Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward

Oriana Ciani, PhD1,2,*, Marc Buyse, ScD3,4, Michael Drummond, PhD5, Guido Rasi, PhD6,7, Everardo D. Saad, MD8, Rod S. Taylor, PhD2

1Evidence Synthesis and Modelling for Health Improvement, Institute of Health Research, University of Exeter Medical School, Exeter, UK; 2Centre for Research on Health and Social Care Management, Bocconi University, Milan, Italy; 3International Drug Development Institute, Louvain-la-Neuve, Belgium; 4CluePoints, Inc., Cambridge, MA, USA; 5Centre for Health Economics, University of York, York, UK; 6European Medicines Agency, London, UK; 7University “Tor Vergata,” Rome, Italy; 8Dendrix Research, Sao Paulo, Brazil

Abstract

The efficacy of medicines, medical devices, and other health technologies should be proved in trials that assess final patient-relevant outcomes such as survival or morbidity. Market access and coverage decisions are, however, often based on surrogate end points, biomarkers, or intermediate end points, which aim to substitute and predict patient-relevant outcomes that are unavailable because of methodological, financial, or practical constraints. We provide a summary of the present use of surrogate end points in health care policy, discussing the case for and against their adoption and reviewing validation methods. We introduce a three-step framework for policymakers to handle surrogates, which involves establishing the level of evidence, assessing the strength of the association, and quantifying relations between surrogates and final outcomes. Although the use of surrogates can be problematic, they can, when selected and validated appropriately, offer important opportunities for more efficient clinical trials and faster access to new health technologies that benefit patients and health care systems. Keywords: clinical outcome assessment, health technology assessment, surrogate end points, validation.

Copyright © 2017, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Market access and coverage policies for drugs, medical devices, and other health technologies ideally should be based on randomized controlled trials or systematic reviews of randomized controlled trials that assess final outcomes relevant to patients, such as survival, morbidity, and health-related quality of life [1]. Nevertheless, regulatory agencies, including the Food and Drug Administration (FDA) in the United States and the European Medicines Agency (EMA), have a long tradition of licensing technologies solely on the basis of evidence of their effects on biomarkers or intermediate end points that act as so-called surrogate end points (Table 1) [2–4]. The role of surrogates is becoming increasingly important in the context of programs initiated by the FDA and the EMA to offer accelerated approval to promising new medicines. The key rationale for the use of a surrogate end point is to predict the benefits of treatment in the absence of data on patient-relevant final outcomes [5]. Evidence from surrogate end points may not only expedite the regulatory approval of new health technologies but also inform coverage and reimbursement decisions. In the United Kingdom, the National Institute for Health and Care Excellence (NICE) has made several recommendations on the basis of cost-effectiveness analyses that relied entirely on treatment effects derived from clinical trials that assessed surrogate end points [6].

Despite the potential appeal of surrogates, their use remains controversial, because they may not capture the combined benefit-risk profile of a technology and because superiority on a surrogate end point may not translate into benefits for patients, or if it did the health care system may not judge the benefits to be good value for money [7–10]. These limitations can be illustrated by the examples of two surrogate end points used in oncology.

Conflicts of interest: O. Ciani and R. S. Taylor conceived the idea for this article, compiled the evidence, and jointly drafted the manuscript. All authors extensively commented on the draft and approved the final version. M. Buyse declares an association with the International Drug Development Institute, Inc., and CluePoints, Inc., and E. D. Saad with Dendrix, Ltd. All other authors declare no conflicts of interest in relation to this article. The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the agencies or organizations with which the authors are affiliated.

* Address correspondence to: Oriana Ciani, Evidence Synthesis and Modelling for Health Improvement, Institute of Health Research, University of Exeter Medical School, South Cloisters, St Luke’s Campus, Exeter EX1 2LU, UK.

E-mail: o.ciani@exeter.ac.uk.

http://dx.doi.org/10.1016/j.jval.2016.10.011
and considered by FDA, as a licensing body, and NICE, as a reimbursement body, in their decision-making activity.

