Article:
Ciani, O., Buyse, M., Drummond, M.F. orcid.org/0000-0002-6126-0944 et al. (3 more authors) (2017) Time to review the role of surrogate endpoints in health policy: state of the art and the way forward. Value in Health. pp. 487-495. ISSN: 1524-4733

https://doi.org/10.1016/j.jval.2016.10.011

Time to review the role of surrogate endpoints in health policy

state of the art and the way forward

Oriana Ciani, PhD;[1,2] Marc Buyse, ScD;[3,4] Michael Drummond, PhD;[5] Guido Rasi, PhD;[6,7] Everardo D Saad, MD;[8] Rod S Taylor, PhD[1]

[1]Evidence synthesis and modelling for health improvement, Institute of Health Research, University of Exeter Medical School, Exeter, UK

[2]Centre for Research on Health and Social Care Management, Bocconi University, Milan, Italy

[3]International Drug Development Institute, Louvain-la-Neuve, Belgium

[4]CluePoints Inc., Cambridge, USA

[5]Centre for Health Economics, University of York, York, UK (Michael Drummond, PhD)

[6]European Medicine Agency, London, UK

[7]University "Tor Vergata", Rome, Italy

[8]Dendrix Research, Sao Paulo, Brazil

Keywords: clinical outcome assessment, surrogate endpoints, validation, health technology assessment

Correspondence to: Dr Oriana Ciani, Institute of Health Research, University of Exeter Medical School, South Cloisters, St Luke's Campus, Exeter EX1 2LU, UK

o.ciani@exeter.ac.uk

Phone: + 39 3496658949

Fax: +39 0258362598

Declaration of interests

MB declares an association with the International Drug Development Institute Inc. and CluePoints Inc.; EDS with Dendrix Ltd. All other authors declare no conflict of interest in relation to this manuscript. The views expressed in this article are the personal views of the author(s) 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.

Acknowledgements

OC is funded by a postdoctoral research fellowship by the University of Exeter Medical School. The funder had no role in collection of data, interpretation of data, writing the report, or the decision to submit the paper for publication. Jemma Lough provided technical editing of the manuscript.

Abstract: 152 words

Main text: 4,500 words (includes placeholders for figures and tables)

Tables: 2

Figures: 2

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. However, market access and coverage decisions are often based on surrogate endpoints, biomarkers, or intermediate endpoints, which aim to substitute and predict patient-relevant outcomes that are unavailable due to methodological, financial, or practical constraints. We provide a summary of the current use of surrogate endpoints in healthcare policy, discussing the case for and against their adoption and reviewing validation methods. We introduce a three-step framework for policy makers 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 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 healthcare systems.

Introduction

Market access and coverage policies for drugs, medical devices, and other health technologies ideally should be based on randomised controlled trials or systematic reviews of randomised 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 based solely on evidence of their effects on biomarkers or intermediate endpoints that act as so-called surrogate endpoints (Table 1).[2-4] The role of surrogates is becoming increasingly important in the context of programmes initiated by the FDA and EMA to offer accelerated approval to promising new medicines. The key rationale for the use of a surrogate endpoint is to predict the benefits of treatment in the absence of data on patient-relevant final outcomes.[5] Evidence from surrogate endpoints may not only expedite the regulatory approval of new health technologies but also inform coverage and reimbursement decisions. In the United Kingdom (UK), the National Institute for Health and Care Excellence (NICE) has made several recommendations based on cost-effectiveness analyses that relied entirely on treatment effects derived from clinical trials that assessed surrogate endpoints.[6]

<<table 1 near here>>

Despite their potential appeal, the use of surrogates remains controversial, because they may not capture the combined risk-benefit profile of a technology and because superiority on a surrogate endpoint may not translate into benefits for patients, or if it did the healthcare 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 endpoints used in oncology 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 based on a favourable effect of the drug on the surrogate endpoint of the rate of tumour 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] However, data from two randomised studies of gefitinib vs placebo that showed no significant survival benefit

became available in 2005,[14,15] and the FDA consequently released new labelling 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 leukaemia (CML) in patients who were newly diagnosed and positive for the Philadelphia chromosome.[17] This approval was based on data from a randomised controlled trial that showed the relative efficacy of dasatinib compared with imatinib on the primary endpoint of confirmed complete cytogenetic response (CCR, surrogate outcome) by 12 months (e.g. 77% vs 66%, p=0·007).[18] However, in deciding about approval of new products, EMA considers their benefit/risk profile, while decisions of Health Technology Assessment (HTA) bodies and payers such as NICE and the Centers for Medicare and 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 chronic myeloid leukaemia represented poor value for money. In a situation where clinical effectiveness information was only available either in terms of biomarker endpoints 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 tyrosine kinase inhibitor (TKI)-treated patients data, naïve to previous pharmacological therapies for CML. Historical data of mid-term 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 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]

As 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 endpoints, give an overview of methods to validate the selection of surrogates, and propose a framework for the appropriate use of surrogates by policy makers. Finally, we identify unanswered questions and key areas for future research.

