Journal of Health Economics 51 (2017) 1–12

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

The role of imperfect surrogate endpoint information in drug approval and reimbursement decisions

Katalin Bognar[a], John A. Romley[b], Jay P. Bae[c], James Murray[c], Jacquelyn W. Chou[a], Darius N. Lakdawalla[b,*]

[a] Precision Health Economics, Los Angeles, CA, United States
[b] University of Southern California, Los Angeles, CA, United States
[c] Eli Lilly & Company, Indianapolis, IN, United States

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form
28 November 2016
Accepted 2 December 2016
Available online 11 December 2016

Keywords:
Surrogate endpoints
Drug approval
Drug reimbursement

ABSTRACT

Approval of new drugs is increasingly reliant on "surrogate endpoints," which correlate with but imperfectly predict clinical benefits. Proponents argue surrogate endpoints allow for faster approval, but critics charge they provide inadequate evidence. We develop an economic framework that addresses the value of improvement in the predictive power, or "quality," of surrogate endpoints, and clarifies how quality can influence decisions by regulators, payers, and manufacturers. For example, the framework shows how lower-quality surrogates lead to greater misalignment of incentives between payers and regulators, resulting in more drugs that are approved for use but not covered by payers. Efficient price-negotiation in the marketplace can help align payer incentives for granting access based on surrogates. Higher-quality surrogates increase manufacturer profits and social surplus from early access to new drugs. Since the return on better quality is shared between manufacturers and payers, private incentives to invest in higher-quality surrogates are inefficiently low.

1. Introduction

Uncertainty surrounds every healthcare decision, particularly when it comes to approval and reimbursement decisions for novel medicines. To reduce this uncertainty, drug developers decide how much and what kind of information to collect about efficacy, real-world effectiveness, and side effects, before seeking approval or reimbursement. At the same time, regulators must decide how much and what kind of information to require before granting approval and market access. Private insurers and other payers for healthcare services must make a similar decision before deciding which therapies to cover and how to cover them.

Health economists and other researchers have studied these issues from a variety of angles. In one prominent example, Claxton et al. utilize the tools of decision theory and value of information theory to develop a framework for assessing whether enough information exists to justify adoption of new medical technologies (Claxton et al., 2005; Claxton et al., 2001). In addition, both researchers and practitioners have proposed and experimented with strategies for making sound healthcare reimbursement decisions in the presence of uncertainty about clinical benefit. For example, concepts of "conditional reimbursement" have been developed that allow payers to reimburse technologies on the basis of preliminary data and to revisit those decisions once more definitive data arrive. While such strategies are not quite commonplace, there is a growing body of evidence on when and how to deploy them most successfully (Carlson et al., 2010; Niezen et al., 2007). In sum, researchers have analyzed the problem of decision making in the presence of incomplete information about clinical benefit, and market participants have begun to devise strategies for making decisions under limited information (Claxton et al., 2001; Claxton, 1999a; Claxton et al., 2015; Claxton et al., 2012; Claxton et al., 2016; Claxton et al., 2002; Eckermann and Willan, 2007; Eckermann and Willan, 2008; Griffin et al., 2011; Hutton et al., 2007).

At the same time, however, discomfort is growing among clinicians and payers about what they see as a slow but inexorable decline in the *quality* of information about new medical technologies. Increasingly, new medical technologies are brought to market on the basis of so-called "surrogate endpoint" data. For example, cancer drugs are often approved based on evidence that drugs increase "progression-free survival," defined as the number of additional months or years until a patient's cancer progresses to a

* Corresponding author.
E-mail address: dlakdawa@usc.edu (D.N. Lakdawalla).

http://dx.doi.org/10.1016/j.jhealeco.2016.12.001

more severe stage. Even though progression-free survival might not be intrinsically valuable to patients, it appears to be correlated with actual survival in several important cases. (Michiels et al., 2016; Shafrin et al., 2016; Beauchemin et al., 2014) Thus, progression-free survival is used as a surrogate for, or correlate of, the outcome patients care about most. A similar example is the use of low-density lipoprotein cholesterol (LDL-C) levels as a surrogate endpoint for cardiovascular mortality (Smith, 2015).

Indeed, drugs to treat high cholesterol as well as cancer have been approved on the basis of these "surrogate endpoints." The use of surrogates allows new drugs to be tested and approved more quickly and more cheaply. For example, it takes longer for a researcher to observe the time it takes a cancer patient to die than to observe the time it takes for their disease to progress. Thus, it would take longer to field a clinical trial measuring life expectancy than progression-free survival. Similarly, changes in LDL-C cholesterol are manifest far before cardiovascular deaths due to elevated LDL-C.

The use of surrogate endpoints continues to grow. For example, the US Food and Drug Administration (FDA) relied on surrogate endpoints in approving roughly 16 drugs per year over 2010–2014 (US Food and Drug Administration, 2015), versus roughly 6 drugs per year over 1998–2008 (US Government Accountability Office, 2009). This is perhaps to be expected as a result of progress in medicine. For example, if cancer patients have few treatment options and expect to die within months, researchers can measure final life-span more rapidly. If, however, patients live a long time when taking currently available drugs, investigators become more willing to tolerate the limitations of surrogate endpoints, in exchange for a substantially shorter or cheaper clinical trial. Malani and Philipson have identified this phenomenon of medical progress making it harder and more expensive to conduct clinical trials (Malani and Philipson, 2011).

Regardless of its underlying causes, the rise of surrogate endpoints has drawn considerable criticism charging that evidence of improvement in a surrogate endpoint should not be used to justify a claim about the effectiveness of a drug at improving patient health (Prasad et al., 2015; Fleming and DeMets, 1996; Kim and Prasad, 2016; Kim and Prasad, 2015). Since surrogates are imperfectly correlated with the final outcomes of interest, surrogate endpoint information provides weaker evidence of the benefit value than does "hard" or final outcome evidence. Yet surrogate endpoints also enable drugs to reach patients in need more quickly and potentially more cheaply.

Economic analysis can help shed some new light on this controversy. Economists will readily recognize how these issues relate to the costs and benefits associated with higher quality information (Griffin et al., 2011). We use standard economic tools to devise a straightforward and systematic framework for studying how "lower quality" surrogate endpoint evidence changes decision making about healthcare technology. In particular, this paper provides a tractable model that: (i) characterizes the benefits of higher quality surrogate endpoints; (ii) identifies the privately optimal access decisions of payers and regulators operating on the basis of imperfect surrogate endpoint information; (iii) describes the interplay between manufacturer price negotiations and the use of surrogate information; and (iv) assesses the social value of improving information quality, in light of the optimal strategies pursued by payers, regulators, and manufacturers.

Several important lessons emerge. First, we show that lower quality surrogate endpoints that are less predictive of final outcomes should lead regulators, payers, and social planners to demand greater evidence of surrogate benefit. Intuitively, decision makers substitute towards demanding a higher level of benefit when faced with a higher degree of measurement error in the clinical endpoint. As a result, measurement error reduces the expected social value from any given, newly developed drug. This in turn leads to more denials of early access and fewer new drug introductions.

Second, from a policy perspective, we show that regulators approve an inefficiently high number of new therapies while payers reimburse too few, a phenomenon that would occur even with "perfect endpoints." While regulators and payers both value clinical benefits to patients, regulators fail to consider the economic cost of using therapy, and payers focus on the price of therapy instead of its lower marginal cost of production. However, lower quality surrogates worsen these two sources of inefficiency. Payers overreact to noise in the surrogate by demanding too much additional evidence of benefit because they fail to internalize the full benefit of allowing more drugs on the market. In contrast, regulators under react by failing to tighten evidence requirements sufficiently because they fail to recognize the full benefit of reducing the number of drugs that come to market. A practical implication of this result is that payers and regulators are most likely to disagree on access when surrogates are of low quality, resulting in lower prices to manufacturers or reduced market access. Conversely, improving the quality of surrogates creates better alignment between payers and regulators when it comes to decisions about drug approval and reimbursement.

Third, pricing and information processing decisions are connected. When price-bargaining between payers and manufacturers is perfectly efficient, payers make socially efficient decisions regarding access to new drugs. Under these circumstances, the total surplus earned jointly by payers and manufacturers reflects the true social surplus. Under efficient Nash-bargaining, payers and manufacturers first maximize this joint surplus and then negotiate over how to divide it. Therefore, pricing efficiency results in social surplus-maximization, which in turn produces efficient use of the available surrogate endpoint information. Thus, an efficient pricing system helps remediate failures in the way information is processed. From a practical standpoint, inefficiencies in drug pricing and price-bargaining are numerous and widespread. However, our analysis suggests an additional benefit of mitigating these common inefficiencies.

