}}$$

where  $\tau_i > 0 \forall i$  and  $\sum_{i=1}^{k_q} \tau_i = 1$ .

The choice of  $\tau_i$  is drawn from an uncountable family of parameters; nevertheless, the LRF represents a consistent estimator of  $R^2_{\text{h}}$  and supports unification across settings by providing a common interpretation regardless of the type of outcome being studied. The appeal of such a consistent treatment of surrogacy evaluation across settings has prompted extensions to other outcome types, including time to event, binary, ordinal, and repeated measures (Table 1). These meta-analytic, information-theoretic methods have been widely adopted as the preferred approach under the CA paradigm.

A natural question that follows is: at what level of such trial  $R^2$  or individual  $R^2$  measures should we conclude that a surrogate is valid? Some have proposed specific cut-off values (for example, the surrogacy evaluation framework of the German Institute for Quality and Efficiency in Health Care [IQWiG], see Table 2), although context will surely also be

TABLE 2 International payers/HTA agencies &amp; handling of surrogate endpoints: recommendations for specific statistical methods/considerations.

Country	Payer/agency	Statistical methods	Citation/web link(s)
United Kingdom	National Institute of Health & Care Excellence (NICE), 2013/9	<p>“When the use of “final” clinical endpoints is not possible and “surrogate” data on other outcomes are used to infer the effect of treatment on mortality and health-related quality of life, evidence in support of the surrogate-to-final endpoint outcome relationship must be provided together with an explanation of how the relationship is quantified for use in modelling. The usefulness of the surrogate endpoint for estimating QALYs will be greatest when there is strong evidence that it predicts health-related quality of life and/or survival. In all cases, the uncertainty associated with the relationship between the endpoint and health-related quality of life or survival should be explored and quantified.</p> <p>Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints:</p> <p>“When data on the final clinical outcome are not available or limited at the licensing stage, and therefore also for the HTA decision-making process, a modelling framework is required to establish the strength of the surrogate relationship between the treatment effects on the surrogate and the final outcome and to predict the likely treatment effect on the final outcome for the new health technology. Multivariate meta-analytic methods provide such a framework as they, by definition, take into account the correlation between the treatment effects on the surrogate and final outcomes as well as the uncertainty related to all parameters describing the surrogate relationship.”</p> <p>“Relying solely on patient-level association is not sufficient when evaluating surrogate endpoints, in particular when individual-level association has been evaluated based on data from a single trial (Fleming and DeMets 1996). A meta-analytic approach based on data from more than one trial to establish the association between the treatment effects on the candidate surrogate endpoint and on the final outcome is more appropriate for evaluation of surrogate endpoints.”</p>	<p><a href="https://www.nice.org.uk/process/pmg9/chapter/foreword">https://www.nice.org.uk/process/pmg9/chapter/foreword</a></p> <p><a href="http://nicedsu.org.uk/multivariate-meta-analysis-td/">http://nicedsu.org.uk/multivariate-meta-analysis-td/</a></p>
Germany	IQWiG, 2011	<p>“There is no “best” method defined, however, correlation-based validation is the ‘preferred’ method, in the sense it has been most widely used in evaluations. Another option discussed is the surrogate threshold effect (STE)”</p> <p>“A correlation of <math>R \geq 0.85</math>; <math>R^2 \geq 0.72</math> measured at the lower bound of the 95% percentage interval allows to conclude that the validation study represents a high reliable result. This interval <math>R &lt; 0.85</math>; <math>R^2 &lt; 0.72</math> to <math>R &gt; 0.7</math>; <math>R^2 &gt; 0.49</math> represents a medium reliable result between surrogate and patient relevant endpoint. If a validation study shows high reliable results with statistically low correlation (<math>R \leq 0.7</math>; <math>R^2 \leq 0.49</math>) measured at the lower bound of the confidence interval then the surrogate is not considered as a valid endpoint.”</p>	<p><a href="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</a></p>
Canada	CADTH, 2017	<p>“Researchers should evaluate and justify the validity of any surrogate endpoints used for parameter estimation. Uncertainty in the association of the surrogate to the final clinical outcome should be reflected in the reference case probabilistic analysis. This uncertainty can also be explored through appropriate scenario analyses. The existence of multiple potential surrogates should be reflected in the analysis of uncertainty.”</p>	<p><a href="https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition">https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition</a></p>

TABLE 2 (Continued)

Country	Payer/agency	Statistical methods	Citation/web link(s)
Australia	PBAC 2008/2016	<p>Transformation of a surrogate to a final outcome – suggested uncertainty analysis – use range of alternative plausible values and present as scenario analysis for economic evaluation</p> <p>Surrogate to Final Outcomes Working Group (STFOWG) report: Information requirements</p> <p>“For a meta-regression of multiple randomised trials: the results for (1) the intercept and coefficient (and their 95% CI), (2) the coefficient of determination (<math>R^2</math> trial), (3) <math>R^2</math> individual (only if individual patient data are available for the trials), and (4) the surrogate threshold effect (STE) as determined by prediction bands.”</p> <p>“An assessment of the implications of the identified uncertainties relating to the PSM to TCO transformation for the structure, the input variables, the results and the sensitivity analyses of the modelled economic evaluation and for the presentation of the stepped economic evaluation.”</p>	<p><a href="https://pbac.pbs.gov.au/information/printable-version-of-guidelines.html">https://pbac.pbs.gov.au/information/printable-version-of-guidelines.html</a></p> <p><a href="https://www.pbs.gov.au/industry/useful-resources/pbac-technical-working-groups-archive/surrogate-to-final-outcomes-working-group-report-2008.pdf">https://www.pbs.gov.au/industry/useful-resources/pbac-technical-working-groups-archive/surrogate-to-final-outcomes-working-group-report-2008.pdf</a></p>