The case for surrogate endpoints

Results from surrogate endpoints generally accrue more quickly than final endpoints, 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 means that trials are also less expensive, which make surrogate endpoints attractive to manufacturers or research sponsors alike. This efficiency can be illustrated in the setting of cardiovascular disease, where the most common final patient-relevant endpoints are mortality and major cardiovascular morbidity (e.g. myocardial infarction, stroke, and hospitalisation due to angina). However, the rates of these final outcomes are 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 endpoint (e.g. carotid artery intima-media thickness, luminal loss) might involve a few hundred patients followed up for weeks or months.[21] Primary endpoints are often discrete whereas surrogates are usually continuous and often repeatedly measured, thus providing more statistical power to detect significant treatment effects.[22] However, it is important to note that smaller sample sizes restrict the likelihood of 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 characterise the safety profile of therapies.
There may also be circumstantial and ethical reasons for the use of surrogate endpoints 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 endpoint could influence the acceptability of a surrogate endpoint.[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 criticised and discouraged.[24] Also, in the case of many treatments developed for rare diseases, the use of surrogate endpoints allow registration clinical trials to achieve the accelerated approval pathway for drug adoption.[25]

To date, most of the focus on the use of surrogate endpoints in healthcare policy making has been in the context of licensing or market authorisation by centralised agencies, including the FDA and EMA. In 1996, the FDA introduced Subpart H, a special regulatory mechanism for drug development

programmes, which allows the organisation to grant marketing approval for new drugs for which well-controlled clinical trials have shown an effect on a surrogate endpoint that is "reasonably likely, based on epidemiological, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit".[26] Applications under Subpart H applications are often candidates for the fast-track programme—an accelerated approval programme 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 endpoints 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.[27] Surrogate endpoints 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 hyperlipidaemia. For the EMA, "conditional approval" and "approval under exceptional circumstances" procedures allow marketing authorisation 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 randomised trials are not available, they also apply when evidence on the final patient-relevant outcome of interest is not available. However, the use of intermediate endpoints 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 endpoints may be preferable in registration RCTs is linked to the possibility of cross-over among trials' arms that could bias the treatment effect observed on the final outcome (e.g. overall survival). In case of cross-over, surrogate endpoints might be preferred to final endpoints to establish the activity of anti-tumoral agents. However, approaches on how to technically handle treatment switching bias have been recently 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.

3. The case against surrogate endpoints

The use of surrogate endpoints in clinical trials means that policy makers 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 based on a reduction in LDL-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 endpoints do not necessarily provide the same answer as final outcomes on the combined risk-benefit profile of a health technology. As mentioned earlier, reliance on tumour 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 licence was subsequently restricted. The Cardiac Arrhythmia Suppression Trial (CAST) is a more commonly cited and more concerning example of surrogate failure. Ventricular arrhythmia was known to be associated with almost four 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; however, results from CAST 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 re-labelled and became indicated for life-threatening ventricular arrhythmias only.[34]

Based on this and other examples of surrogate failures for regulatory approvals, in the late 1990s Fleming and De Mets illustrated how the use of surrogate endpoints might lead to inappropriate conclusions about the risk-benefit profiles of treatments (figure 1).[7] Failure of surrogate endpoints 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 endpoint may fail because it does not lie in the same pathophysiological process that results in the final endpoint (figure 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 tumour. Surrogate endpoints may also fail because the health technology may affect only the pathway mediated through the surrogate endpoint (figure 1b; for example, when encainide or flecainide were used to supress cardiac arrhythmias) or only pathways independent of the surrogate endpoint (figure 1c; for example, when CD4 counts were used as potential surrogate endpoints for death from HIV infection). In each case, the treatment effect observed on the surrogate endpoint would capture only part of the whole effect on

the final endpoint. As a result, a false-positive conclusion (ie, the technology is effective when it is not) or a false-negative conclusion (ie, the intervention is not effective when it is) may occur based on observations of the surrogate endpoint. The intervention itself may also affect the final outcomes through unintended, unanticipated, and unrecognised mechanisms of action that operate independently from the disease process (figure 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 haemoglobin,[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 (figure 1e); however, the treatment effects observed on the surrogate endpoint could still yield misleading information in relation to the magnitude of the effect of the treatment on the final endpoint, that could be either underestimated or overestimated.

<<figure 1 near here>>

Reliance on trials of a surrogate endpoint has more often shown to substantially overestimate the treatment effect of health technologies. This problem was reported by Ridker and Torres after reviewing 324 consecutive cardiovascular trials.[38] They observed that trials with primary endpoints that were surrogates were more likely to report a positive treatment effect (77/115 trials, 67%) than trials that reported final patient-relevant primary outcomes (113/209 trials, 54%) (p=0·02). A meta-epidemiological study involving 185 randomised controlled trials that used surrogate endpoints 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 endpoints were twice as likely to report positive treatment effects as trials that reported final outcomes (52/84 trials, 62% *vs* 37/101 trials, 37%, p<0·01). Furthermore, trials that used surrogates found treatment effects that were, on average, 28–48% larger than trials that used corresponding final outcomes. This "surrogate endpoint 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 endpoints 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 a variety of statistical methods to validate surrogates.[40] However, the complexity of many of these statistical methods means that their uptake in practice is relatively low. Below we describe a three-step process to validate and use surrogate-based evidence for use in healthcare decision making.