In May 2003, the FDA approved the tyrosine kinase inhibitor gefitinib for patients with non–small-cell lung cancer on the basis of a favorable effect of the drug on the surrogate end point of the rate of tumor response [11]. The initial approved indication was the treatment of patients who were refractory to established cancer treatments—both a platinum-based regimen and docetaxel [12,13]. Nevertheless, data from two randomized studies of gefitinib versus placebo that showed no significant survival benefit became available in 2005 [14,15], and the FDA consequently released new labeling for gefitinib, which limited its use only to continuation in patients who had already taken the medicine for the disease and whose doctor believed it was helping them [16].

In a second example, the EMA approved the second-generation tyrosine kinase inhibitor dasatinib for treatment of the “chronic phase” of chronic myeloid leukemia (CML) in patients who were newly diagnosed and positive for the Philadelphia chromosome [17]. This approval was based on data from a randomized controlled trial that showed the relative efficacy of dasatinib compared with imatinib on the primary end point of confirmed complete cytogenetic response (CCR, surrogate outcome) by 12 months (e.g., 77% vs. 66%; P = 0.007) [18]. In deciding about approval of new products, however, EMA considers their benefit-risk profile, whereas decisions of health technology assessment (HTA) bodies and payers such as NICE and the Centers for Medicare & Medicaid Services in the United States are based on a broader value-for-money evaluation. When NICE appraised the drug in March 2012, it concluded that first-line use of dasatinib for the treatment of CML represented poor value for money. In a situation in which clinical effectiveness information was available either in terms of biomarker end points or as immature data on overall survival, the evidence review group systematically looked for evidence supporting the adoption of CCR at 12 months as reliable predictors of overall survival by looking at data of patients treated with tyrosine kinase inhibitor, naive to previous pharmacological therapies for CML. Historical data of midterm survival (i.e., up to 7 years since the start of the treatment), conditional to achievement of CCR at 12 months post-treatment, were identified and used to predict and extrapolate long-term survival curves for the dasatinib-treated cohort of patients. The analyses showed a small estimated incremental gain in survival (final outcome) extrapolated from the observed improvement on CCR (22.7 years vs. 21.3 years) and a patient cost of £30,477 per year, which equated to a cost per quality-adjusted life-year (QALY) of more than £200,000 [19].

Because the issues introduced are likely to intensify in a context of promotion of accelerated approval for medicines, raising greater challenges for those bodies seeking to assess the costs and benefits of new health technologies, in this policy perspective we discuss the case for and against the use of surrogate end points, give an overview of methods to validate the selection of surrogates, and propose a framework for the appropriate use of surrogates by policymakers. Finally, we identify unanswered questions and key areas for future research.

### The Case for Surrogate End Points

Results from surrogate end points generally accrue more quickly than from final end points, thus allowing for clinical trials with shorter follow-up periods and smaller sample sizes [20]. Reducing trial sample size and duration ensures faster patient access to new therapies and it means that trials are also less expensive, which make surrogate end points attractive to manufacturers or research sponsors alike. This efficiency can be illustrated in the setting of cardiovascular disease, for which the most common final patient-relevant end points are mortality and major cardiovascular morbidity (e.g., myocardial infarction, stroke, and hospitalization due to angina). The rates of these final outcomes are, however, typically low, particularly in populations with early-stage cardiovascular disease, thus requiring a definitive trial involving thousands or tens of thousands of patients followed up for several years. In contrast, a trial powered on a surrogate primary end point (e.g., carotid artery intima-media thickness and luminal loss) might involve a few hundred patients followed up for weeks or months [21]. Primary end points are often discrete, whereas surrogates are usually continuous and often repeatedly measured, thus providing more statistical power to detect significant treatment effects [22]. It is, however, important to note that smaller sample sizes restrict the likelihood of