Abbreviations: PSM, proposed surrogate measure; TCO, target clinical outcome.



important. Alternatively, the surrogate threshold effect (STE) has been created as a practical measure to define the minimum level of efficacy required on the surrogate to conclude that efficacy would also be present on the clinical endpoint.<sup>23</sup>

#### 2.4.4 | Bayesian approaches

As the meta-analytic approaches discussed above evolved, a further CA strategy was proposed by Daniels and Hughes<sup>8</sup> using a Bayesian random effects meta-analysis. In essence, the model focuses on trial-level effects as follows:

$$\theta_i \mid \gamma_i \sim N(\alpha + \beta\gamma_i; \tau^2)$$

where  $\theta_i$  and  $\gamma_i$  are the treatment effects for trial  $i$  on T and S respectively. Minimally informative priors are placed on the fixed effects and the regression coefficients. An informative surrogate consistent with the Prentice criteria would have  $\alpha = 0$  and  $\beta \neq 0$ ; a perfect surrogate would also require  $\tau^2 = 0$ . At first glance it appears that use of the Bayesian approach may address some of the computational issues originally identified with the meta-analytic approach involving the fitting of a joint mixed model with random intercepts and treatment effects; however, it has been noted that, as the number of trials in the meta-analysis increases, inconsistent results are likely to arise due to the increasing number of nuisance parameters on which uninformative priors are being placed.<sup>24</sup>

Note that Bayesian approaches will also be readily applicable to most, if not all, of the methodologies outlined in Table 1. For example, Li et al (2010)<sup>25</sup> implement the principal stratification approach using a Bayesian imputation technique to address problems of identifiability, and Elliott et al<sup>26</sup> extend this further to incorporate missing data in the potentially observable final outcome under the scenarios of ignorable and non-ignorable missingness. Bayesian inference also lends itself well to the regulatory and reimbursement decision-making requirements discussed in Section 3.

## 2.5 | Towards a unified approach

We conclude this methodology overview with an update on recent advances which have identified previously unrecognized connections between the causal effects and causal association principles. Alonso et al (2015)<sup>27</sup> make a case for the use of expected causal effects, rather than individual causal effects, on the grounds that the expected causal effects are readily estimable and acceptable to regulatory authorities. They then explain that the results of a CA information theory-based analysis generating an individual causal association and trial-level surrogacy measure might be anticipated to be consistent with that of the individual causal effect. Further investigation will be needed to demonstrate whether this is the case across all types of surrogate and clinical endpoints. A subsequent development<sup>28</sup> establishes a relationship between the individual causal association (CA) and the surrogate predictive function (CE) and implements a Monte Carlo approach to address identifiability problems. In the context of a binary surrogate and binary clinical outcome, Alonso et al (2018)<sup>29</sup> offer a maximum entropy strategy from information theory to aid with identifiability in estimating the CE surrogate prediction function and the individual causal association (CA), and suggest that these measures may be used jointly to assess surrogacy. Thus, strands from the CE and CA paradigms have been combined in the quest for surrogacy evaluation. Alonso et al (2019)<sup>30</sup> emphasize that this new measure of individual causal association does not substitute for the original  $R^2_{\text{indiv}}$  which retains distinct value in its own right. Van der Elst et al (2021)<sup>31</sup> go on to extend the original surrogate threshold effect metric from the CA paradigm to the new individual causal association in the CE context.

## 2.6 | The next steps

It is clear from the above that the pharmaceutical statistician has a comprehensive range of methods at their disposal with which to quantify the efficacy of an intervention, based on evidence that includes data on one or more putative surrogate outcomes. We now go on to discuss the ways in which this evidence might be synthesized, alongside other factors such as quality, safety and cost-effectiveness, to inform the decision-making of regulatory authorities (on licensing) and payer and health technology assessment (HTA) agencies (on reimbursement).

### 3 | USE OF SURROGATE ENDPOINTS IN HEALTHCARE POLICY

Access to drugs and other medical healthcare technologies across international health-care systems depends on overcoming two evidentiary hurdles. First, regulatory bodies, such as the Food and Drug Administration (FDA) in United States, European Medicines Agency (EMA), and the Medicines and Healthcare products Regulatory Agency (MHRA) in United Kingdom, determine whether a healthcare technology should receive a license for entry to the market and be available for use in clinical practice. Such a license is typically based on three criteria: quality (are the raw materials, equipment, and the technical knowledge required to process, package, and distribute the therapy sufficient?), efficacy (does the therapy have demonstrable health benefit?) and safety (Is the therapy safe?). Second, subsequent to or in parallel with licensing, payers (often supported by HTA agencies) decide whether the therapy should be reimbursed, be that through a public funding or a private insurance system. Payer bodies, such as National Institute for Health and Care Excellence (NICE) in United Kingdom, Institute for Quality and Efficiency in Health Care (IQWiG) in Germany, or Centre for Medicare and Medicaid (CMS) in United States, raise two additional evidentiary criteria: comparative effectiveness (does the therapy provide sufficient health benefit over the longer term compared to standard of care?) and cost-effectiveness (does the benefit relative to the cost of therapy provide good value for money?).<sup>32</sup> Surrogate endpoints can play a key role in both the licensing and reimbursement settings.<sup>4,33</sup>