Finally, we show that greater quality in surrogate endpoints benefits manufacturers and payers. This circumstance leads to a classic free-riding problem in which no single party has the incentive to undertake sufficient investments in improved quality. As a result of free-riding, the benefits of improved surrogates will exceed costs on the margin. Therefore, some degree of public-sector investment or subsidies for investment is called for to improve the quality of available surrogate endpoints.

While focused primarily on surrogate endpoints, our analysis also relates to the broader literature on the reliability of information about the benefit of new technologies. For example, clinical efficacy measured in clinical trials may not faithfully represent the "effectiveness" that will ultimately accrue to real-world patients because trials are conducted under constrained conditions, such as aggressive monitoring or mitigation of safety issues or adverse events (Soares et al., 2005). Outcomes in true real-world circumstances might vary from idealized randomized trial effects (Claxton et al., 2005).

Our study grows out of the decision-theoretic research that has emerged to provide a framework for evaluating the imperfect evidence available for informing adoption decisions (Claxton et al., 2005). As efforts to improve regulatory efficiency and decrease research costs continue, regulators are increasingly faced with imperfect information, one particular form being surrogate endpoints (Claxton et al., 2016; McKenna et al., 2015). Under such uncertainty, value of information analysis is particularly salient in decision making (Claxton et al., 2005; Claxton et al., 2002; Griffin et al., 2011). Regulators must consider a range of competing issues:

uncertainty regarding the drug's potential for harm; the cost to patients of less timely access to beneficial new medicines; and the ease or difficulty of reversing their adoption decisions (Claxton et al., 2016; Eckermann and Willan, 2007; Eckermann and Willan, 2008). Within the economic literature on drug regulation, Manski (2009) argues that imperfect clinical evidence calls for an adaptive rather than an "all-or-nothing" approach by the FDA (Manski, 2009). Viscusi and Zeckhauser (2015) have analyzed how cognitive biases cause regulators to respond less than optimally to uncertainty about clinical evidence (Viscusi and Zeckhauser, 2015). Our interest is in how varying the quality of clinical evidence changes privately and socially optimal decisions made by regulators, payers, and manufacturers when access and pricing are based on surrogate endpoint information.

Our study is organized as follows. In Section 2, we present the economic environment that we study theoretically. Section 3 describes our model of socially and privately optimal decision making about drug access in the presence of surrogate endpoints. Section 4 characterizes the solution for regulators, payers, and the social planner, and assesses the efficiency of regulator and payer decisions. In Section 5, we describe the implications of information quality for manufacturer pricing behavior and the division of value between manufacturers and payers. Section 6 summarizes the policy implications of all our analyses. Section 7 presents a few salient case studies illustrating several outcomes predicted by the model and suggests directions for future empirical research, and Section 8 concludes.

2. Economic environment

Consider a new drug with uncertain clinical benefit. There exists surrogate endpoint evidence of benefit, which is correlated imperfectly with the true benefit that patients value. A regulatory body must decide whether to approve this drug on the basis of the surrogate endpoint information, or whether to wait for final outcome data, which will take longer to produce. If the drug is approved by regulators, the manufacturer must decide how to price it, and payers need to determine whether they should cover the drug and provide it to enrollees, before the final outcome data is available. The payer and manufacturer decisions are again based on surrogate endpoint information.

We distinguish between "regulators," who focus solely on approving or rejecting drugs without actually paying for them, and "payers," who must pay for drugs and decide whether to cover them. For instance, the Food and Drug Administration (FDA) is the regulator in the US, the European Medicines Agency (EMA) is the regulator in Europe, and payers could be private insurance companies, or public payers like Medicaid or the United Kingdom's National Health Service.

In our setup, a favorable decision will be analogous to the "approval with research" recommendation used by the National Institute for Health and Clinical Excellence in the UK (Longworth et al., 2013), and also analogous to the comparison of "adopt and trial" and "delay and trial" strategies found in the broader literature (Eckermann and Willan, 2007; Eckermann and Willan, 2008).

The new therapy provides patients the true benefit, B, a random variable with a well-defined mean and variance. For simplicity, we treat this benefit as one-dimensional. Our theoretical results do not presume a particular type of true benefit since decision makers may vary in what they ultimately care about. Possible candidates include: mortality rate, life-years gained, quality-adjusted life years gained, healthcare costs, or adverse events avoided (Sanders et al., 2016; National Institute for Health and Care Excellence, 2013). In addition to the true benefit, information on surrogate endpoint benefit, B^{SE}, is collected in and reported from a clinical trial before evidence on the true or "final" benefit becomes available. Benefits may be in comparison to the existing standard of care, or relative to placebo.

The variables B and B^{SE} follow some joint distribution for a given disease area and/or mechanism of action. We are especially interested in the distribution of B, conditional on the realized value of B^{SE}; this conditional density proves to be key, and is denoted as $f(B|B^{SE})$.[1]

Regulators, payers, and manufacturers are all assumed to know the joint and conditional distributions of B and B^{SE}, so there is no asymmetric information. The conditional density is based on past clinical and scientific knowledge in the disease area and drug mechanism of action, and reflects the predictive power of the surrogate endpoint with respect to the final outcome. Note that this predictive power varies both by surrogate/final outcome pairs as well as by disease area. For example, blood pressure and LDL-C are considered strong predictors of acute myocardial infarction, stroke, and cardiovascular death (Ingelsson et al., 2007). Viral load is considered a weaker predictor of outcomes for patients suffering from viral diseases (Medicare Payment Advisory Commission, 2015). Progression-free survival is a stronger predictor of overall survival in breast cancer (Michiels et al., 2016; Beauchemin et al., 2014) than in gastric cancer (Paoletti et al., 2013).

In terms of prediction, we formally assume that the strict monotone likelihood ratio property (MLRP) applies to the density of the true outcome conditional on the surrogate benefit. Thus, a drug with a larger surrogate endpoint benefit is more likely to produce a larger final outcome benefit. Effectively, this assumption ensures that surrogate endpoint improvement is always valuable to the innovator on the margin. MLRP is a frequently imposed condition in information economics and in contract theory (Athey, 2002; Bolton and Dewatripont, 2005; Mas-Colell et al., 1995). In this context, we are assuming that a drug with a greater surrogate benefit is viewed as more effective than an otherwise identical drug with a smaller surrogate benefit.

We wish to analyze the decisions made when the surrogate endpoint benefit has been realized but the final outcome value has not been. This is the salient period for approval, coverage, and initial pricing decisions. Intrinsically, surrogate endpoints are used only in circumstances when they can be collected more quickly than the final outcome. (If surrogates were more difficult to collect and imperfectly correlated with the final outcome, there would be no reason to collect them.) Both the regulator and the payer develop optimal decision rules for "early access," that is, approval or denial of the drug for patient use and coverage as a function of the observed surrogate endpoint improvement. We refer to this observed level of improvement as the surrogate "signal." Fig. 1 presents the decision tree that illustrates the timing of the regulator and payer decisions about early access; the drug manufacturer also negotiates an initial price based on the surrogate endpoint information. In the terminal stage, decision makers revisit decisions about early access and could, for example, eliminate access to a drug for which the final outcome evidence proves unfavorable. To focus our analysis on key results, early access is assumed not to affect optimal access decisions in the terminal stage; we return to this assumption in the conclusion.

One nuanced issue is that the surrogate endpoint is itself measured with error in a clinical trial. Our analysis here makes the simplifying assumption that the surrogate endpoint benefit is measured accurately in the trial. In an appendix, we address sampling variability in clinical trials, and a requirement of statistical signif-

[1] We assume that the univariate as well as the joint density functions are continuous, twice differentiable and strictly positive.

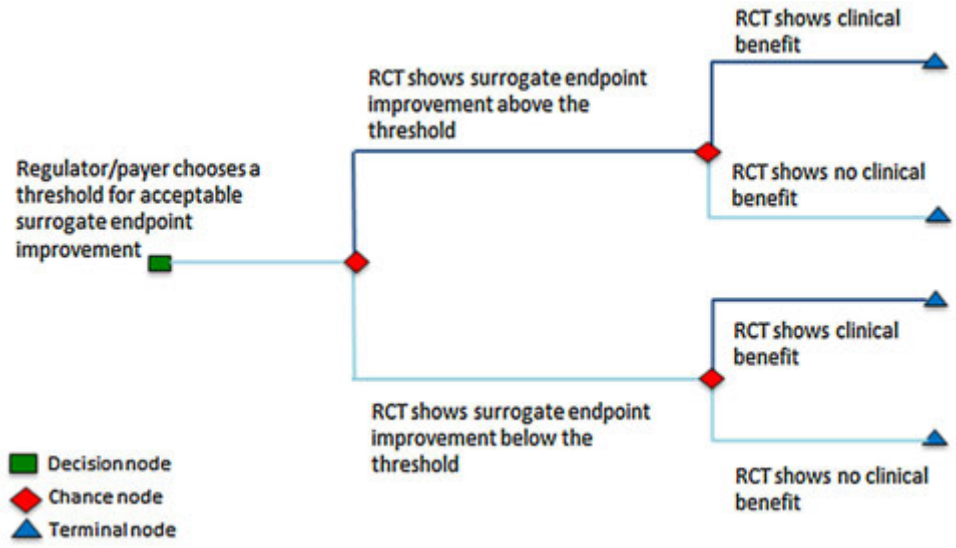

Fig. 1. Decision analytic framework.

icance; closed-form results are unavailable — except for a special case — but we demonstrate that key results continue to hold.