### Table 1 – Outcome and end point definitions.

<table>
<thead>
<tr>
<th>End point</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomarker</td>
<td>Characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention</td>
<td>$H_bA_{1c}$, C-peptide</td>
</tr>
<tr>
<td>Patient-relevant (final) end point</td>
<td>Characteristic or variable that reflects how patients feel or function or how long they survive</td>
<td>Diabetic foot: mortality, health-related quality of life</td>
</tr>
<tr>
<td>Intermediate end point</td>
<td>End point is, or is felt to be, of value to patients but does not represent the ultimate patient-relevant final outcome of interest</td>
<td>Hypoglycemic symptoms</td>
</tr>
<tr>
<td>Surrogate end point</td>
<td>Biomarker or intermediate end point intended to substitute and predict for patient-relevant final end point</td>
<td>$H_bA_{1c}$ and glucose control as surrogate for diabetes complications and mortality</td>
</tr>
</tbody>
</table>

$H_bA_{1c}$, glycated hemoglobin; LDL-cholesterol, low-density lipoprotein cholesterol; SBP, systolic blood pressure.

* Definitions adapted from the Biomarkers Definition Working Group [5].
identifying safety issues (especially if they are rare). It has been stated that “there is no surrogate for safety” [23], meaning that usually long-term observations of the adverse events of interest are needed to fully characterize the safety profile of therapies.

There may also be circumstantial and ethical reasons for the use of surrogate end points in clinical trials of new and emerging treatments—the seriousness of the condition, the availability of alternative therapies, or the difficulty of studying the final end point could influence the acceptability of a surrogate end point [2,23]. At the end of the last century, when the AIDS epidemic was a global concern, hastening decision making about the efficacy of new therapies for HIV infection because of lack of clinical benefit data was criticized and discouraged [24]. Also, in the case of many treatments developed for rare diseases, the use of surrogate end points allows registration clinical trials to achieve the accelerated approval pathway for drug adoption [25].

To date, most of the focus on the use of surrogate end points in health care policymaking has been in the context of licensing or market authorization by centralized agencies, including the FDA and the EMA. In 1996, the FDA introduced subpart H, a special regulatory mechanism for drug development programs, which allows the organization to grant marketing approval for new drugs for which well-controlled clinical trials have shown an effect on a surrogate end point that is “reasonably likely, based on epidemiological, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit” [29]. Subpart H applications are often candidates for the fast-track program—an accelerated approval program to expedite the review of interventions for life-threatening diseases or those with irreversible morbidity, which has been in place since 1992 [2]. A recent review showed that pivotal trials using surrogate end points as their primary end point formed the exclusive basis of FDA approvals for 91 of 206 (44%) indications for novel therapeutic agents between 2005 and 2012 [27]. Surrogate end points were used in virtually all trials of agents approved through the accelerated approval pathway, most of which were for the treatment of cancer, infectious diseases, and metabolic diseases, including cardiovascular disease, diabetes mellitus, and hyperlipidemia. For the EMA, “conditional approval” and “approval under exceptional circumstances” procedures allow marketing authorization to be granted when comprehensive data cannot be provided at the time of the submission [28,29]. Although these procedures refer primarily to situations when data from randomized trials are not available, they also apply when evidence on the final patient-relevant outcome of interest is not available. Nevertheless, the use of intermediate end points that are not the final clinical outcomes of relevance should only be the basis for granting a marketing authorization when they are “agreed” to be surrogates or to be sufficiently informative by the scientific and regulatory community.

Another reason why surrogate end points may be preferable in the registration of randomized controlled trials is linked to the possibility of crossover among trials’ arms that could bias the treatment effect observed on the final outcome (e.g., overall survival). In case of crossover, surrogate end points might be preferred to final end points to establish the activity of antitumoral agents. Nevertheless, approaches on how to technically handle treatment switching bias have recently been reported in the literature [30] and, from a payer’s perspective, it is important to consider that a treatment effect that would reflect the sequence of all available treatments would still be of value.