#### 3.1 | Licensing

Acceptance based on surrogate outcomes was initially limited to specific licensing circumstances. In response to the AIDS epidemic of the 1980s, the FDA introduced its “accelerated approval” regulation in 1992 to approve new drugs or biologicals intended for treatment of “serious illnesses” that offer meaningful therapeutic benefit compared with existing treatment on the basis of a documented treatment effect on a surrogate or non-ultimate endpoint.<sup>34</sup> Accelerated approval focuses on unmet medical need and can include chronically debilitating conditions, where the number of patients in trials is likely to be small and there is no existing licensed therapy, or orphan medicines, used against rare diseases, often also without existing treatments. Although initially developed for accelerated approval for life threatening diseases without treatment options, agencies such as the FDA and EMA have steadily expanded use of surrogates far beyond this original intent. In 2018, 73% of FDA licensed drugs (43/59) received expedited approval and many of these were for conditions such as gout and hypertension that were neither life-threatening nor lacking in existing treatments.<sup>35</sup> A review of FDA cancer drug approvals between 1992 and 2019 found 194 drug authorizations for 132 drugs that were based on surrogate endpoints.<sup>36</sup> Only half (89, 46%) of these approvals were based on accelerated licensing pathways, the remainder being under the regular licensing process. Furthermore, pivotal trials using surrogates as their primary endpoint formed the exclusive basis of FDA approvals for 91 of 206 (44%) indications for novel therapeutic agents between 2005 and 2012.<sup>37</sup>

The potential risk of using surrogate endpoints in the regulatory setting has long been highlighted,<sup>38</sup> based on cases where drugs licensed on the basis of surrogates later turned out to be harmful and increase overall mortality. As the authors note, these effects can be the results of unintended effects of surrogate endpoints in the pathway of the specific disease being treated and action of the therapy (Figure 1). There may exist (1) non-causal associations: a surrogate endpoint that is thought to be causal might simply be associated; (2) multiple causal pathways: a surrogate endpoint may lie within just one of several causal pathways of disease, thereby any changes in the surrogate may have an uncertain, and potentially negligible, effect on the desired patient outcome; (3) unintended outcomes: the therapy may exert effects outside the disease process that can have unmeasured harm or unmeasured benefit on patient outcomes. Even outcomes which appear biologically plausible as surrogates, such as nonfatal myocardial infarction as a surrogate for all-cause or cardiovascular mortality, can fail to achieve acceptable levels of surrogacy.<sup>39</sup>

While it cannot completely prevent the unintended negative final health outcome consequences of using surrogates, it has been proposed that this risk can be reduced by restricting licensing claims to new therapies for which there is an appropriately validated surrogate endpoint.<sup>35</sup> However, use of statistically validated surrogates does not appear to have been consistently implemented in the regulatory setting. In the above survey of FDA cancer drugs,<sup>36</sup> the strength of association between the surrogate endpoint (such as progression free survival or tumour response) and the final outcome of all-cause mortality was often weak or absent. A total of 61% of approvals had no documented correlation between the surrogate and mortality, 16% had poor correlation ( $r \leq 0.7$ ), 2% had medium correlation ( $0.70 < r < 0.85$ ), 5% had high correlation ( $r \geq 0.85$ ), and 17% had varied levels of correlation across multiple validation studies. The

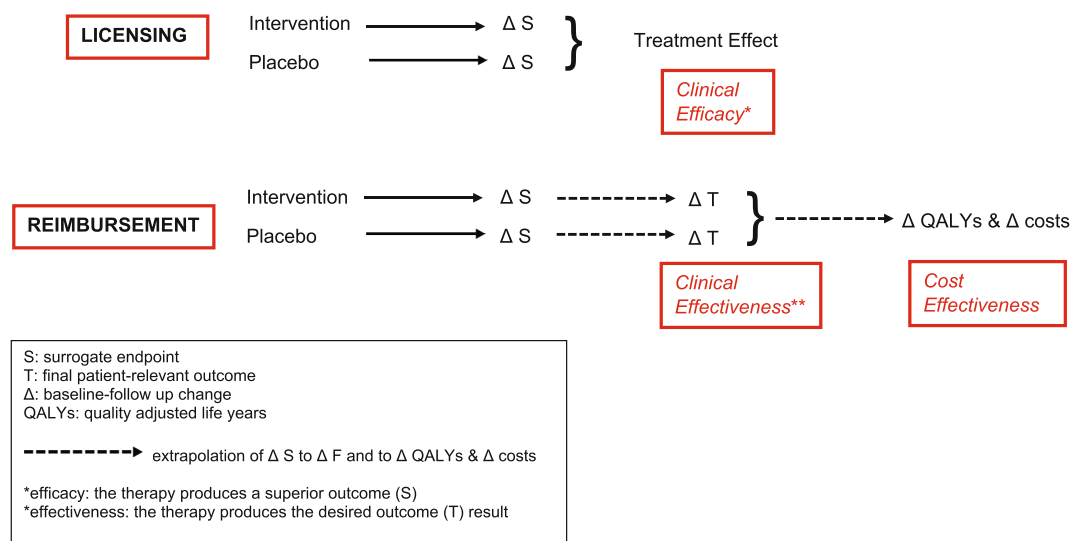


FIGURE 1 Comparison of role of surrogate endpoints in licensing and reimbursement.