In analyzing decision making, we consider a representative patient who derives a constant value v for each unit of final outcome benefit (e.g., QALYs), but derives no direct utility from the surrogate endpoint benefit. A constant value per unit of benefit is consistent with much contemporary practice in health technology assessment, but is an increasingly imperfect approximation when patients are more risk-averse and variability in the potential benefit is large. It is straightforward in principle to incorporate diminishing returns into our analytic framework, but in practice closed-form results are difficult to obtain without imposing a number of unsatisfactory simplifications. Risk aversion could induce caution among decision makers about approving a drug based on a surrogate endpoint signal, because approving a "bad" drug is costlier than denying approval of a "good" drug (however, under circumstances explained below, greater uncertainty can promote access to a drug with a low-quality surrogate.)

We assume that decision makers value final outcome benefits as patients do. A payer internalizes clinical benefits accruing to its customers, while also bearing financial costs associated with treatment. In covering a drug with final outcome benefit b, the payer's utility is $vb - p$, where p is the reimbursement price. Some payers make an initial coverage decision based on the final outcome benefit that can be expected in view of the surrogate endpoint signal, and given an announced reimbursement price for the drug; this type of decision is analyzed in Sections 3 and 4. An example of such a payer would be the UK's National Health Service. Other payers may simultaneously decide whether to cover a drug and negotiate over the reimbursement price. In Section 5, we analyze this interaction between a price-negotiating payer and manufacturer, based on the Nash-bargaining framework (Lakdawalla and Yin, 2015). US private payers would typically fall into this latter category.

The regulator is assumed to care only about clinical benefits, and so its utility from approving treatment is simply vb. By contrast, a social planner also accounts for production cost, and so its utility is $vb - c$, assuming a constant marginal cost of production c. Neither the FDA nor the EMA bears the financial costs of treatment, and neither is charged with considering them.[2] One implication of our assumption is that the regulator can assess benefits in purely clinical terms, in contrast with a decision maker that must monetize benefits so as to be commensurate with costs. Yet outside the US and perhaps even in the US, a regulator may place some weight on cost. The fundamental assumption here — depending on the particular analytical result — is that the regulator places less than a dollar's worth of weight on a dollar of production cost, or that the regulator cares about production cost but not about the reimbursement price. This seems reasonable, since we define the regulator as an approval body that does not directly bear the cost of paying for the drugs. It is also worth noting that the marginal production cost of most drugs — especially conventional "small-molecule" drugs — is quite modest.

It is possible that treatment produces benefits that are not enjoyed by patients themselves. For example, employers may share in any gains in workplace productivity. A social planner and regulator focus on social value, and should account for such benefits in the value parameter v. Insofar as a payer ignores benefits which do not accrue to its enrollees, it will be more likely to deny coverage of a drug that a regulator (and social planner) would approve, reinforcing a key result in Section 3. It is also possible that the "surrogate endpoint" is valuable to the patient in and of itself, independent of how it predicts the final endpoint. For example, disease progression may be a surrogate for patient survival, but patients may also wish to delay progression, independent of survival benefits. If this is the case, the marginal benefit of increases in the surrogate will be higher than what we model, all else equal.

Abstracting from these issues, the alternative perspectives on the utility from treatment with a new drug with final outcome benefit b can be summarized as follows:

$$u_r(b; p, c) = vb \tag{1}$$

$$u_p(b; p, c) = vb - p$$

$$u_s(b; p, c) = vb - c$$

in which the subscripts r, p, and s refer to the regulator, payer and social planner, respectively. It will be convenient below to con-

[2] Public or private payers may consider costs in a secondary review of whether to allow access to the drugs; for example, the UK's National Health Service may choose on cost grounds not to grant access to a drug that was approved by the EMA. As noted earlier in this section, we include costs in a payer's utility function.

sider the general utility function $u(b; p, c)$, which nests the three perspectives above as special cases.

3. Access decisions

A regulator that is weighing whether to approve a drug before final outcome evidence is available from a clinical trial, based on a surrogate endpoint signal of a particular strength, will face one of two possibilities: (1) the drug ends up being non-inferior to a comparator therapy, so that incremental benefit is non-negative ($b \geq 0$); or (2) the drug is inferior to a comparator therapy ($b < 0$). The regulator's expected utility from early approval can be expressed in terms of these two kinds of outcomes:

$$E\left[u_r(B; p, c)|b^{SE}\right] = E\left[vB|B \geq 0, b^{SE}\right] * Pr\left[B \geq 0|b^{SE}\right] - E\left[-vB|B < 0, b^{SE}\right] * Pr\left[B < 0|b^{SE}\right] \quad (2)$$

where $E\left[u_r(B; p, c)|b^{SE}\right]$ and $Pr\left[B \geq 0|b^{SE}\right]$ stand for expected utility and probability conditional on the event that the surrogate endpoint signal is b^{SE} (formally, $E\left[u_r(B; p, c)|B^{SE} = b^{SE}\right]$ and $Pr\left[B \geq 0|B^{SE} = b^{SE}\right]$).

Expected utility consists of a difference between two terms. The first term represents the expected upside from early approval of a drug that proves to have positive final outcome benefit, weighted by its likelihood. The second term is the downside from a drug with negative final outcome benefit.

The payer faces an analogous but distinct trade-off in deciding whether to cover a drug. From its standpoint, the two key cases are: (1) the drug ends up being worth its price in terms of final outcome improvement, i.e., $vb \geq p$; or (2) the drug ends up being worth less than its price, i.e., $vb < p$. The payer's expected utility from early coverage based on the surrogate signal can thus be expressed as:

$$E\left[u_p(B; p, c)|b^{SE}\right] = \left(E\left[vB - p|vB \geq p, b^{SE}\right] * Pr\left[vB \geq p|b^{SE}\right]\right) - (E[-(vB - p)|vB < p, b^{SE}] * Pr\left[vB < p|b^{SE}\right] \quad (3)$$

The first term is the upside from a drug with value exceeding price, and the second is the downside from a drug with price exceeding value. While a payer and a regulator face the same conceptual trade-off, the way they perceive benefits and costs differs. This difference will lead to distinctive decision rules.

We now characterize private and social thresholds for drug access. The MLRP assumption implies that the expected utilities from early access in Eqs. (2) and (3) are monotonically increasing in the level of surrogate endpoint benefit. Thus, the optimal decision rules for both regulators and payers are such that approval or coverage will be granted if and only if the surrogate benefit signal is strong enough, that is, higher than a particular threshold value. The therapy will not be made available if the surrogate evidence is below the threshold, as shown in Fig. 2 for a regulator's approval decision. The optimal thresholds for a regulator and a payer respectively maximize the clinical benefits and net benefits in terms of the final outcome benefit that can be expected according to the surrogate signal.

Formally, the general form of the problem faced by a decision maker with objective function $u(b; p, c)$ is

$$\max_{\tau} E\left[u(B; p, c)|b^{SE} \geq \tau\right] = \max_{\tau} \int_{\tau}^{\infty} \left(\int_{\mathbb{R}} u(b; p, c) f\left(b|b^{SE}\right) db\right) f_{SE}\left(b^{SE}\right) db^{SE} \quad (4)$$

in which τ denotes the threshold for access.

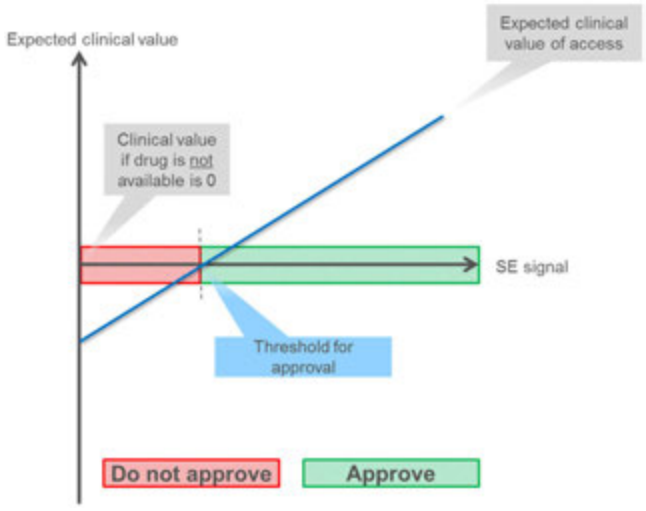

Fig. 2. Optimal approval decision for a regulator.