**The Case against Surrogate End Points**

The use of surrogate end points in clinical trials means that policymakers must extrapolate from these findings to estimate the true benefits to patients and health systems, which results in uncertainty about the health and economic value of the health technology in question. For example, for a trial assessing the efficacy of statins on the basis of a reduction in low-density lipoprotein cholesterol, regulatory agencies may want to predict what is the corresponding effect in terms of stroke events prevented, whereas HTA bodies may want to predict what is the corresponding QALY gain in the relevant population. A key pitfall is that surrogate end points do not necessarily provide the same answer as final outcomes on the combined benefit-risk profile of a health technology. As mentioned earlier, reliance on tumor response as a surrogate outcome led to gefitinib initially being licensed for non–small-cell lung cancer, but this drug was later found to have no benefit in terms of overall survival and the license was subsequently restricted. The Cardiac Arrhythmia Suppression Trial is a more commonly cited and more concerning example of surrogate failure. Ventricular arrhythmia was known to be associated with almost 4 times the risk of death related to cardiac complications, particularly sudden death [31,32]. Two drugs, encainide and flecainide, were found to suppress arrhythmias effectively and were approved by the FDA; results from the Cardiac Arrhythmia Suppression Trial, however, later showed that the use of these anti-arrhythmic agents was associated with 2.5 times higher mortality in patients with asymptomatic or mildly symptomatic ventricular arrhythmia after myocardial infarction [33]. Both drugs were later relabeled and became indicated for life-threatening ventricular arrhythmias only [34].

On the basis of this and other examples of surrogate failures for regulatory approvals, in the late 1990s Fleming and DeMets [7] illustrated how the use of surrogate end points might lead to inappropriate conclusions about the benefit-risk profiles of treatments (Fig. 1). Failure of surrogate end points may occur for various reasons, but it is often difficult to determine which of the mechanisms illustrated in Figure 1 might underlie the failure. No clear pattern exists between types of failure and different diseases, as shown by the following examples. A putative surrogate end point may fail because it does not lie in the same pathophysiological process that results in the final end point (Fig. 1A); for example, using prostate biopsies as a surrogate for death from prostate cancer when biopsy detects only latent disease and death is due to aggressive forms of this tumor. Surrogate end points may also fail because the health technology may affect only the pathway mediated through the surrogate end point (Fig. 1B; e.g., when encainide or flecainide was used to suppress cardiac arrhythmias) or only pathways independent of the surrogate end point (Fig. 1C; e.g., when CD4 counts were used as potential surrogate end points for death from HIV infection). In each case, the treatment effect observed on the surrogate end point would capture only part of the whole effect on the final end point. As a result, a false-positive conclusion (i.e., the technology is effective when it is not) or a false-negative conclusion (i.e., the intervention is not effective when it is) may occur on the basis of observations of the surrogate end point. The intervention itself may also affect the final outcomes through unintended, unanticipated, and unrecognized mechanisms of action that operate independently from the disease process (Fig. 1D). For example, rosiglitazone was approved in 1999 by the FDA and in 2000 by the EMA as an oral combination therapy for patients with type 2 diabetes on the basis that it reduced levels of glycated hemoglobin [35], but it was later found to increase significantly the risk of myocardial infarction and mortality [36,37]. Finally, in some situations, the surrogate could lie in the only causal pathway of the disease process and thus would entirely capture the intervention’s effect on the final outcome (Fig. 15); nevertheless, the treatment effects observed on the surrogate end point could still yield misleading information in relation to the magnitude of the effect of the treatment on the final end point, which could be either underestimated or overestimated.
Reliance on trials of a surrogate end point has more often shown to substantially overestimate the treatment effect of health technologies. This problem was reported by Ridker and Torres [38] after reviewing 324 consecutive cardiovascular trials. They observed that trials with primary end points 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). A meta-epidemiological study involving 185 randomized controlled trials that used surrogate end points or patient-relevant outcomes and that were reported in six high-impact general medical journals was specifically designed to confirm or refute this observation by comparing the treatment effects from the trials that used surrogates and those that used final outcomes [39]. This analysis found that trials that used surrogate end points were twice as likely to report positive treatment effects as trials that used 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 final outcomes. This “surrogate end point bias” was not explained by differences between the two groups of trials in terms of the risk of bias or other characteristics.