FDA's current table of surrogate endpoints is based on either surrogate endpoints accepted by FDA in their previous approvals, or those the FDA consider may be acceptable for future approval application.<sup>40</sup> However, rather paradoxically, the surrogate endpoints listed in the FDA table do not necessarily have evidence of their formal statistical validation.<sup>41</sup> A study of EMA authorizations issued between 2011 and 2018 for products assessed by expedited pathways, found the majority of approvals were based on pivotal trials that reported non-validated surrogate endpoints.<sup>42</sup>

Although we have focused on the role and challenges of surrogate endpoints in licensing for drug and biologic therapies, it is important to acknowledge that these considerations equally apply to other therapies, such as medical devices.<sup>43</sup>

### 3.2 | Payers/reimbursement/HTA

As outlined above, reimbursement decisions by the payer and HTA agencies (hereafter referred to as payer for brevity), such as NICE in the United Kingdom, are typically made based on estimates of clinical effectiveness and cost-effectiveness of new therapies relative to current standard of care. When only the surrogate endpoint from a clinical trial programme (and the regulatory process) is available to assess the cost-effectiveness of medical technology, payers are therefore required to extrapolate the long-term estimates of clinical effectiveness that can include many parameters including costs and utilities, to populate a health economic decision model.<sup>44</sup>

Although for a given new therapy, the same trial evidence of surrogacy is likely to be used in reimbursement as licensing, there is an important fundamental difference in the decision process between settings (Figure 1). Licensing approval of a therapy is likely to follow from demonstration of statistical superiority (typically against placebo in the phase 3 trial setting) on a validated surrogate endpoint, that is, "clinical efficacy" (the therapy produces a superior outcome result). However, in the payer setting, the magnitude of the predicted treatment effect on the final patient relevant outcome based on the observed treatment effect on the surrogate endpoint is also key. The relative magnitude of treatment effect on the surrogate needs to be sufficient to predict a meaningful long-term improvement in the clinical outcome, (e.g., achievement of surrogate threshold effect – see Table 2, which summarizes the statistical considerations for surrogate evaluation recommended by international payer and HTA agencies) that is, "clinical effectiveness" (the therapy produces the desired outcome[s] result). If this can be shown, the payer then needs to judge that the therapy is "cost effective." In the case of NICE this requires the extrapolation of final outcome into quality adjusted life years (QALYs), with demonstration that the incremental cost per QALY gained with therapy is acceptable.<sup>45</sup> A key aspect of the reimbursement decision by payers is therefore to assess the magnitude of therapeutic benefit of a medical technology (relative to additional cost over current standard care). In this context, payers face the further challenge that trials of a surrogate endpoint can lead to substantial overestimation of the treatment effects compared to clinical outcomes. In

TABLE 3 Hierarchy of evidence for surrogate endpoints.

	Definition	Source of evidence	Statistical metrics
Level 3	Biological plausibility	Clinical data and understanding of disease (surrogate endpoints on final patient relevant outcome disease pathway)	Not applicable
Level 2	Observational association	Epidemiological studies of relationship between surrogate endpoint & final patient relevant outcome	Correlation between surrogate endpoint and final patient relevant outcome
Level 1	Interventional/treatment effect association	Randomised controlled trial(s) <sup>a</sup> with treatment change in surrogate endpoint and final patient relevant outcome <sup>b</sup>	Trial level R <sup>2</sup> /Spearman's correlation Surrogate threshold effect (STE)

Note: Adapted from Ciani et al, 2017<sup>4</sup>; Taylor & Elston, 2009<sup>48</sup>; Bucher et al, 1999.<sup>49</sup>

<sup>a</sup>Individual participant or trial level meta-analysis of multiple randomised controlled trial or single large randomised controlled trial with surrogate-final outcome association assessed by trial site.

<sup>b</sup>Evidence should be randomised controlled trials from same disease indication, line of therapy, class of treatment/intervention and comparator therapy that surrogate endpoint is applied. If extrapolating from different disease indication, line of therapy, class of treatment/intervention and comparator therapy then evidence may not qualify as level 1.

their review of 324 cardiovascular trials, Ridker and Torres found that trials with primary endpoints that were surrogates were more likely to report a positive treatment effect (77 of 115 trials; 67%) than trials that reported final patient-relevant primary outcomes (113 of 209 trials; 54%,  $P = 0.02$ ).<sup>46</sup> Many of these trials informed licensing decisions. A meta-epidemiological study involving 185 randomized controlled trials reported in six high-impact general medical journals that used surrogate endpoints or patient-relevant outcomes compared the treatment effects from trials that used surrogates and those that used final outcomes.<sup>47</sup> This analysis found that trials using surrogate endpoints were more likely to report positive treatment effects as trials using final outcomes (52 of 84 trials [62%] vs. 37 of 101 trials [37%],  $P < 0.01$ ). Furthermore, trials that used surrogates found treatment effects that were, on average, 28% to 48% larger than trials that used corresponding clinical outcomes. So, in addition to the requirement for surrogate validity (i.e., treatment effect on surrogate has a strong association with clinical outcome effect), in order to assess clinical and cost effectiveness payers must also know the magnitude of the surrogate-clinical outcome association (and the uncertainty around this) in order to model reliably the long-term clinical and cost-effectiveness of a therapy.<sup>4</sup>