1. Establish the level of evidence—The first step in the process is to consider the hierarchy of available evidence. Table 2 shows a three-step framework for healthcare policy makers to consider the suitability of surrogates, which was proposed by a previous article[6] based on the publication by Bucher and colleagues.[41] In this framework, level 3 evidence for a surrogate is based on biological plausibility alone, while evidence is considered to be level 2 when a strong association exists between the surrogate and the final endpoint across cohorts or at the level of the individual patient. However, as Fleming and DeMets noted, "a correlate does not a surrogate make",[7] 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 randomised controlled trials. Trial-based evidence of a final outcome is usually not available for a new healthcare technology for which surrogate endpoints are used, 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 is available, a different class. This element highlights the importance of the specificity of the surrogate outcome validity, in relation to the treatment, to the indication and to the context of the proposed use.[44]

<<table 2 near here>>

When searching for the evidence supporting the link between a putative surrogate and a related final outcome, it important to recognise 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 emphasised that the validity of a

surrogate endpoint 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 endpoints for long-term outcome in pulmonary tuberculosis found different results when separate analysis on trials from two geographical regions (i.e., East Africa and East Asia) were performed.[50]

2. 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 randomised controlled trials on the subject of interest. The most reliable approach is to perform a meta-analysis using patient-level data from all randomized 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 the final outcome.[52] For example, the strength of the association between treatment effects on the surrogate (e.g. mean difference) and 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 (R^2),[53] which are particularly strict rules for the acceptability of putative surrogate endpoints when applied in practice. Whilst 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, policy makers should take into account the greater uncertainty with observational evidence in making their decisions.

3. 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).[55] Estimating the expected effect on the final outcome is useful to decide whether a surrogate endpoint is of practical interest or not. This is the case, e.g., of progression-free survival, that is the time elapsed between randomization (or treatment initiation) and documented tumour progression or death. Tumour 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] On the other hand, the use of progression-free survival as a surrogate endpoint 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 if the balance of benefits and risk is positive when informing the design of registration trials, while reimbursement agencies usually consider long-term effectiveness or cost-effectiveness. In this step, therefore, not only is the direction and significance of the treatment effect important, but also 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 UK, Sweden, Australia, and Canada, an assessment of value is formalised in a cost-effectiveness analysis.[58] It has been estimated that 27% to 50% of all submissions to NICE in the UK, the Pharmaceutical Benefits Advisory Committee in Australia, and the Common Drug Review in Canada are based on surrogate endpoints.[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 endpoint 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 healthcare 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 beta-blocker) or device (e.g. renal denervation therapy). Looking at many trials of antihypertensive drugs, a meta-regression analysis has been undertaken 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 modelled 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]

<<figure 2 near here>>

How do current surrogate endpoints measure up?

To date, few empirical assessments have investigated the adequacy of evidence for specific surrogate endpoints 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 endpoint in clinical trials based on observational cohorts only showing strong correlations between SVR and multiple clinically important outcomes, such as development of hepatocellular carcinoma, end-stage liver complications, and mortality.[54] Reimbursement agencies currently lack the necessary evidence to confirm the relationship between treatment effects observed on SVR and final outcomes in chronic hepatitis C. [61] On the other end, oncology has a long tradition of using surrogates.[62-65] Two frequently used surrogates for overall survival are progression-free survival and time to progression. Progression-free survival is a composite endpoint for which the events of interest are documented tumour progression or death. For time to progression, patients with no prior documentation of disease progression are censored at the time of death. In the past decade, results of a number of meta-analyses of randomised controlled trials quantifying the statistical association between progression-free survival or time to progression

and overall survival in cancer have been published. Based on a systematic review of these meta-analyses, Ciani and colleagues recently sought to assess the suitability of and quality of evidence for progression-free survival and time to progression as surrogates for advanced solid tumours.[53] Although the highest evidence was level 2 in less prevalent diseases such as brain tumour, level 1 evidence had consistently supported some level of association between effects of treatment on both progression-free survival and time to progression compared with overall survival in the four most frequently evaluated advanced solid tumours: colorectal cancer, non-small-cell lung cancer, breast cancer, and ovarian cancer. However, with few exceptions, the strength of the associations between the surrogates and final outcome tended to be relatively low, with $\rho<0{\cdot}7$ across tumour types. This finding of low levels of associations between surrogate endpoints and survival in oncology was also confirmed by Prasad and colleagues.[66]

Even if the evidence is considered to be level 1, a low strength of association between the surrogate and the final outcome can have important implications for decision making. Two specific tools have recently been developed to operationalise components of the three-stage validation framework described above and to help policy makers make appropriate choices when selecting surrogate endpoints. The first of these tools, the Biomarker Surrogacy Evaluation Schema (BSES3),[46] is based on a validation scheme originally developed by clinicians and researchers working in rheumatology.[67] The second tool, another framework for the validation of surrogate endpoints in oncology, was published in 2011 by the Institute of Quality and Efficiency in Healthcare (IQWiG)—a health technology assessment agency that assesses the benefits and harms of drug and non-drug technologies on behalf of the German Federal Joint Committee and Federal Ministry of Health.[68] For a surrogate to be deemed acceptable for reimbursement, both the IQWiG and BSES3 tools require a high level of association with the final outcome (i.e. $\rho\geq0{\cdot}85$, $R^2\geq0{\cdot}60$). Accordingly, only progression-free survival in metastatic colorectal and ovarian cancer treated with cytotoxic agents would be judged to have achieved acceptable evidence of surrogacy.