How Do We Appropriately Select, Validate, and Apply Surrogates?

The potential failure of surrogate end points means that the validity of the relation between the surrogate and the final outcome needs to be established clearly in advance. Several authors have described various statistical methods to validate surrogates [40]. Nevertheless, the complexity of many of these statistical methods means that their uptake in practice is relatively low. We describe here a three-step process to validate and use surrogate-based evidence for use in health care decision making.

Establish the Level of Evidence

The first step in the process is to consider the hierarchy of available evidence. Table 2 presents a three-step framework for health care policymakers to consider the suitability of surrogates, which was proposed in a previous article [6] based on the publication by Bucher et al. [41]. In this framework, 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 final end point across cohorts or at the level of the individual patient. Nevertheless, as Fleming and DeMets [7] noted, “a correlate does not a surrogate make,” and so associations at the level of the individual patient do not directly validate surrogate measures, although they may identify good prognostic markers [42,43]. The highest level of evidence (level 1) therefore relates to evidence showing that technologies that improve the surrogate also improve the final outcome across many randomized controlled trials. Trial-based evidence of a final outcome is usually not available for a new health care technology for which surrogate end points are used, and so this evidence needs to be sourced from other trials of the same or a similar technology—for example, in the case of drugs, trials should be of drugs from within the same class or, if that is not available, a different class. This element highlights the importance of the specificity of the surrogate outcome validity in

**Table 2 – Hierarchy of evidence for surrogate end point validity**

<table>
<thead>
<tr>
<th>Hierarchy level</th>
<th>Requirement</th>
<th>Source of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment effect on surrogate corresponds to treatment effect on final outcome</td>
<td>Randomized controlled trials showing that changes in the surrogate are associated with commensurate changes in the final outcome</td>
</tr>
<tr>
<td>2</td>
<td>Consistent association between surrogate and final outcome</td>
<td>Epidemiological/observational studies</td>
</tr>
<tr>
<td>3</td>
<td>Biological plausibility of relation between surrogate and final outcome</td>
<td>Pathophysiological studies and understanding of the disease process</td>
</tr>
</tbody>
</table>

* Adapted from Elston and Taylor [6] and Bucher et al. [41].
relation to the treatment, to the indication, and to the context of the proposed use [44].

When searching for the evidence supporting the link between a putative surrogate and a related final outcome, it is important to recognize that the surrogacy status of a biomarker is likely to be specific to the context of its use and to the intervention [45]. Several authors have emphasized that the validity of a surrogate end point shown in a particular intervention cannot be assumed to apply to another class, particularly when the two have different off-target effects profiles [41,46,47]. Furthermore, the use of active or inactive control interventions may also influence the surrogate to final outcome relationship [48,49]. Finally, contextual or environmental factors may play a role; for example, a recent evaluation of sputum culture results during treatment as potential surrogate end points for long-term outcome in pulmonary tuberculosis found different results when separate analyses on trials from two geographical regions (i.e., East Africa and East Asia) were performed [50].