To assist payers in appropriate use of surrogate endpoints in their decision-making, two evaluative frameworks have been developed. The first was originally proposed by Taylor and Elston,<sup>48</sup> based on the JAMA checklist for clinical trials,<sup>49</sup> and updated by Ciani in 2017.<sup>4</sup> We summarize the evaluative framework in Table 3. Here, level 3 evidence for a surrogate is based on biological plausibility alone, whereas evidence is considered to be level 2 when a strong association exists between the surrogate and the clinical endpoint across cohorts or at the level of the individual patient. The highest level of evidence (level 1) requires evidence that technologies that improve the surrogate also improve the clinical outcome across many randomized controlled trials. In 2008, Lassere proposed a multidimensional hierarchical evidence schema for evaluating the status of biomarkers as surrogate endpoints - Biomarker-Surrogacy Evaluation Schema (BSES).<sup>50</sup> BSES extends beyond the Ciani framework to include the dimensions of biological, epidemiological, statistical, clinical trial and risk-benefit evidence of the biomarker clinical endpoint relationships. BSES systematically evaluates and ranks the surrogacy status of biomarkers and surrogate endpoints using defined levels of evidence. The schema incorporates three independent domains: study design (the highest score being  $\geq 3$  RCTs in the appropriate drug classes assessing the relationship between surrogate and final outcome), target outcome (the highest score being for a clinical outcome of mortality), and statistical evaluation (the highest score occurring where an intervention change in the surrogate strongly predicts change in the clinical outcome based on a statistical measure such as  $R^2_{\text{trial}}$ ,  $R^2_{\text{indiv}}$  or STE). Each domain ranks items from 0 to 5. An additional category called "penalties", due to lack of evidence, evidence to the contrary or evidence of harm, incorporates additional considerations of biological plausibility, benefit-risk and generalizability. The total score (0-15) determines the level of evidence, with Level 1 the strongest and Level 5 the weakest. To be designated a Level 1 or 2 surrogate, a biomarker or intermediate outcome must also meet the rank of at least 3 within the study design, target outcome and statistical strength domains, and there must be no RCT evidence that use of the biomarker caused patient harm.

Despite a clear preference for patient-relevant final endpoints and therefore caution around the use of surrogates, there has been no universal uptake of these evaluative frameworks by international payers.<sup>4</sup> Furthermore, few payers

have specific methodological guidance around the assessment of medical technologies based on surrogate outcomes. Grigore et al reported that across 73 HTA agencies, only 29 (40%) had methods guidance that specifically referred to surrogate endpoints. Furthermore, only a very small number of these agencies (NICE in United Kingdom, Australia Pharmaceutical Benefits Advisory Committee (PBAC), Canadian Agency for Drugs and Technologies in Health (CADTH), and German Institute for Quality and Efficiency in Health Care (IQWiG)) had developed more detailed analytic methods and prescriptive criteria for the acceptance of surrogate endpoints, including proposed statistical methods for surrogate validation<sup>43</sup> (see Table 2). While these agencies each acknowledge the lack of methodological consensus including statistical analytic approaches around the validation of surrogates, there is clear emphasis on their preference for multiple randomized trial data supporting the link between surrogate and clinical endpoints (supported by meta-analysis/meta-regression). NICE methods, informed by their technical advisory center (Decision Support Unit, DSU) report,<sup>51</sup> recommend the use of Bayesian multivariate meta-analytic methods to take into account the correlation between the treatment effects on the surrogate and clinical outcomes. However, only one agency (IQWiG) discusses acceptable cut-off numerical values of the level of association between surrogate and final outcome and indicates that the lower bound of the 95% percentage interval must include a correlation ( $r$ )  $\geq 0.85$  or regression-based model  $R^2 \geq 0.72$  to indicate high validity. NICE have recently updated their methods, processes, and topic selection for health technology evaluation.<sup>52</sup> As part of this update, NICE commissioned a number of technical support papers that included considerations on the handling of surrogate outcomes that include the following specific recommendations: “In all cases, the uncertainty associated with the relationship between the surrogate endpoint and the final outcome should be quantified and presented. It should also be incorporated through probabilistic sensitivity analysis and can be further explored in scenario analysis.” and “Extrapolation of a surrogate to final to a different population or technology of a different class or with a different mechanism of action needs thorough justification.”<sup>53</sup>

#### 4 | THE WAY FORWARD?

In the light of the technical developments and healthcare policy considerations outlined above, we make the following observations and suggestions for areas that merit prioritized attention going forward:

1. With the increasing patient and societal pressure for faster access to therapies (as highlighted by the development of vaccines and therapies in response to the COVID-19 pandemic), there is likely to be increased future pressure for more efficient clinical trial design, including use of surrogate endpoints, to inform licensing and reimbursement of new therapies.
2. Avoid the use of accelerated/expedited access approaches based on non-validated surrogates, and limit their use (based on validated surrogate endpoints) to those circumstances where such expedited evaluation is necessary and warranted, for example, diseases/conditions that severely impair and have high unmet treatment needs. The UK Early Access to Medicines Scheme (EAMS)<sup>54</sup> is one such framework which targets use through defined eligibility criteria.
3. Given growing consensus regarding statistical methods for the evaluation and validation of surrogate endpoints, increase the regulatory pressure to apply surrogate evaluation frameworks with statistical validation at their heart as part of licensing and reimbursement decision-making, to focus on therapies with validated surrogate endpoints.
4. Consider quantitative decision analytic frameworks to assist reimbursement decision-making that include a transparent quantification of the predicted effect on the clinical outcome and related health outcomes (such as QALYs) of an observed trial-based change in the surrogate. Such models could incorporate surrogate outcome evidence informed by Bayesian multivariate meta-analysis.<sup>51</sup>
5. Increase application of conditional models of licensing and reimbursement where therapies are reassessed using confirmatory trials based on final patient relevant outcomes (and withdrawn if these fail to show predicted benefit and economic outcomes). Conditional assurance may be used to estimate the probability of success in a confirmatory trial where initial approval has been based on a surrogate outcome.<sup>55</sup>
6. Improved reporting of randomized trials using surrogate endpoints is needed including the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013 and CONSORT (Consolidated Standards of Reporting Trials) 2010 statements. The ongoing work on extensions to these documents (SPIRIT-SURROGATE and CONSORT-SURROGATE) provides a timely opportunity to engage researchers, regulators and other key stakeholders in the challenges of the use of surrogates in clinical trials.<sup>56</sup>

## 5 | CONCLUDING REMARKS

Since the initial forays into surrogate endpoint evaluation over 30 years ago, development of multiple methodologies has rapidly taken place and continues apace. As understanding of how to interpret the outputs generated by these methods grows, we must now move to incorporate them formally as a quantitative input to licensing and reimbursement decision-making. A Bayesian approach to this would provide a natural framework through which to take full account of the uncertainties inherent where the accumulated evidence incorporates surrogate outcome data.

### CONFLICT OF INTEREST

The authors have no interests to declare.

### DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### ORCID

Christopher J. Weir  <https://orcid.org/0000-0002-6494-4903>

### REFERENCES

1. Grieve AP. 25 years of Bayesian methods in the pharmaceutical industry: a personal, statistical bummel. *Pharm Stat.* 2007;6(4):261-281.
2. Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med.* 2017;15:134. doi:10.1186/s12916-017-0902-9
3. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001;69:89-95. doi:10.1067/mcp.2001.113989
4. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to review the role of surrogate endpoints in health policy: state of the art and the way forward. *Value Health.* 2017;20:487-495.
5. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA.* 1999;282:790-795.
6. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* 1989;8:431-440.
7. Ensor H, Lee RJ, Sudlow C, Weir CJ. Statistical approaches for evaluating surrogate outcomes in clinical trials: a systematic review. *J Biopharm Stat.* 2016;26(5):859-879.
8. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med.* 1997;16:1965-1982.
9. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics.* 2002;58:21-29.
10. Taylor JMG, Wang Y, Thiebaut R. Counterfactual links to the proportion of treatment effect explained by a surrogate marker. *Biometrics.* 2005;61:1102-1111.
11. Qu YM, Case M. Quantifying the indirect treatment effect via surrogate markers. *Stat Med.* 2006;25:223-231.
12. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics.* 2000;1:49-67.
13. Alonso A, Molenberghs G. Surrogate marker evaluation from an information theory perspective. *Biometrics.* 2007;63:180-186.
14. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics.* 2009;65:530-538.
15. Gilbert PB, Qin L, Self SG. Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Stat Med.* 2008;27:4758-4778.
16. Buyse M, Molenberghs G, Paoletti X, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biom J.* 2016;58(1):104-132.
17. Conlon A, Taylor J, Li Y, Diaz-Ordaz K, Elliott M. Links between causal effects and causal association for surrogacy evaluation in a gaussian setting. *Stat Med.* 2017;36:4243-4265. doi:10.1002/sim.7430
18. Wolfson J, Gilbert P. Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. *Biometrics.* 2010;66:1153-1161.
19. Li Y, Taylor JMG, Elliott MR, Sargent DJ. Causal assessment of surrogacy in a meta-analysis of colorectal cancer trials. *Biostatistics.* 2011;12:478-492.
20. Weir IR, Rider JR, Trinquart L. Counterfactual mediation analysis in the multistate model framework for surrogate and clinical time-to-event outcomes in randomized controlled trials. *Pharm Stat.* 2021;21:1-13. doi:10.1002/pst.2159
21. Alonso A, Molenberghs G, Geys H, Buyse M, Vangeneugden T. A unifying approach for surrogate marker validation based on Prentice's criteria. *Stat Med.* 2006;25(2):205-221. doi:10.1002/sim.2315
22. Shannon CE, Weaver W. *A Mathematical Theory of Communication.* The Bell System Technical Journal; 1948.
23. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat.* 2006;5:173-186.

24. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. 2002;21:589-624. doi:10.1002/sim.1040
25. Li Y, Taylor JMG, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics*. 2010;66:523-531.
26. Elliott MR, Li Y, Taylor JMG. Accommodating missingness when assessing surrogacy via principal stratification. *Clin Trials*. 2013;10(3):363-377. doi:10.1177/1740774513479522
27. Alonso AA, Van der Elst W, Molenberghs G, et al. On the relationship between the causal-inference and metaanalytic paradigms for the validation of surrogate endpoints. *Biometrics*. 2015;71:15-24.
28. Alonso AA, Van der Elst W, Meyvisch P. Assessing a surrogate predictive value: a causal inference approach. *Stat Med*. 2017;36:1083-1098.
29. Alonso A, Van der Elst W, Molenberghs G. A maximum entropy approach for the evaluation of surrogate endpoints based on causal inference. *Stat Med*. 2018;37:4525-4538. doi:10.1002/sim.7939
30. Alonso A, Van Der Elst W, Molenberghs G, et al. A reflection on the causal interpretation of individual-level surrogacy. *J Biopharm Stat*. 2019;29(3):529-540. doi:10.1080/10543406.2019.1579221
31. Van der Elst W, Alonso A, Coppenolle H, et al. The individual level surrogate threshold effect in a causal-inference setting with normally distributed endpoints. *Pharm Stat*. 2021;20(6):1216-1231. doi:10.1002/pst.2141
32. Cohen J, Cairns C, Paquette C, Faden L. Comparing patient access to pharmaceuticals in the UK and US. *Appl Health Econ Health Policy*. 2006;5:177-187.
33. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Use of surrogate endpoints in healthcare policy: a proposal for adoption of a validation framework. *Nat Rev Drug Discov*. 2016;15:516.
34. Kesselheim AS, Wang B, Franklin JM, Darrow JJ. Trends in utilization of FDA expedited drug development and approval programs, 1987-2014: cohort study. *BMJ Open*. 2015;351:h4633.
35. Lenzer J, Brownlee S. Should regulatory authorities approve drugs based on surrogate endpoints? *BMJ Open*. 2021;374:n2059.
36. Chen EY, Haslam A, Prasad V. FDA acceptance of surrogate endpoints for cancer drug approval: 1992-2019. *JAMA Intern Med*. 2020;180:912-914.
37. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA Intern Med*. 2014;311:368-377.
38. Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613.
39. O'Fee K, Deych E, Ciani O, et al. Assessment of nonfatal myocardial infarction as a surrogate for all-cause and cardiovascular mortality in treatment or prevention of coronary artery disease: a meta-analysis of randomized clinical trials. *JAMA Intern Med*. 2021;181(12):1575-1587. doi:10.1001/jamainternmed.2021.5726
40. United States Food and Drug Administration. accessed 16 January 2022. <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>.
41. Gyawali B, Hey SP, Kesselheim AS. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. *E Clin Med*. 2020;21:100332.
42. Schuster Bruce C, Brhlikova P, Heath J, McGettigan P. 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 Med*. 2019;16:e1002873.
43. Grigore B, Ciani O, Dams F, et al. Surrogate endpoints in health technology assessment: an international review of methodological guidelines. *Pharmacoeconomics*. 2020;38:1055-1070.
44. Joore M, Grimm S, Boonen A, de Wit M, Guillemin F, Fautrel B. Health technology assessment: a framework. *RMD Open*. 2020;6(3):e001289.
45. National Institute for Health and Care Excellence. accessed 17 January 2022, <https://www.nice.org.uk/process/pmg6/chapter/assessing-cost-effectiveness>.
46. Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000-2005. *JAMA Intern Med*. 2006;295:2270-2274.
47. Ciani O, Buyse M, Garside R, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ Open*. 2013;346:f457.
48. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK health technology assessment reports. *Health Technol Assess*. 2009;13(8):1-50.
49. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA, Evidence-Based Medicine Working Group. 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 endpoints. Evidence-based medicine working group. *JAMA*. 1999;282:771-778.
50. Lassere MN. The biomarker-surrogacy evaluation schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res*. 2008;17:303-340.
51. Bujkiewicz S, Achana F, Papanikos T, et al. NICE DSU technical support document 20: multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints, 2019; <http://www.nicedsu.org.uk>.
52. National Institute for Health and Care Excellence. accessed 16 January 2022. <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/changes-to-health-technology-evaluation>.