That progression-free survival and time to progression have been shown to perform poorly as surrogates in oncology does highlight a more general current scepticism by payers about using surrogates in their coverage decisions.[69] This view was reflected in the guideline “Endpoints used in relative effectiveness assessment of pharmaceuticals” published by the European Network for Health

Technology Assessment (EUnetHTA) in February 2013,[70] which recommended that European health technology appraisal agencies should be very cautious about using surrogates and should use them only if they have been appropriately validated according to approaches recognised for evaluation of surrogates. This caution was echoed in NICE's most recent update of guidance on methods for manufacturers to which our work contributed. This guidance now requires a review of the evidence to support the validity of the surrogate endpoint, a clear description of how the relationship between the surrogate and final outcome has been quantified in an economic model, and a full exploration of the additional uncertainty associated with using surrogates to predict cost-effectiveness (e.g. a probabilistic sensitivity analysis that explicitly considers the uncertainty of the predicted treatment effect on the final outcome given the observed treatment effect on the surrogate).[71] It is important to note that where cost-effectiveness informs reimbursement decisions, the problem of using surrogates in economic models is often only one aspect of missing or incomplete data. For example, the collection of a final outcome, but only in the short term, requires extrapolation to assess the long-term effect on the final outcome which has its own pros and cons. Hence, the framework presented here represents only one aspect of what is often a chain of evidentiary uncertainties that need to be considered to make sense of the available evidence and estimate the incremental cost-effectiveness ratio.

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 endpoints. Although we have shown that the use of surrogates for healthcare policy making can be problematic, we also argue that, when selected and used 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 healthcare systems. Based on knowledge of epidemiology, clinical trials, and statistical methods, we have outlined a three-step framework to guide the selection of surrogate endpoints by policy makers. However, some unanswered questions 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 policy making around coverage and reimbursement given the costs and constraints of gathering multi-trial datasets-some of which belong to corporate sponsors. However, trial-level data are inadequate to perform gold standard analyses of surrogacy.[52,55] Published reports may provide unreliable or biased summary statistics (e.g. results on a non-intervention-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 European Medicines Agency and the pharmaceutical industry.[72-74]

How can we improve the quality of the future evidence base for surrogates?—As outlined above, validation of a surrogate endpoint is rather a chicken-and-egg problem, in that it requires evidence of the relation between the surrogate and the final outcome. As 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. However, moving forward, there is a unique opportunity to link the coverage and licensing of technologies based on surrogate endpoints to a conditional decision[75,76] that incentivises extension of follow-up in order 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 endpoints, through a process that will enable the fast developing framework of medicine's adaptive pathways to patients (MAPPs).[77,78]

How can we improve the uptake of surrogate validation frameworks into policy making?—We have described a three-stage framework for the use of surrogates in policy making, and a small number of existing tools[68,79] are available to guide the use of surrogates. However, currently little or no evidence is available around their acceptability and the likelihood of their uptake by key stakeholders in the

healthcare policy-making 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. compared acceptance of surrogates, from hemoglobin A1c to sustained virological response, for the approval of thirteen drugs across seven international regulatory and HTA authorities showing high variability in consideration and assessment of surrogate endpoints. 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 non quantifiable benefit for the medicines under assessment compared to 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 harmonisation of different approaches, particularly across the current 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 randomised 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 endpoint and only a third of publications discuss the potential limitations of the surrogate endpoint or evidence about its validity.[83] Reports of clinical trials should therefore state whether their collected outcomes are surrogate endpoints 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 current Consolidated Standards of Reporting Trials (CONSORT) statement.[83]

Conclusions

The potential for surrogate endpoints to impact on healthcare policy and the consequent diffusion of technologies into practice is illustrated by the fact 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 endpoints.[58] With increasing societal pressure for faster access to therapies, the use of surrogates in healthcare 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 healthcare 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. However, the use of appropriately validated surrogate endpoints provides an unmissable opportunity to speed up access to innovative technologies that offer important benefits for patients and healthcare systems and to improve efficiency within the research and development environment.

Contributors

OC and RST conceived the idea for the paper, compiled the evidence, and jointly drafted the manuscript. All authors extensively commented on the draft and approved the final version.

References

1 Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998; **2**: i-iv, 1-74.

2 Chakravarty A. Regulatory aspects in using surrogate markers in clinical trials. In: Burzykowski T, Molenberghs G, Buyse M, eds. The evaluation of surrogate endpoints. New York: Springer Science+Business Media, 2005.