Assess the Strength of Association
Having established the level of evidence, the second step is to assess the strength of the association between the surrogate and the final outcome. Among several approaches to address this issue, regression-based and meta-analytic approaches dominate the field [43]. Establishing the strength of an association for level 1 surrogacy usually requires a meta-analysis of all randomized controlled trials on the subject of interest. The most reliable approach is to perform a meta-analysis using patient-level data from all randomized controlled trials of this treatment [51]. When patient-level data are available, two levels of association can be estimated: the association between the surrogate and the final outcome, and the association between the effect of treatment (drug or technology) on the surrogate and on the final outcome [52]. For example, the strength of the association between treatment effects on the surrogate (e.g., mean difference) and the final (e.g., log odds ratio) outcome is usually quantified through the correlation coefficient or its square (called the coefficient of determination), both of which range from 0 to 1.0. Thresholds set to identify good surrogates can be as high as 0.8 for correlation coefficients (r or ρ) or 0.65 for coefficients of determination (R²) [53], which are particularly strict rules for the acceptability of putative surrogate end points when applied in practice. Although ideally level 1 evidence should be used to establish the surrogacy status of a biomarker, it may be that only level 2 evidence is available [19,54]. If so, policymakers should take into account the greater uncertainty with observational evidence in making their decisions.

Quantify the Relation between the Surrogate and the Final Outcome
The final step relates to predicting and quantifying the relation between the surrogate and the final outcome, and between the observed effect on the surrogate and the expected effect on the final outcome. For cost-effectiveness analyses, these would be the expected impact on QALYs. A quantitative approach has been proposed to support this objective, using an extension of the meta-analytic approach to surrogate assessment. The approach consists of estimating the “surrogate threshold effect,” which is the magnitude of treatment effect on the surrogate that would predict a statistically significant treatment effect, or with appropriate extension a clinically meaningful effect, on the final outcome (see Technical Appendix A in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.10.011) [55]. Estimating the expected effect on the final outcome is useful to decide whether a surrogate end point is of practical interest. This is the case, for example, of progression-free survival, which is the time elapsed between randomization (or treatment initiation) and documented tumor progression or death. Tumor progression is preferably determined by radiographic evidence, but in some cases worsening of symptoms and signs of disease may also constitute evidence of progression. In gastric cancer, progression-free survival is not an acceptable surrogate for overall survival in advanced disease; indeed, the surrogate threshold effect suggests that only very large effects on progression-free survival are likely to predict significant effects on overall survival [56]. But then, the use of progression-free survival as a surrogate end point may be limited by the fact that progression may have different implications in different settings, according to whether it is symptomatic, whether salvage therapy is available, and whether its occurrence heralds imminent death. Estimating the expected effect on the final outcome is also crucial for decisions on coverage and reimbursement for health technologies. Regulators are usually focused on early evidence of safety and efficacy to determine whether the balance of benefits and risk is positive when informing the design of registration trials, whereas reimbursement agencies usually consider long-term effectiveness or cost-effectiveness. In this step, therefore, not only are the direction and significance of the treatment effect important, but also is its magnitude [57]. Decisions around market access and reimbursement are normally based on an assessment of the incremental value of the technology in question for the final outcome relative to that for the existing usual or standard of care. In many settings, including the United Kingdom, Sweden, Australia, and Canada, an assessment of value is formalized in a cost-effectiveness analysis [58]. It has been estimated that 27% to 50% of all submissions to NICE in the United Kingdom, to the Pharmaceutical Benefits Advisory Committee in Australia, and to the Common Drug Review in Canada are based on surrogate end points [58]. Whether decisions on market access and reimbursement are based on a formal economic evaluation or on the magnitude of the clinical benefit, the effect of the treatment on the surrogate end point needs to be large enough to predict an improvement in the final outcome (i.e., length or quality of life) before the technology can be concluded to be of value to patients and health care systems. Figure 2 provides a schema showing how to apply this three-step surrogate validation framework to a hypothetical assessment of an antihypertensive drug (e.g., a β-blocker) or a device (e.g., renal denervation therapy). Looking at many trials of antihypertensive drugs, a meta-regression analysis has been conducted to allow conclusions to be drawn about the strength of the association between the reduction in blood pressure (surrogate) and adverse (final) outcome [46]. Although not based on this meta-analysis, previous economic analyses that modeled the reduction in cardiovascular risk from the observed effect of treating blood pressure have shown that antihypertensive drug treatments have an excellent cost-effectiveness profile [59,60].