The formulation above treats the surrogate endpoint benefit as known. In reality, it too is observed with error that depends on the sampling variability in the trial. For the sake of comparison, the following decision problem incorporates sampling variability in the trial SE:

$$\max_{\tau} \int_{\mathbb{R}} \int_{\tau-\varepsilon}^{+\infty} \int_{\mathbb{R}} u(b; p, c) f\left(b, b^{SE}, \varepsilon\right) dbdb^{SE}d\varepsilon$$

in which b^{SE} is the true SE, ε is statistical noise, and the measured SE, $\hat{b^{SE}} = b^{SE} + \varepsilon$. This more complex problem is analyzed in the appendix, where we show that our key results from the simpler formulation continue to obtain, albeit without elegant closed-form solutions.

Applying Leibniz's rule to the problem without sampling variability, the derivative of the objective function in Eq. (4) with respect to the threshold is

$$-\left(\int_{\mathbb{R}} u(b; p, c) f(b|\tau) db\right) f_{b^{SE}}(\tau) = -E\left[u(B; p, c)|\tau\right] f_{SE}(\tau) \quad (5)$$

Recognizing that $f_{SE}\left(b^{SE}\right)$ is everywhere positive, the optimal threshold τ^* satisfies the first-order condition:

$$E\left[u(B; p, c)|\tau^*\right] = 0 \quad (6)$$

The uniqueness of the optimum is ensured by the MLRP assumption. For the social planner, the first order condition is

$$vE\left[B|\tau_s^*\right] = c \quad (7)$$

However, for the regulator, Eq. (6) becomes $vE\left[B|B^{SE} = \tau_r^*\right] = 0$, which implies an inefficiently low approval threshold. Since the social planner places more weight on cost, the lower-threshold regulator approves "too many" drugs based on surrogate endpoints.

By similar reasoning, a payer will demand a stronger surrogate signal than the social planner before covering the drug, since the drug's reimbursement price exceeds its production cost. Thus, a payer approves "too few" drugs based on surrogate endpoints. This rank ordering of decision makers' approval/coverage thresholds is illustrated in Fig. 3. Later, we describe how variation in the quality of surrogates influences the extent of inefficient over- and under-approval.

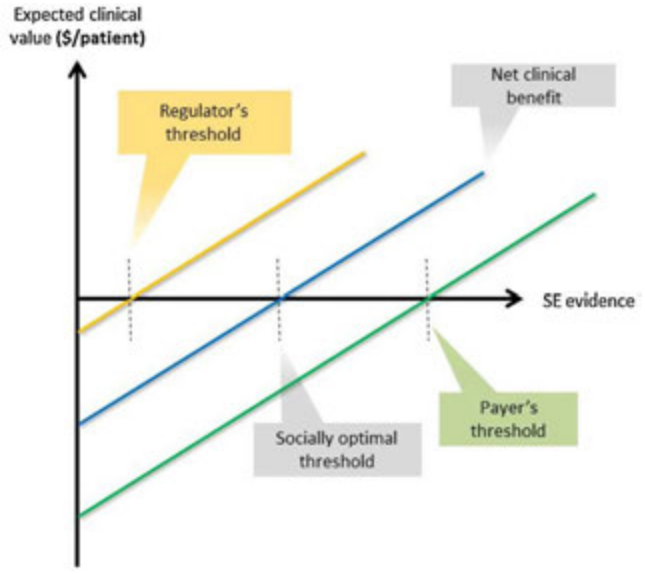

Fig. 3. Optimal thresholds in the three perspectives.

4. Determinants of access decision rules

The privately and socially optimal decision rules depend on various factors, including the quality of the surrogate endpoint in predicting the final outcome benefit; prior expectations of surrogate and final outcome benefits; the value of a final outcome benefit; and the payer and social cost of treatment with a new drug. Here, we explore how these factors affect decision making. In what follows, we assume joint normality of surrogate and final outcome benefits to obtain closed-form-expressions for access thresholds.

4.1. Conditional expected benefit under joint normality

For all decision makers, utility is linear in its arguments, and expected utility depends on the expectation of the final outcome benefit conditional on the strength of the surrogate endpoint signal. Under joint normality of the surrogate and final outcome benefits, a standard result is that the conditional expectation of the final outcome benefit can be written as (Greene, 2012):

$$E[B|b^{SE}] = E[B] + \rho\frac{\sigma_B}{\sigma_{SE}}\left(b^{SE} - E\left[B^{SE}\right]\right), \tag{8}$$

in which ρ is the correlation between the surrogate and final outcome benefits, $E[B]$ and $E\left[B^{SE}\right]$ are the expectations of each prior to the trial, and σ_{SE} and σ_B are their respective standard deviations.

If the surrogate signal turns out as expected $\left(b^{SE} = E\left[B^{SE}\right]\right)$, the signal has no information content, and the expected final outcome benefit is simply its prior expectation. If the surrogate signal turns out to be one unit above its expectation, this new and favorable information predicts that the final outcome benefit will be $\rho\frac{\sigma_B}{\sigma_{SE}}$ units higher, on average, than its own expectation. The term $\rho\frac{\sigma_B}{\sigma_{SE}}$ increases in the strength of the correlation between the endpoints, and corresponds to the coefficient that would be obtained by regressing final outcome benefits on surrogate endpoint benefits, an exercise that is designed to maximize predictive power. We refer to this term as the "quality" of the surrogate endpoint signal. For a higher quality surrogate, one unit of improvement leads to a larger upward revision in the expected final outcome improvement. In cancer, for example, progression-free survival is a higher-quality surrogate for overall survival for breast tumors (Michiels et al., 2016; Beauchemin et al., 2014) than for gastric tumors (Paoletti et al., 2013).

4.2. The regulator's approval threshold

The regulator will approve the drug if $E\left[B|b^{SE}\right] \geq 0$. Applying Eqs. (1) and (8) to Eq. (6), this condition is equivalent to approving drugs with $b^{SE} \geq \tau_r^*$, where:

$$\tau_r^* = E\left[B^{SE}\right] - \frac{E[B]}{\rho\frac{\sigma_B}{\sigma_{SE}}} \tag{9}$$

There are two cases to analyze here, although the intuition is similar in both. First is the case where $E(B) > 0$, so the drug is thought to be beneficial ex ante. In this case, the trial ascertains whether there is any "bad news" that would revise beliefs downward and result in a rejection. For a higher quality surrogate (e.g., higher $\frac{\sigma_B}{\sigma_{SE}}$), a smaller amount of "bad news" will cause the regulator to reject the product. Therefore, the threshold τ_r^* will move upwards. So, for example, the same amount of bad news could lead a regulator to approve a gastric cancer drug based on progression-free survival, but to reject a breast cancer drug, because this surrogate is a stronger predictor of overall survival for breast cancer (Michiels et al., 2016; Beauchemin et al., 2014) than for gastric cancer (Paoletti et al., 2013).

When $E(B) < 0$, the trial provides the decision makers with an option to make use of the drug if it proves to be more beneficial than expected. Thus the issue is whether there is "good news" in the trial surrogate endpoint that would revise beliefs about benefit upward and result in approval. For a higher quality surrogate, a smaller amount of "good news" will cause the regulator to approve the product. Therefore, the threshold τ_r^* moves down. Note that in both cases, when the surrogate is of higher quality, the regulator requires a smaller amount of "news" to change her beliefs.

To summarize, the regulator sets a higher approval threshold if

1 Its prior expectation of the final outcome benefit ($E[B]$) is relatively low
2 Its prior expectation of the *surrogate* outcome benefit ($E\left[B^{SE}\right]$) is relatively *high*
3 The surrogate is of relatively low quality, and $E[B] < 0$
4 The surrogate is of relatively *high* quality, and $E[B] > 0$.

The approval threshold does not depend on v, the value of the final outcome benefit to patients, because the regulator does not weigh the financial cost of the new drug against its clinical benefit. Thus, there is no need to "monetize" clinical benefit in the regulator's decision.

4.3. The payer's decision

The payer will grant coverage to the drug if $vE\left[B|b^{SE}\right] \geq p$. This condition is equivalent to allowing access for drugs with $b^{SE} \geq \tau_p^*$, where:

$$\tau_p^* = E\left[B^{SE}\right] - \frac{E[B]}{\rho\frac{\sigma_B}{\sigma_{SE}}}\left(1 - \frac{p/E[B]}{v}\right) \tag{10}$$

As with the regulator, the payer sets a higher threshold for access if the prior expectation of the final outcome benefit is relatively low or the prior expectation of the surrogate endpoint benefit is relatively high.