3 Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)* 2005; **24**: 67-78.

4 Lathia CD, Amakye D, Dai W, et al. The value, qualification, and regulatory use of surrogate end points in drug development. *Clin Pharmacol Ther* 2009; **86**: 32-43.

5 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69**: 89-95.

6 Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports. *Int J Technol Assess Health Care* 2009; **25**: 6-13.

7 Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; **125**: 605-13.

8 Gotzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care* 1996; **12**: 238-46.

9 Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ* 2011; **343**: d7995-d.

10 Moynihan R. Surrogates under scrutiny: fallible correlations, fatal consequences. *BMJ* 2011; **343**.

11 US Food and Drug Administration. Label and approval history. Iressa. 2003. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Search.Overview&DrugName=IRESSA&CFID=18154657&CFTOKEN=b218e5813ec4050d-EC2283AF-A7EA-AF68-D7A67591BFE7CA5F (accessed 21 November 2012).

12 Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *J Clin Oncol* 2003; **21**: 2237-46.

13 Kris MG, Natale RB, Herbst RS, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* 2003; **290**: 2149-58.

14 Thatcher N, Chang A, Parikh P, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 2005; **366**: 1527-37.

15 Kelly K, Gaspar L, Chansky K, et al. Low incidence of pneumonitis on SWOG 0023: A preliminary analysis of an ongoing phase III trial of concurrent chemoradiotherapy followed by consolidation docetaxel and gefitinib/placebo maintenance in patients with inoperable stage III non-small cell lung cancer. *J Clin Oncol (Meeting Abstracts)* 2005; **23**: 7058.

16 US Food and Drug Administration. Gefitinib (marketed as Iressa) information. 2005. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm110473.htm.

17 EMA. EPAR summary for the public: Sprycel. 2014. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Summary_for_the_public/human/000709/WC500056993.pdf.

18 Kantarjian H, Shah NP, Hochhaus A, et al. Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med* 2010; **362**: 2260-70.

19 Ciani O, Hoyle M, Pavey T, et al. Complete cytogenetic response and major molecular response as surrogate outcomes for overall survival in first-line treatment of chronic myelogenous leukemia: a case study for technology appraisal on the basis of surrogate outcomes evidence. *Value Health* 2013; **16**: 1081-90.

20 Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: cardiovascular diseases. *Stat Med* 1989; **8**: 415-25.

21 Buhr KA. Surrogate end points in secondary analyses of cardiovascular trials. *Prog Cardiovasc Dis* 2012; **54**: 343-50.

22 Friedman LM, Furberg CD, DeMets DL. What Is the Question? In: Friedman LM, Furberg CD, DeMets DL, eds. Book title. Fourth ed. New York: Springer, 2010.

23 Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA* 1999; **282**: 790-5.

24 Machado SG, Gail MH, Ellenberg SS. On the use of laboratory markers as surrogates for clinical endpoints in the evaluation of treatment for HIV infection. *J Acquir Immune Defic Syndr* 1990; **3**: 1065-73.

25 Miyamoto BE, Kakkis ED. The potential investment impact of improved access to accelerated approval on the development of treatments for low prevalence rare diseases. *Orphanet J Rare Dis* 2011; **6**: 49.

26 US Government Printing Office. Code of Federal Regulations, Title 21, Chapter 1 - FOOD AND DRUGS ADMINISTRATION, Part 314, Applications for fda approval to market a new drug, Subpart H, Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses. Washington DC: US Government Printing Office; 1997.

27 Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA* 2014; **311**: 368-77.

28 European Commission. Commission regulation (EC) No 507/2006. Brussels; 2006.

29 European Commission. Regulation (EC) no 726/2004 of the European Parliament and of the Council 2004.

30 Ishak KJ, Proskorovsky I, Korytowsky B, Sandin R, Faivre S, Valle J. Methods for adjusting for bias due to crossover in oncology trials. *Pharmacoeconomics* 2014; **32**: 533-46.

31 Bigger JT, Jr. Long-term continuous electrocardiographic recordings and electrophysiologic testing to select patients with ventricular arrhythmias for drug trials and to determine antiarrhythmic drug efficacy. *Am J Cardiol* 1986; **58**: 58C-65C.

32 Norris RM, White HD. Therapeutic trials in coronary thrombosis should measure left ventricular function as primary end-point of treatment. *Lancet* 1988; **1**: 104-6.

33 Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. *N Engl J Med* 1989; **321**: 406-12.

34 Bigger JT, Jr. The events surrounding the removal of encainide and flecainide from the Cardiac Arrhythmia Suppression Trial (CAST) and why CAST is continuing with moricizine. *J Am Coll Cardiol* 1990; **15**: 243-5.

35 Gale EA. Lessons from the glitazones: a story of drug development. *Lancet* 2001; **357**: 1870-5.

36 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007; **356**: 2457-71.

37 Nissen SE, Wolski K. Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Arch Intern Med* 2010; **170**: 1191-201.

38 Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000-2005. *JAMA* 2006; **295**: 2270-4.

39 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* 2013; **346**: f457.

40 Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; **25**: 183-203.

41 Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA* 1999; **282**: 771-8.