How Do Present Surrogate End Points Measure Up?
To date, few empirical assessments have investigated the adequacy of evidence for specific surrogate end points or groups of surrogates, particularly in terms of reimbursement policy. A timely example is with the use of sustained virological response (SVR) for the approval of new direct-acting antiviral agents (i.e., boceprevir and telaprevir) in chronic hepatitis C. The FDA has considered SVR at 12 weeks a valid surrogate for a primary end point in clinical trials on the basis of observational cohorts showing only strong correlations between SVR and multiple clinically important outcomes, such as development of hepatocellular carcinoma, end-stage liver complications, and mortality [54]. At present reimbursement agencies lack the necessary evidence to confirm the relationship between treatment effects
have been published. On the basis of a systematic review of these results of a number of meta-analyses of randomized controlled studies and against the use of surrogate end points. Although we have shown that the use of surrogates for health care policymaking can be problematic, we also argue that when selected and used appropriately, surrogates can be useful in decision making. Specifically, surrogates can help to reduce the uncertainty associated with evaluating the effectiveness of new treatments by focusing on earlier endpoints that are more likely to reflect the clinical benefit of the intervention. However, it is important to note that surrogates should be used with caution and only if they have been appropriately validated according to the criteria established by Prasad et al. [66].

Fig. 2 – Framework for the evaluation of surrogate end points in an HTA setting. HRQOL, health-related quality of life; HTA, health technology assessment.

Unanswered Questions and Areas for Future Research

In this review, we have sought to bring together key evidence for and against the use of surrogate end points. Although we have shown that the use of surrogates for health care policymaking can be problematic, we also argue that when selected and used appropriately, surrogates can be useful in decision making. Specifically, surrogates can help to reduce the uncertainty associated with evaluating the effectiveness of new treatments by focusing on earlier endpoints that are more likely to reflect the clinical benefit of the intervention. However, it is important to note that surrogates should be used with caution and only if they have been appropriately validated according to the criteria established by Prasad et al. [66].
appropriately, they can not only offer important opportunities for the more efficient conduct of clinical trials but also allow faster access to new health technologies, which would benefit patients and health care systems. On the basis of knowledge of epidemiology, clinical trials, and statistical methods, we have outlined a three-step framework to guide the selection of surrogate end points by policymakers. Some unanswered questions, however, remain, and these present opportunities for further methodological research to inform future handling of surrogates.

**Do we need individual patient-level data to fully validate surrogates?**

For practical reasons, trial-level (or aggregate) data are the usual evidence available for policymaking around coverage and reimbursement given the costs and constraints of gathering multitrtrial data sets, some of which belong to corporate sponsors. Trial-level data are, however, inadequate to perform criterion standard analyses of surrogacy [52,55]. Published reports may provide unreliable or biased summary statistics (e.g., results on a nonintervention-to-treat population, or reporting of odds ratio rather than hazard ratio for time-to-event data). Most aggregate-data analyses fail to take into account estimation error (i.e., a finite number of finite-sample-size trials in the meta-analytic data) by using "errors-in-variables" regression models, or use only simple linear regression models, irrespective of the functional form of the data. Access to patient-level data is desirable but should be further encouraged by data-sharing and data transparency initiatives led by both the FDA or EMA and the pharmaceutical industry [72–74].

**How can we improve the quality of the future evidence base for surrogates?**

As outlined earlier, validation of a surrogate end point is rather a chicken-and-egg problem, in that it requires evidence of the relation between the surrogate and the final outcome. Because the final outcome for a specific new technology will not yet be available—and hence the surrogate is used as replacement—trials of older technologies in the same technology class are often used to provide that evidence. Nevertheless, moving forward, there is a unique opportunity to link the coverage and licensing of technologies on the basis of surrogate end points to a conditional decision [75,76] that incentivizes extension of follow-up to accrue the relevant evidence on final outcomes to inform later and more definitive decisions on coverage. In-depth knowledge of the natural history of diseases combined with analyses on existing baseline data, emerging large data networks, or past trials helps to identify surrogate end points through a process that will enable the fast developing framework of medicine’s adaptive pathways to patients [77,78].