In addition, the payer sets a higher threshold if the monetary value of final outcome improvement is relatively low, or the reimbursement price of the drug is relatively high. These conditions on v and p have the natural implication that a payer sets a higher threshold if the prior expectation of the net benefit to the payer ($vE[B] - p$) is relatively low. Relatedly, the payer also sets a higher threshold if the expected cost effectiveness of the new drug is unfavorable from the payer's perspective (that is, $p/E[B]$ is relatively high). Thus, in

contrast with a regulator, a payer's decision making (in terms of the access threshold imposed on new drugs) is driven in part by its drug acquisition costs. As with the regulator, a payer sets a higher threshold when the surrogate is relatively low quality and the prior expectation of benefit is negative, or the surrogate is higher quality and expected benefit is positive. However, a payer's perspective on benefit is net of price (that is, $vE[B]-p$).[3]

4.4. The socially optimal decision

The social planner sets the approval threshold at

$$\tau_s^* = E\left[B^{SE}\right] - \frac{E[B]}{\rho\frac{\sigma_B}{\sigma_{SE}}}\left(1 - \frac{c/E[B]}{v}\right) \tag{11}$$

This expression is identical to a payer's threshold, with cost replacing price. The social planner weighs benefits and costs in setting its access threshold, but from the perspective of society. Thus, for example, the planner will set a higher threshold when the "social" cost-effectiveness ratio is expected to be less favorable (i.e., $c/E[B]$ is higher.)

4.5. Determinants of inefficiency

The decisions of a regulator and payer depart from that of the social planner, as was shown in Fig. 3. The regulator's threshold is weakly less stringent than the efficient threshold, because it fails to consider cost. Based on Eqs. (9) and (11), the difference in thresholds is:

$$\tau_s^* - \tau_r^* = \left(\frac{c}{v\rho\frac{\sigma_B}{\sigma_{SE}}}\right) > 0 \tag{12}$$

This expression suggests, intuitively, that the degree of misalignment rises with cost because the regulator fails to consider the marginal cost of producing the drug. Since both decision makers care about the value of the clinical benefit, the misalignment decreases with v, when costs are held constant. A less obvious implication is that a lower-quality (i.e., more poorly correlated) surrogate endpoint exacerbates this inefficiency, leading to a greater deviation in the regulator's threshold. Intuitively, a regulator fails to account for the cost of treatment with new drugs when setting its threshold; as a consequence, a regulator strengthens its approval threshold too little in response to less information in the surrogate endpoint. Speaking heuristically, the marginal cost of mistakenly approving a drug is lower for the regulator than the social planner. Thus, the regulator responds less than the planner does to increases in the likelihood of mistakes, and thus to reductions in the quality of the surrogate endpoint.

In contrast to regulators, payers impose a threshold that is too stringent, because they "over-internalize" costs, accounting for the reimbursement price rather than the social cost. From Eqs. (10) and (11), the difference in thresholds is

$$\tau_p^* - \tau_s^* = \left(\frac{p-c}{v\rho\frac{\sigma_B}{\sigma_{SE}}}\right) > 0 \tag{13}$$

Once again, it is fairly intuitive that the degree of misalignment rises with the manufacturer markup, $p-c$, since this markup is the wedge between payer and social costs. As with the regulator decision rule, a lower-quality surrogate endpoint exacerbates the extent of inefficiency.

A final implication follows from the difference between the payer and regulator thresholds:

$$\tau_p^* - \tau_r^* = \left(\frac{p}{v\rho\frac{\sigma_B}{\sigma_{SE}}}\right) > 0 \tag{14}$$

Specifically, a regulator and a payer disagree the most about access thresholds when surrogate endpoint quality is low. It is important to emphasize that this result does not depend on the signs of the prior expectations of benefit for the two decision makers.[4]

5. Manufacturer pricing behavior

In the preceding section, we analyzed the decision making of a payer that sets a threshold for covering a new drug based on a surrogate endpoint signal, given a fixed reimbursement price. This may be applicable, for instance, to some public payers that require manufacturers to announce a price and then determine whether the drug will be covered. We now consider the case of a payer that simultaneously determines coverage and negotiates the price with the manufacturer. Many private payers, and some public payers, solve this problem.

To focus squarely on the role of the surrogate endpoint signal, we abstract from complexities relating to negotiation between a manufacturer and multiple payers. For example, we rule out sequential bargaining games, and instead consider a single payer. To simplify further, we assume that sales of a covered drug are inelastic with respect to the price. Thus, a profit-maximizing manufacturer seeks to maximize net revenue $p-c$, while the payer maximizes its expected utility given the surrogate endpoint signal. It is worth emphasizing that our setup produces relatively simple pricing rules that might not match more complex real-world pricing environments.

The payer and manufacturer will jointly agree to launch the drug only if their joint surplus from doing so is positive. If this fails to be true, they cannot both find it worthwhile to launch the drug. This simple point has an important implication, because the sum of the payer's and manufacturer's surplus, $\left(v\left(E[B]+\rho\frac{\sigma_B}{\sigma_{SE}}\left(b^{SE}-E\left[B^{SE}\right]\right)\right)-p\right)+(p-c)$, is equal to social surplus,$v\left(E[B]+\rho\frac{\sigma_B}{\sigma_{SE}}\left(b^{SE}-E\left[B^{SE}\right]\right)\right)-c$. Therefore, it follows that drugs are launched only if the expected social surplus from doing so is positive. This is the same criterion that the social planner uses. Therefore, in a simple and efficient setting with Nash-bargaining, zero bargaining costs, and payers that fully internalize the value of drugs, payers and manufacturers launch the efficient number of drugs, given the quality of the surrogate endpoint information. Note that this would be true even if a regulator approved too many drugs in an earlier stage — the market would filter out the excess efficiently. One can also use similar reasoning to show that payers make efficient benefit design decisions under ideal Nash-bargaining (Lakdawalla and Sood, 2013). Intuitively, payers and manufacturers find it optimal to maximize joint surplus first, and then bargain over how to divide it. To be sure, real-world price negotiation is unlikely to be perfectly efficient for a variety of reasons discussed below. However, this theoretical result demonstrates the connection between efficient pricing and the efficient use of the available surrogate endpoint information, and vice-versa.

[3] It is possible for the expected final outcome benefit to be positive while the expected benefit *net* of the reimbursement price is negative. In this case, a payer responds to a higher quality signal by lowering its coverage threshold, while a regulator raises its approval threshold.

[4] If the expected final outcome benefit is positive while the expected net benefit is negative, the result here follows from the results in Sections 4.2 and 4.3. The result here also holds when the decision makers alter their threshold decisions in the same direction when surrogate quality rises. Intuitively, the difference in thresholds is driven entirely by the decision makers' contrasting perspectives on costs, under the maintained assumption that a regulator and payer value clinical benefits to patients equally.

In light of these results, we can — without loss of generality — abstract from payers' access and benefit design decisions when setting up the bargaining problem, because these problems are solved optimally in an earlier and independent stage. This avoids cumbersome notation without changing the analytical results for pricing. Denoting the Nash-bargaining leverage of the manufacturer as $\delta \in (0, 1)$, the Nash-equilibrium is the solution to

$$\max_p (p-c)^{\delta}\left\{v\left(E[B]+\rho\frac{\sigma_B}{\sigma_{SE}}\left(b^{SE}-E\left[B^{SE}\right]\right)\right)-p\right\}^{1-\delta}$$

in which the expression in curly brackets is the payer's expected utility under joint normality. This problem has the first-order condition:

$$\left(1-\delta\right)(p^*-c)=\delta\left\{v\left(E[B]+\rho\frac{\sigma_B}{\sigma_{SE}}\left(b^{SE}-E\left[B^{SE}\right]\right)\right)-p^*\right\}$$

which can be restated as

$$(p^*-c)=\delta\left\{v\left(E[B]+\rho\frac{\sigma_B}{\sigma_{SE}}\left(b^{SE}-E\left[B^{SE}\right]\right)\right)-c\right\}$$

The expression in curly brackets is the total social surplus that can be expected from drug coverage based on a surrogate endpoint signal. In equilibrium, the manufacturer's share of this surplus is equal to its leverage, the standard Nash-bargaining result. The price, manufacturer profits, payer profits, and total social surplus are all higher if

1 The surrogate is of relatively high quality ($\rho\frac{\sigma_B}{\sigma_{SE}}$), and the surrogate endpoint signal is "good news" ($b^{SE} > E\left[B^{SE}\right]$)
2 The surrogate is of relatively low quality, and the surrogate endpoint signal is "bad news" ($b^{SE} < E\left[B^{SE}\right]$)
3 There is more "good news" or less "bad news" in terms of the strength of the surrogate endpoint signal (i.e., $b^{SE} - E\left[B^{SE}\right]$ is higher)
4 The prior expectation of the final outcome benefit is relatively favorable ($E[B]$ is larger).