53. National Institute for Health and Care Excellence. accessed 16 January 2022 <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/chte-methods-consultation>.
54. Medicines and Healthcare products Regulatory Agency. accessed 16 January 2022, <https://www.gov.uk/guidance/apply-for-the-early-access-to-medicines-scheme-eams>.
55. Temple JR, Robertson JR. Conditional assurance: the answer to the questions that should be asked within drug development. *Pharm Stat*. 2021;20:1102-1111.
56. EQUATOR network. Reporting guidelines under development for clinical trials, accessed 29 March 2022 <https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-clinical-trials/>.
57. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med*. 1992;11:167-178.
58. Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med*. 1997;16:1515-1527.
59. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*. 1998;54:1014-1029.
60. Tibaldi F, Abrahantes JC, Molenberghs G, et al. Simplified hierarchical linear models for the evaluation of surrogate endpoints. *J Stat Comput Simul*. 2003;73:643-658.
61. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J R Stat Soc Ser C Appl Stat*. 2001;50(4):405-422.
62. Renfro LA, Shi Q, Sargent DJ, Carlin BP. Bayesian adjusted R2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Stat Med*. 2012;31(8):743-761.
63. Ghosh D, Taylor JMG, Sargent DJ. Meta-analysis for surrogacy: accelerated failure time models and semicompeting risks modeling. *Biometrics*. 2012;68(1):226-232. doi:10.1111/j.1541-0420.2011.01638.x
64. Tibaldi F, Barbosa FT, Molenberghs G. Modelling associations between time-to-event responses in pilot cancer clinical trials using a Plackett-dale model. *Stat Med*. 2004;23:2173-2186.
65. Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Stat Med*. 2001;20:3023-3038.
66. Renard D, Geys H, Molenberghs G, Burzykowski T, Buyse M. Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biom J*. 2002;44:921-935.
67. Alonso A, Geys H, Molenberghs G, Vangeneugden T. Investigating the criterion validity of psychiatric symptom scales using surrogate marker validation methodology. *J Biopharm Stat*. 2002;12:161-178.
68. Alonso A, Geys H, Molenberghs G, Kenward MG, Vangeneugden T. Validation of surrogate markers in multiple randomized clinical trials with repeated measurements. *Biom J*. 2003;45:931-945.
69. Alonso A, Geys H, Molenberghs G, Kenward MG, Vangeneugden T. Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: canonical correlation approach. *Biometrics*. 2004;60:845-853.
70. Pryseley A, Tilahun A, Alonso A, Molenberghs G. Using earlier measures in a longitudinal sequence as a potential surrogate for a later one. *Comput Stat Data Anal*. 2010;54:1342-1354.
71. Renard D, Geys H, Molenberghs G, et al. Validation of a longitudinally measured surrogate marker for a time-to-event endpoint. *J Appl Stat*. 2003;30(2):235-247. doi:10.1080/0266476022000023776
72. Alonso A, Molenberghs G, Burzykowski T, et al. Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics*. 2004;60:724-728.
73. Qu Y, Case M. Quantifying the effect of the surrogate marker by information gain. *Biometrics*. 2007;63:958-962.
74. Miao XP, Wang YC, Gangopadhyay A. An entropy-based nonparametric test for the validation of surrogate endpoints. *Stat Med*. 2012;31:1517-1530.
75. Alonso A, Molenberghs G. Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. *Stat Methods Med Res*. 2008;17:497-504.
76. Pryseley A, Tilahun A, Alonso A, Molenberghs G. Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. *Clin Trials*. 2007;4:587-597.
77. Ensor H, Weir CJ. Evaluation of surrogacy in the multi-trial setting based on information theory: an extension to ordinal outcomes. *J Biopharm Stat*. 2020;30(2):364-376.
78. Ensor H, Weir CJ. Separation and the information theory surrogate evaluation approach: a penalised likelihood solution. *Pharm Stat*. 2021;21:1-14. doi:10.1002/pst.2152
79. Conlon A, Taylor JMG, Elliott MR. Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal. *Biostatistics*. 2014;15:266-283.
80. Zigler CM, Belin TR. A Bayesian approach to improved estimation of causal effect Predictiveness for a principal surrogate endpoint. *Biometrics*. 2012;68(3):922-932.
81. Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. *Biometrics*. 2008;64:1146-1154.
82. Qin L, Gilbert PB, Follmann D, Li D. Assessing surrogate endpoints in vaccine trials with case-cohort sampling and the cox model. *Ann Appl Stat*. 2008;2:386-407.
83. Chen H, Geng Z, Jia JZ. Criteria for surrogate end points. *J R Stat Soc Ser B Stat Methodol*. 2007;69:919-932.
84. Ju CA, Geng Z. Criteria for surrogate end points based on causal distributions. *J R Stat Soc Ser B Stat Methodol*. 2010;72:129-142.



85. Deslandes E, Chevret S. Assessing surrogacy from the joint modelling of multivariate longitudinal data and survival: application to clinical trial data on chronic lymphocytic leukaemia. *Stat Med*. 2007;26:5411-5421.
86. Ghosh D, Elliott MR, Taylor JMG. Links between analysis of surrogate endpoints and endogeneity. *Stat Med*. 2010;29:2869-2879.
87. Emsley R, Dunn G, White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat Methods Med Res*. 2010;19:237-270.
88. Chen SX, Leung DHY, Qin J. Information recovery in a study with surrogate endpoints. *J Am Stat Assoc*. 2003;98:1052-1062.
89. Chen SX, Leung DHY, Qin J. Improving semiparametric estimation by using surrogate data. *J R Stat Soc Ser B Stat Methodol*. 2008;70:803-823.
90. Benda N, Gerlinger C. Sperm count as a surrogate endpoint for male fertility control. *Stat Med*. 2007;26:4905-4913.
91. Baker SG, Sargent DJ, Buyse M, Burzykowski T. Predicting treatment effect from surrogate endpoints and historical trials: an extrapolation involving probabilities of a binary outcome or survival to a specific time. *Biometrics*. 2012;68(1):248-257. doi:[10.1111/j.1541-0420.2011.01646.x](https://doi.org/10.1111/j.1541-0420.2011.01646.x)
92. Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. *Appl Stat*. 2001;50:375-387.
93. Lin H, McCulloch CE, Mayne ST. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med*. 2002;21:2369-2382. doi:[10.1002/sim.1179](https://doi.org/10.1002/sim.1179)
94. Taylor JMG, Wang Y. Surrogate markers and joint models for longitudinal and survival data. *Control Clin Trials*. 2002;23:626-634. doi:[10.1016/S0197-2456\(02\)00234-9](https://doi.org/10.1016/S0197-2456(02)00234-9)

**How to cite this article:** Weir CJ, Taylor RS. Informed decision-making: Statistical methodology for surrogacy evaluation and its role in licensing and reimbursement assessments. *Pharmaceutical Statistics*. 2022;21(4):740-756. doi:[10.1002/pst.2219](https://doi.org/10.1002/pst.2219)