**How can we improve the uptake of surrogate validation frameworks into policymaking?**

We have described a three-stage framework for the use of surrogates in policymaking, and a small number of existing tools [68,79] are available to guide the use of surrogates. At present however, little or no evidence is available around their acceptability and the likelihood of their uptake by key stakeholders in the health care policymaking process, including payers, patient groups, industry, and clinicians [80]. Heterogeneous approaches to validation of surrogates also lead to decisions on reimbursement that often vary across agencies and diseases, which poses an issue of equity of access for patients across different jurisdictions [81]. Rocchi et al. [81] compared acceptance of surrogates, from glycated hemoglobin $A_1c$ to SVR, for the approval of 13 drugs across seven international regulatory and HTA authorities showing high variability in consideration and assessment of surrogate end points. An analogue comparison was performed across three major EU countries on six products approved in type 2 diabetes, hepatitis C, and oncology [82]. The analysis concluded that IQWIG was more often inclined to recognize no or nonquantifiable benefit for the medicines under assessment compared with the French Haute Autorité de Santé, where request of prospective or observational additional studies is likely, or NICE and the Scottish Medicines Consortium, where, however, low surrogate to final outcomes associations lead to highly uncertain incremental cost-effectiveness ratios. Research is therefore urgently needed to better understand the barriers to implementation of surrogate evaluation tools and harmonization of different approaches, particularly across the present licensing and reimbursement divide.

**How can we improve the use and reporting of surrogates in clinical trials?**

There are also important implications for use and reporting of surrogates for the clinical trials community. A review of randomized controlled trials published in high-impact general medical journals showed that 43% of authors do not explicitly state that their primary outcome was a surrogate end point and only a third of publications discuss the potential limitations of the surrogate end point or evidence about its validity [83]. Reports of clinical trials should therefore state whether their collected outcomes are surrogate end points and provide a clear rationale for the selection of these surrogates, including reference to biological plausibility and evidence of validation. It has been suggested that guidance on surrogates should be incorporated into the present Consolidated Standards of Reporting Trials statement [83].

**Conclusions**

The potential for surrogate end points to have an impact on health care policy and the consequent diffusion of technologies into practice is illustrated by the fact that the primary outcome of more than 40% of pivotal trials used as the basis for approval of new indications is a surrogate that aims to substitute for and predict a final patient-relevant outcome [27]. In the case of specific diseases, such as oncology, this proportion increases to two-thirds of all trials. In terms of reimbursement decisions, a substantial proportion of all submissions to the main HTA agencies are based on surrogate end points [58]. With increasing societal pressure for faster access to therapies, the use of surrogates in health care policy is likely to increase. The recent data-sharing initiatives will greatly facilitate the evaluation of surrogates using patient-level data from industry-sponsored randomized clinical trials.

Relying on surrogates rather than final patient-relevant outcomes increases the uncertainty when making decisions about licensing and coverage of health care technologies. Furthermore, using putative surrogates that are not validated may raise serious ethical concerns. Surrogates can result in market access for technologies that turn out to offer no true health benefit—or even harm—to patients and can result in overestimation of treatment effects, which can lead to inappropriate decisions on coverage. The use of appropriately validated surrogate end points, however, provides an unmissable opportunity to speed up access to innovative technologies that offer important benefits for patients and health care systems and to improve efficiency within the research and development environment.
Acknowledgments

We thank Jemma Lough for providing assistance in the technical editing of the manuscript.

Source of financial support: O. Ciani is funded by a postdoctoral research fellowship by the University of Exeter Medical School. The funder had no role in the collection and interpretation of data, in writing the report, or in the decision to submit the manuscript for publication.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2016.10.011 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES


60 Gluud C, Brok J, Gong Y, Knoppers BM. Data sharing, year 1—data sharing, year 2. BMJ 2014;349:g6332.