The latter two conditions specifically imply that the expected final outcome benefit is greater. Where the final outcome benefit is relative to a standard of care defined by the efficacy of other drugs, the reimbursement price is higher when the new drug's expected benefit over prior treatments is greater. The availability of therapeutic alternatives may further affect new drug pricing by altering the relative bargaining power of a payer and manufacturer.

A key result here is that a manufacturer benefits from an improvement in the quality of a surrogate endpoint — through accelerated approval and coverage decisions — and so does a payer, as one would expect. The fact that each captures only a share of the increase in social surplus has important policy implications.

6. Policy implications

In Section 4.5, we showed that a higher-quality surrogate endpoint helps improve alignment between drug approval and reimbursement decisions. Section 5 showed that efficient pricing leads to efficient coverage thresholds being set by payers, given the available surrogate endpoint information. From a policy perspective, this latter result suggests that inefficient prices preserve distortions in the way information is processed and likely result in excessive denials of access to drugs. For example, this result might apply if prices are set administratively in a manner that does not match market outcomes, or if payers fail to internalize the benefits accruing to patients (i.e., v differs for payers and patients). From a practical standpoint, such inefficiencies in pricing are more likely to be the rule than the exception, but our analysis reveals the linkage between price regulation and the processing of clinical information by the marketplace. Policies that move towards more efficient pricing also lead to more efficient processing of surrogate information and more efficient access decisions.

We also showed that efficient pricing "cleans up" inefficiency in regulator behavior. Even if regulators approve too many drugs, efficient pricing will result in optimal access. Thus, regulators can afford to take more risks on approving drugs with uncertain benefits, when pricing is more efficient, and vice-versa. For instance, physician-administered cancer drugs treating older populations – whose prices are administratively set by Medicare Part B – might warrant greater caution than antidepressants reimbursed primarily by private payers.

Section 5 also showed how higher quality increases the total social surplus from early access to new drugs, because higher-quality surrogates lead to more accurate decisions based on surrogate endpoint information. Mathematically, this effect appears as higher social value per new drug, and accrues both to payers and manufacturers.

The sharing of surplus between payers and manufacturers results in a classic free-riding problem, in which neither party faces efficient incentives to invest in improving the quality of surrogate endpoints. Thus, absent public-sector investments or subsidies for private research into improved endpoints, the following market failures are likely to occur:

1 There are drugs left "in limbo," with regulatory approval but coverage denial;
2 For each drug that has been developed and brought to trial, society can expect to realize less value;
3 Investments in pharmaceutical innovation are lower because the expected return on investment for the manufacturer is dampened.

The third point follows because higher quality surrogates lead to greater profit (Section 5), and in turn greater profits lead to greater investments in innovation according to a variety of models in the literature (Nordhaus, 1969; Loury, 1979; Green and Scotchmer, 1995). Thus, firms will be more likely to innovate in areas with higher quality surrogates. This creates an interaction between the quality of evidence and the quantity of innovation. Malani and Philipson make a complementary point about the cost of evidence, with a particular focus on the context of HIV treatment (Malani and Philipson, 2011).

Therefore, improvements in the quality of surrogates can lead to fewer denials of early coverage of drugs approved by regulators, higher expected social value from each drug that enters trials, and more drugs developed and brought to trial. Moreover, public policies that stimulate improvements in surrogate quality may be needed in order to achieve these aims, because private benefits are shared across many parties.

Finally, it is useful to note that the problem of interest here — free riding in the costly production of high quality surrogates — arises much more generally. For example, where a manufacturer has invested in evidence that demonstrates efficacy of a drug, and another manufacturer seeks approval of a new drug with a similar mechanism of action but fewer side effects, the second manufacturer will be able to rely to some degree on the first manufacturer's costly evidence in expediting its own approval.[5]

[5] We thank an anonymous referee for this point.

7. Case studies and empirical context

To illustrate some of the practical implications of our theoretical framework, we present a few historical case studies drawn from FDA approval and payer rejection experience. Causal inference regarding the effects of endpoint quality lies beyond the scope of this paper. We aim instead to provide a few real-world examples of the various kinds of outcomes predicted by the theory and suggest a path forward for future empirical research in this area.

First, our theory predicted that rejections might occur for drugs with low-quality surrogates, and particularly if the drug underperformed prior expectations. For example, rociletinib was investigated for use in a genotype of non-small cell lung cancer known as EGFR+. The surrogate endpoint in this case was "objective response rate" (ORR), which is the fraction of patients whose tumors shrank by a prespecified size. ORR is a surrogate for overall survival in cancer, but it is weakly correlated with overall survival in non-small cell lung cancer (Blumenthal et al., 2015). In spite of a relatively strong ORR benefit — 30.2% of patients exhibited tumor shrinkage — an FDA advisory committee voted against accelerated approval, instead recommending to wait for further evidence of benefit (US Food and Drug Administration, 2016).

Second, our framework suggests that the misalignment between regulator and payer preference will result in some drugs "left in limbo," i.e., approved for use but not covered by payers. One example is aclidinium bromide for the treatment of acute episodes among patients with chronic obstructive pulmonary disease (COPD). The surrogate endpoint in this case is "forced expiratory volume in one second" (FEV-1), a measure of how much air is expelled from the lungs within one second. FEV-1 is considered predictive in COPD of mortality, hospitalizations, and exacerbations of disease. (Niewoehner et al., 2000; Paul and David, 2014) However, it is still thought to be a low-quality surrogate (Paul and David, 2014; Vestbo et al., 2008). This drug was approved by the FDA, but its use has been limited among insurers. For example, some private insurers do not cover it at all (Cigna, 2016), while others subject it to dispensing limits (BlueCross BlueShield of Illinois, 2016) or other access restrictions (Harvard Pilgrim Health Care, 2016). Another example is belimumab, which was approved to treat lupus on the basis of a surrogate endpoint, but subsequently denied reimbursement by the top five national payers in Europe who found the surrogate too weakly predictive of final benefit (Marinoni, 2012). In these cases, regulators perceive enough benefit for approval, but payers do not always see fit to provide access.

Third, the framework suggests the possibility of drugs approved and covered on the basis of surrogates alone, particularly when the drug generates a strong signal using a high-quality surrogate. Ruxolitinib is a first-in class treatment for myelofibrosis, approved on the basis of percent reduction in spleen size as a surrogate endpoint (US Food and Drug Administration, 2011a). Patients with myelofibrosis experience substantial symptom burden, which can be managed and alleviated with spleen size reduction (Mascarenhas and Hoffman, 2013; MPN Research Foundation, 2016). Scientific understanding of myelofibrosis, a type of chronic leukemia, as a disease is still evolving, with spleen size reduction and patient reported outcomes serving to be the best endpoints for myelofibrosis treatments (Mascarenhas and Hoffman, 2013; MPN Research Foundation, 2016). Strengthening the signal of its surrogate endpoint, ruxolitinib has also demonstrated the potential correlation between spleen reduction and long term mortality in preliminary findings from clinical trials (Verstovsek et al., 2012a; Verstovsek et al., 2012b). The quality of and strong signal from the surrogate endpoint for ruxolitinib has resulted in its widespread coverage and placement on preferred drug lists (Express Scripts, 2016; CVS Caremark, 2016).

Fourth, the theory presupposes that approvals are sometimes granted based on surrogate endpoints that are later shown to be poor predictors of final outcome benefit. One example is bevacizumab, which was approved in 2008 to treat breast cancer on the basis of clinical trials showing a progression-free survival benefit of 5.5 months (US Food and Drug Administration, 2011b). Progression-free survival is thought to be a good quality surrogate for overall survival in breast cancer (Michiels et al., 2016; Beauchemin et al., 2014). The FDA granted "accelerated approval" on the basis of this information, but required that additional clinical evidence be collected in the meantime. When the follow-up information became available, there was no evidence that patients lived longer when treated with bevacizumab, and the FDA withdrew its approval for breast cancer in 2011 (US Food and Drug Administration, 2011b; US Food and Drug Administration, 2011c).