42 Baker SG, Kramer BS. Evaluating surrogate endpoints, prognostic markers, and predictive markers: Some simple themes. *Clin Trials* 2015; **12**: 299-308.

43 Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. *Nat Rev Clin Oncol* 2010; **7**: 309-17.

44 Perspectives on Biomarker and Surrogate Endpoint Evaluation: Discussion Forum Summary. Washington DC: 2011 by the National Academy of Sciences, 2011.

45 IOM. Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease Washington DC, 2010.

46 Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). *BMC Med Res Methodol* 2012; **12**: 27.

47 De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. *Control Clin Trials* 2001; **22**: 485-502.

48 Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med* 1997; **16**: 1965-82.

49 Shanafelt TD, Loprinzi C, Marks R, Novotny P, Sloan J. Are chemotherapy response rates related to treatment-induced survival prolongations in patients with advanced cancer? *J Clin Oncol* 2004; **22**: 1966-74.

50 Phillips PP, Fielding K, Nunn AJ. An evaluation of culture results during treatment for tuberculosis as surrogate endpoints for treatment failure and relapse. *PLoS One* 2013; **8**: e63840.

51 Burzykowski T, Molenberghs G, Buyse M. The Evaluation of Surrogate Endpoints. New York: Springer Science+Business Media, Inc.; 2005.

52 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.

53 Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumours: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care* 2014; **30**: 13.

54 Mishra P, Murray J, Birnkrant D. Direct-acting antiviral drug approvals for treatment of chronic hepatitis C virus infection: Scientific and regulatory approaches to clinical trial designs. *Hepatology* 2015; **62**: 1298-303.

55 Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006; **5**: 173-86.

56 Buyse M, Molenberghs G, Paoletti X, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biom J* 2015.

57 Tsoi B, Masucci L, Campbell K, Drummond M, O'Reilly D, Goeree R. Harmonization of reimbursement and regulatory approval processes: a systematic review of international experiences. *Expert Rev Pharmacoecon Outcomes Res* 2013; **13**: 497-511.

58 Clement FM, Harris A, Li JJ, Yong K, Lee KM, Manns BJ. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 2009; **302**: 1437-43.

59 Wisloff T, Selmer RM, Halvorsen S, Fretheim A, Norheim OF, Kristiansen IS. Choice of generic antihypertensive drugs for the primary prevention of

cardiovascular disease--a cost-effectiveness analysis. *BMC Cardiovasc Disord* 2012; **12**: 26.

60 Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes: UKPDS 40. UK Prospective Diabetes Study Group. *BMJ* 1998; **317**: 720-6.

61 Gluud C, Brok J, Gong Y, Koretz RL. Hepatology may have problems with putative surrogate outcome measures. *J Hepatol* 2007; **46**: 734-42.

62 Ellenberg S, Hamilton JM. Surrogate endpoints in clinical trials: cancer. *Stat Med* 1989; **8**: 405-13.

63 Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *Int J Clin Oncol* 2009; **14**: 102-11.

64 Dunn BK, Akpa E. Biomarkers as surrogate endpoints in cancer trials. *Semin Oncol Nurs* 2012; **28**: 99-108.

65 Berghmans T, Pasleau F, Paesmans M, et al. Surrogate markers predicting overall survival for lung cancer: ELCWP recommendations. *Eur Respir J* 2012; **39**: 9-28.

66 Prasad V, Kim C, Burotto M, Vandross A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Intern Med* 2015; **175**: 1389-98.

67 Lassere MN, Johnson KR, Boers M, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol* 2007; **34**: 607-15.

68 IQWiG. Validity of surrogate endpoints in oncology. Executive Summary. Cologne, Germany: IQWiG, 2011.

69 Velasco Garrido M, Mangiapane S. Surrogate outcomes in health technology assessment: an international comparison. *Int J Technol Assess Health Care* 2009; **25**: 315-22.

70 EUnetHTA. Endpoints used in REA of pharmaceuticals - Surrogate Endpoints, 2013.

71 NICE. Guide to the methods of technology appraisal 2013. 2013. http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9.

72 European Medicines Agency. Publication and access to clinical-trial data. EMA/240810/2013. London; 2013.

73 Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. *N Engl J Med* 2013; **369**: 475-8.

74 Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, year 1--access to data from industry-sponsored clinical trials. *N Engl J Med* 2014; **371**: 2052-4.

75 Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002; **360**: 711-5.

76 Carlson JJ, Sullivan SD, Garrison LP, Neumann PJ, Veenstra DL. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. *Health Policy* 2010; **96**: 179-90.

77 Eichler HG, Baird LG, Barker R, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clin Pharmacol Ther* 2015; **97**: 234-46.

78 Eichler HG, Oye K, Baird LG, et al. Adaptive licensing: taking the next step in the evolution of drug approval. *Clin Pharmacol Ther* 2012; **91**: 426-37.

79 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**: 303-40.

80 Schievink B, Lambers Heerspink H, Leufkens H, De Zeeuw D, Hoekman J. The use of surrogate endpoints in regulating medicines for cardio-renal disease: opinions of stakeholders. *PLoS One* 2014; **9**: e108722.

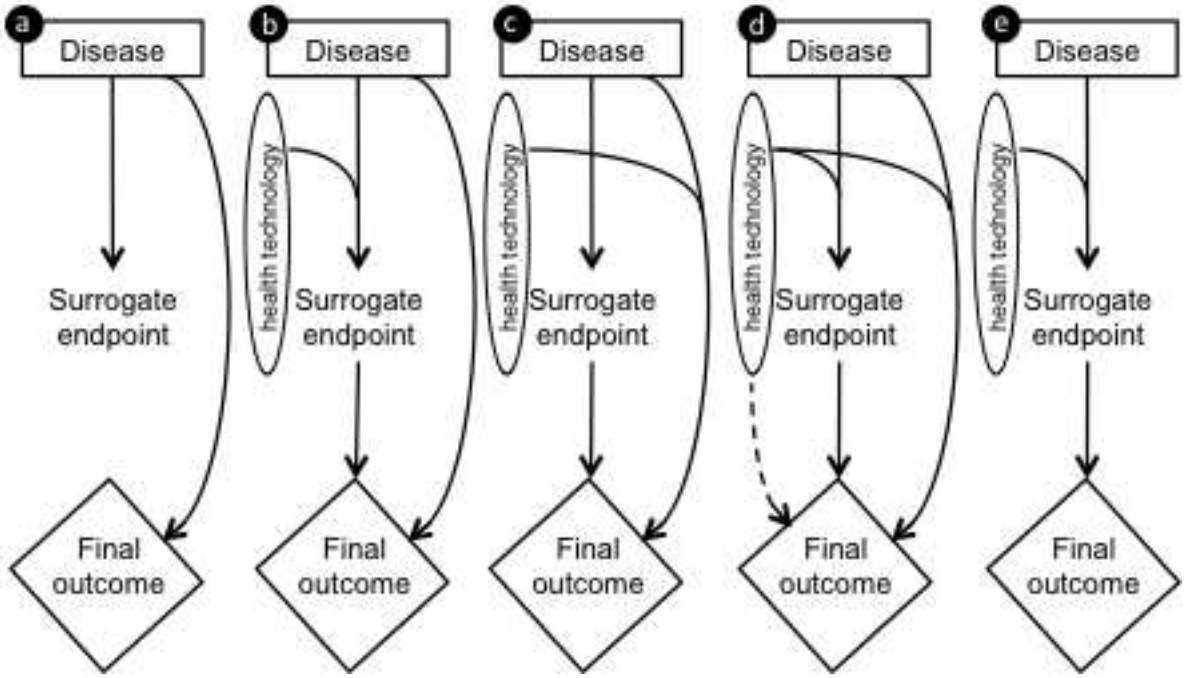
81 Rocchi A, Khoudigian S, Hopkins R, Goeree R. Surrogate outcomes: experiences at the Common Drug Review. *Cost Eff Resour Alloc* 2013; **11**: 31.

82 Beckerman R, Chowdhury C. The value of surrogate endpoints in the EU3: CBPartners, 2013.

83 la Cour JL, Brok J, Gotzsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. *BMJ* 2010; **341**: c3653.