As noted, these examples do not systematically test our theoretical predictions, but they illustrate their connection to real-world outcomes. More systematic empirical testing could proceed along several lines. To stimulate further research, we suggest a few paths forward, recognizing that a full development of these empirical strategies lies beyond the scope of this paper. First, one might view the correlation between surrogates and final endpoints as exogenous in a particular disease area. For instance, progression-free survival is known to be a better correlate of overall survival in some cancer tumor types than in others. Exploiting this fact, one might proceed by estimating a model of approval and reimbursement thresholds as a function of this correlation. These thresholds are latent variables that can be modeled using data on binary approval and reimbursement decisions in the usual way. The theory predicts that decision makers call for more stringent evidence of surrogate benefit within tumor types that have lower-quality surrogates. To conduct such a study, one would need to cull the literature for estimated correlations between the surrogate endpoint and the final endpoint, across a range of diseases. These data could then be combined with approval data.

Second, one could test the prediction that payers and regulators behave differently. One could collect reimbursement data measuring the fraction of payers that gave any access to the drug, or preferred access to the drug. A complication here is deciding how to measure "any access" and "preferred access," but this seems a surmountable obstacle for an empirical study. Given this information, one could assess empirically whether and to what extent payers demand more stringent evidence of benefit than regulators, when faced with disease areas that have lower-quality surrogates.

Finally, one could assess the effects of free-riding incentives on investments into surrogates. Here, pharmaceutical pipeline data could be used to ascertain the surrogate endpoints used in clinical trials for a given disease area. One could simply count the number of surrogates and determine whether the arrival of new surrogate measures is more likely in disease areas where fewer firms are competing to develop or market new drugs. This implication is supported by theory, because free-riding incentives are always exacerbated by the presence of more competitors.

8. Conclusion

Uncertainty is inevitable when evaluating new medical treatments. We have evaluated one particular source of uncertainty, imperfect information about clinical benefit (such as quality-adjusted life years). In some contexts, the availability of reliable information has improved, due to lower costs for particular data points or analytic techniques. Yet in our setting, there is reason to be concerned about the quality of information. As medical technologies continue to improve, uncertainty about clinical benefit may become more important. Longer lives and better health mean,

fortunately, that clinical trial participants will be slower to die or develop significant comorbid diseases. Thus, it will become increasingly costly to collect final outcome data. On one hand, this may lead to longer and more expensive clinical trials, an eventuality that imposes its own costs on patients waiting for new therapies to be introduced. On the other hand, this trend may hasten the substitution towards imperfect surrogate endpoints. The question of how best to incorporate surrogate endpoints into regulatory, reimbursement, and pricing decisions will become increasingly important.

Our analysis reveals three main themes for policy makers. First, imprecision in surrogate endpoint measures leads to excessive denials of early access to new drugs, lower social value for each drug that is developed and brought to trial, and lower levels of innovation. This tendency is exacerbated by inefficiency in drug pricing, and by misalignment between the objectives of payers and of patients. Thus, policies that improve the efficiency of pricing also help mitigate inefficiencies associated with surrogate endpoints. At the same time, regulators can afford to take more risks on "less well-proven" benefits, when the drug will be priced in a well-functioning private market, and vice-versa for drugs with inefficient pricing. Moreover, private incentives to invest in better surrogate endpoints are likely to be insufficient because the benefits are shared across payers and manufacturers. This misalignment creates a role for public policy to stimulate some (finite) degree of improvement in surrogate endpoints through direct investment in or subsidies for research.

Surrogate endpoints can be based on biomarkers. Thus support for the identification of new biomarkers through gene expression profiling, and for the empirical validation of identified biomarkers, can help to improve the quality of surrogate endpoint information. As a noteworthy example, the Biomarkers Consortium has launched an effort targeting autism spectrum disorder (National Institutes of Health, 2015); public partners within this consortium include the National Institutes of Health, the FDA, and the Centers for Medicare and Medicaid Services (The Biomarkers Consortium, 2011). In diabetes and related diseases, the protein adiponectin offers promise as a surrogate, but requires further validation (Lim et al., 2014).

A second theme for policy makers from our analysis is that economic factors can and should be incorporated into clinical evidence requirements. Greater leniency is warranted for products with greater expected social value, and vice versa. The role of economic factors in reimbursement decisions is often fairly explicit, but regulators typically disavow a connection between approval based on clinical factors and the economic value of a new therapy. Agencies like the FDA and EMA may need to begin considering costs. On the other hand, the theory also suggests that an efficient pharmaceutical pricing regime can substitute for this requirement. With ideally efficient pricing, payers replicate the socially efficient approval threshold. Here, even if regulators approve too many drugs, payers would apply an efficient filter. Thus, it would appear that policy makers can consider introducing economic criteria into drug approval decisions, and/or reforms that improve the efficiency of the pricing regime, e.g., less reliance on administratively determined prices. While the latter is a more daunting policy problem to solve, it also has a broad array of social benefits.

Finally, imperfect surrogate endpoint information can be efficiently incorporated into approval and reimbursement processes. Decision makers would be prudent to call for higher standards of evidence when faced with poorer-quality surrogates or lower expectations of clinical benefit from new products. However, rather than resisting the emergence of surrogate endpoints, healthcare decision makers should search for ways to make the best possible use of the information available.

One possibility is to utilize new regulatory decision schemes that move beyond traditional binary outcomes of approval or rejection. For example, surrogate endpoints could be particularly favorable in adoption decision schemes such as "coverage with evidence development" (CED), also known as "performance-based risk sharing agreements" (PBRSA). CED gives conditional availability for promising technologies with the requirement that further evidence of benefit is produced (Hutton et al., 2007). The primary aim of such schemes is to reduce uncertainty in outcomes, efficacy, cost, economic benefit such that an appropriate price that aligns the benefit to the manufacturer and the value to the patient (Carlson et al., 2010; Garrison et al., 2013). Such increased regulatory flexibility for surrogate endpoint data would enable patients to receive clinical benefits using technologies that otherwise would have been rejected by regulatory bodies. CED is currently being used by the Centers for Medicare and Medicaid (CMS) to assess the efficacy of a variety of technologies including NaF-18 PET scans for identification of bone metastasis of cancer (Centers for Medicare & Medicaid Services, 2015). Likewise, in 2015, Harvard Pilgrim Health Care entered into a PBRSA with Amgen for its PCSK9 inhibitor known as evolocumab in order to assess real-world efficacy and performance and provide justification for the high price tag (Reinke, 2016). As final outcome data becomes increasingly scarce and costly, it makes sense to substitute towards surrogate endpoint evidence.

Future research on surrogate endpoints can be taken in a number of worthwhile directions. Our framework addressed early access to a drug, based on a surrogate endpoint signal of a final outcome benefit for which evidence was not yet available. In reality some trials follow multiple surrogates, or may produce an early and thus noisy signal of the final outcome to accompany the surrogate evidence, presenting a decision maker with the problem of how best to combine the signals. Where multiple surrogates are available but only some are reported in a trial, a decision maker will have to consider whether the reported measure is susceptible to any bias. In addition to clinical benefit, a trial is typically informative about safety and adverse events, and in general a decision maker will need to decide how to trade off favorable and unfavorable outcomes, based on their importance to patients and the comparative quality of the signals. A decision maker could be faced with the problem of prioritizing two drugs for the same disease. In such a situation, it is possible that a drug with a noisier signal merits priority, precisely because its greater variability in the final outcome creates option value in the form of upside risk (Sanchez et al., 2012).[6]

The last point also highlights the potential interaction between risk-preferences and decision making in this context. Our framework utilized a number of assumptions, including risk neutrality and in some instances joint normality of the surrogate and final outcome benefit. Earlier in the text, we conjectured that risk-aversion may induce greater caution about acting on low-quality signals. In this section, we also discussed how lower-quality signals might also be more valuable. Numerical analysis could be used to explore decision making in the presence of some degree of risk aversion, and under alternative distributional assumptions. In addition, our framework was parsimonious in the flow of information. In reality, evidence accumulates over the course of a trial. These dynamics confront a regulator and others with an ongoing decision process vis-à-vis drug access, based on the evidence thus far and the evidence still likely to come; complexities such as reversion to the mean would become relevant to decision making. Our simple dynamics also assumed that early treatment did not affect the value of treatment once final outcome evidence became available; effectively, there was no cost to reversing a decision to grant early access if the final outcome proved unfavorable. This assumption

[6] We thank an anonymous referee for this point.

may not be appropriate in some circumstances, and a number of studies have analyzed the impact of reversal costs on optimal decision making (Claxton et al., 2012; Claxton et al., 2016; Eckermann and Willan, 2007; Eckermann and Willan, 2008).

Future research on surrogate endpoints can be taken in other worthwhile directions. Our framework addressed early access to a drug, based on a surrogate endpoint signal of a final outcome benefit for which evidence was not yet available. In reality, some trials follow multiple surrogates. Others may produce early, noisy measures of the final outcome to accompany the surrogate evidence, presenting a decision maker with the problem of how best to combine the signals. In addition to clinical benefit, a trial is typically informative about safety and adverse events. When an effective drug also poses safety risks, decision makers will need to weigh the trade off and the comparative quality of the signals. Along another dimension, certain patients may value a surrogate endpoint independently of the final outcome; for example, even if progression-free survival were not predictive of overall survival, more advanced disease typically produces more severe symptoms.