***Figure 1*: Illustrations of different mechanisms for failure of surrogate endpoint**

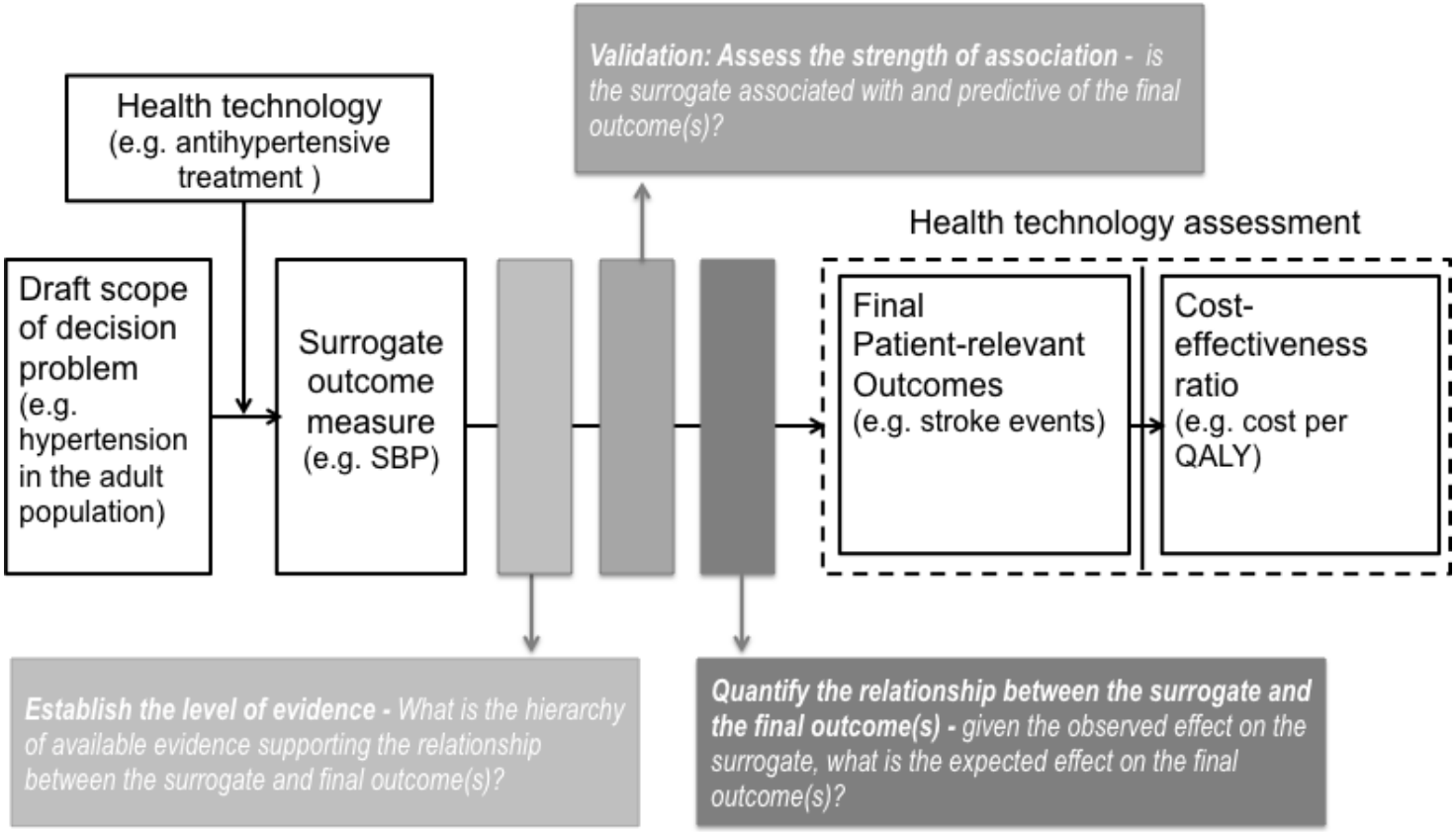
Adapted from Fleming and DeMets[7]



a) The surrogate is not in the causal pathway of the disease process; b) Of several causal pathways of disease, the health technology only affects the pathway mediated through the surrogate; c) The surrogate is not in the pathway of the health technology's effect, or is insensitive to it; d) The health technology has mechanisms of actions independent of the disease process (dotted line); e) The surrogate lies in the only causal pathway of the disease process, however, the treatment effect observed on the surrogate endpoint could underestimate or overestimate the treatment effect on the final outcome.

Figure 2: Framework for the evaluation of surrogate endpoints in a health technology assessment setting

HRQoL = health related quality of life



The graph shows the sequence of actions to implement in a health technology assessment of a drug technology when surrogate outcomes evidence is available. For example, systolic blood pressure is the surrogate endpoint for a major cardiovascular event (e.g. a stroke) and, in a cost-effectiveness evaluation, for health-related quality of life (HRQoL) and mortality. After an initial scope of the decision problem, the first step requires systematic review of the evidence explaining the relationship between the surrogate outcome and the final patient relevant outcome (establish the level of evidence). A level-2 evidence would show that groups of patients with higher average blood pressure have worse cardiovascular outcomes. A level-1 evidence would show that treatment-induced changes on the surrogate (i.e. reduction in blood pressure), would correspond to treatment-induced changes on the final outcome (i.e. a reduction in stroke events). This evidence will then be assessed to define whether the surrogate is associated with and predictive of the final patient relevant outcome (validation). If so, a quantification of the estimated effect on the final outcome given the observed effect on the surrogate outcome in the setting of interest will be performed (quantification), and could be used as input in a cost-effectiveness analysis.

***Table 1*: Outcome and endpoint definitions**

Adapted from Biomarkers Definition Working Group[5]

Endpoint	**Definition***	**Example (Diabetes mellitus)**	**Example (Cardiovascular disease)**
Biomarker	Characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention	HbA_{1C}, C-peptide	LDL-cholesterol, C-reactive protein, cardiac troponins
Patient-relevant (final) endpoint	Characteristic or variable that reflects how patient feel or function or how long they survive	Diabetic foot mortality, health-related quality of life	Stroke, myocardial infarction mortality, health-related quality of life
Intermediate endpoint	Endpoint is, or is felt to be, of value to patients but does not represent the ultimate patient-relevant final outcome of interest	Hypoglycemic symptoms	Exercise capacity
Surrogate endpoint	Biomarker or intermediate endpoint intended to substitute and predict for patient-relevant final endpoint	HbA_{1c} and glucose control as surrogate for diabetes complications	SBP as surrogate for major cardiovascular events in patients with hypertension

		and mortality	

HbA_{1C}=glycated haemogolobin, LDL-cholesterol=low-density lipoprotein cholesterol, SBP=systolic blood pressure

*Definitions adapted from the Biomarkers Definition Working Group[5]

***Table 2*: Hierarchy of evidence for surrogate endpoint validity**

Adapted from Elston and Taylor[6] and Bucher and colleagues[41]

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