This study has also taken the choice of surrogate endpoint and the structure of clinical trials as given. In some important cases, drug developers may be able to structure trials to their own advantage. One possibility is that multiple surrogates are available for the final outcome benefit of a particular drug. In such a situation, drug developers may possess superior information, and specify advantageous surrogates in designing trials. Regulators and payers would then have to consider the susceptibility of reported surrogates to bias. Another possibility is that a pharmaceutical company with multiple drugs in development for a disease may be more likely to bring to trial the drug that is expected to perform well with respect to a surrogate, recognizing that a superior surrogate signal will predict a superior final outcome benefit. The strategic design of trials, together with the appropriate response of regulators and payers, represent an important and interesting topic for further inquiry.

This study has focused on thresholds for drug access based on the magnitude of a surrogate endpoint signal, yet access may also impose a threshold for statistical significance of the trial evidence. In fact such a requirement can result in denial when access would have been welfare-enhancing (Claxton et al., 2001; Claxton, 1999a; Claxton, 1999b). Still, many real-world decision makers demand statistical significance. Our framework can naturally accommodate such a requirement. In some situations, the “signal threshold” for the magnitude of the surrogate data will be determinative, while in others the significance threshold is. As one would expect, the significance threshold is likely to play less of a role as the size of a trial grows.

The extended framework allows for investigation of the implications of imperfect clinical evidence for the design of clinical trials, and more generally, the optimal degree of investment into evidence quality (Claxton et al., 2015). The question becomes, in the context of the broader literature, the expected value of sample information (Claxton et al., 2001; Claxton, 1999a; Eckermann and Willan, 2007; Eckermann and Willan, 2008; Griffin et al., 2011), and speaks to how much money to invest in growing the size of a trial, given the cost of recruitment and the benefit of uncovering an improved signal. It is possible that a higher-quality surrogate would be a cost-effective substitute for a larger trial in some clinical situations. In any event, as a practical matter, continued investigation into patient registries could aid in the collection of both clinical benefits and safety events for long-term post-approval data.

The economic implications have not been fully explored, even as they have drawn an increasing amount of attention from researchers in the clinical and basic sciences. We hope our study represents a first step towards broader economic literature on surrogate endpoints as a source of imperfect information about clinical effects. Much remains to be done in order to evaluate how the advent of surrogate endpoints is likely to influence the behavior of key stakeholders in the healthcare system and how it should inform health policy.

Acknowledgements

Support for this research was provided by Eli Lilly & Company. Precision Health Economics is a consultancy to the health insurance and life science industries. Darius Lakdawalla is the Chief Scientific Officer and a stockholder at Precision Health Economics. John Romley is a consultant at Precision Health Economics. Technical assistance on model development was provided by Oliver Diaz. Administrative and editorial support was provided by Alison Silverstein, Chelsey Kamson, and Taylor Watson.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhealeco.2016.12.001.

References

Athey, S., 2002. Monotone comparative statics under uncertainty. Q. J. Econ., 187–223.
Beauchemin, C., et al., 2014. Progression-free survival as a potential surrogate for overall survival in metastatic breast cancer. Oncol. Targets Ther. 7, 1101–1110.
BlueCross BlueShield of Illinois, 2016. Standard Dispensing Limits (DL). July 24, 2016, Available from: http://www.bcbsil.com/PDF/rx/rx_dispensing_limits_std_il.pdf.
Blumenthal, G.M., et al., 2015. Overall response rate, progression-Free survival, and overall survival with targeted and standard therapies in advanced non–Small-Cell lung cancer: a US food and drug administration trial-Level and patient-Level analyses. J. Clin. Oncol.
Bolton, P., Dewatripont, M., 2005. Contract Theory. MIT press.
CVS Caremark, 2016. Value Formulary Effective as of 07/01/2016 [cited 2016 August 23]; Available from: https://www.caremark.com/portal/asset/Value_Formulary.pdf.
Carlson, J.J., et al., 2010. Linking payment to health outcomes: a taxonomy and examination of performance-based reimbursement schemes between healthcare payers and manufacturers. Health Policy 96 (3), 179–190.
Centers for Medicare & Medicaid Services, 2015. Decision Memo for Positron Emission Tomography (NaF-18) to Identify Bone Metastasis of Cancer (CAG-00065R2).
Cigna, 2016. Prescription Drug List Changes: Starting January 1., pp. 2016.
Claxton, K., et al., 2001. Bayesian value-of-information analysis. Int. J. Technol. Assess. Health Care 17 (01), 38–55.
Claxton, K., Sculpher, M., Drummond, M., 2002. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). Lancet 360 (9334), 711–715.
Claxton, K., Cohen, J.T., Neumann, P.J., 2005. When is evidence sufficient? Health Aff. (Millwood) 24 (1), 93–101.
Claxton, K., et al., 2012. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. Health Technol. Assess. 16 (46), 1–323.
Claxton, K., et al., 2015. How to estimate the health benefits of additional research and changing clinical practice. BMJ 351, h5987.
Claxton, K., et al., 2016. A comprehensive algorithm for approval of health technologies with, without, or only in research: the key principles for informing coverage decisions. Value Health.
Claxton, K., 1999a. Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals. Health Econ. 8 (3), 269–274.
Claxton, K., 1999b. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J. Health Econ. 18 (3), 341–364.
Eckermann, S., Willan, A.R., 2007. Expected value of information and decision making in HTA. Health Econ. 16 (2), 195–209.
Eckermann, S., Willan, A.R., 2008. The option value of delay in health technology assessment. Med. Decis. Making.
Express Scripts, 2016. Preferred Prescriptions Drug List [cited 2016 August 23]; Available from: https://www.express-scripts.com/preferredprescriptions.
Fleming, T.R., DeMets, D.L., 1996. Surrogate end points in clinical trials: are we being misled? Ann. Intern. Med. 125 (7), 605–613.
Garrison, L.P., et al., 2013. Performance-based risk-sharing arrangements—good practices for design, implementation, and evaluation: report of the ISPOR good practices for performance-based risk-sharing arrangements task force. Value Health 16 (5), 703–719.

Green, J.R., Scotchmer, S., 1995. On the division of profit in sequential innovation. Rand J. Econ. 26 (1), 20–33.
Greene, W.H., 2012. Econometric Analysis, 7 ed. Pearson Education.
Griffin, S.C., et al., 2011. Dangerous omissions: the consequences of ignoring decision uncertainty. Health Econ. 20 (2), 212–224.
Harvard Pilgrim Health Care, 2016. P&T Committee Updates 2014 July 24, Available from: https://www.harvardpilgrim.org/portal/page?_pageid=253,5297993&_dad=portal&_schema=PORTAL.
Hutton, J., et al., 2007. Coverage with evidence development: an examination of conceptual and policy issues. Int. J. Technol. Assess. Health Care 23 (4), 425–435.
Ingelsson, E., et al., 2007. Clinical utility of different lipid measures for prediction of coronary heart disease in men and women. JAMA 298 (7), 776–785.
Kim, C., Prasad, V., 2015. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US food and drug administration approvals. JAMA Intern. Med. 175 (12), 1992–1994.
Kim, C., Prasad, V., 2016. Strength of validation for surrogate end points used in the US food and drug administration's approval of oncology drugs. Mayo Clin. Proc.
Lakdawalla, D., Sood, N., 2013. Health insurance as a two-Part pricing contract. J. Public Econ. 102, 1–12.
Lakdawalla, D., Yin, W., 2015. Insurers' negotiating leverage and the external effects of Medicare part D. Rev. Econ. Stat. 97 (2), 314–331.
Lim, S., Quon, M.J., Koh, K.K., 2014. Modulation of adiponectin as a potential therapeutic strategy. Atherosclerosis 233 (2), 721–728.
Longworth, L., et al., 2013. When does NICE recommend the use of health technologies within a programme of evidence development?: a systematic review of NICE guidance. Pharmacoeconomics 31 (2), 137–149.
Loury, G.C., 1979. Market structure and innovation. Q. J. Econ. 93 (3), 395–410.
MPN Research Foundation, 2016. Primary Myelofibrosis (PMF) [cited 2016 August 23]; Available from: http://www.mpnresearchfoundation.org/Primary-Myelofibrosis.
Malani, A., Philipson, T.J., 2011. Can Medical Progress Be Sustained? Implications of the Link Between Development and Output Markets. National Bureau of Economic Research.
Manski, C.F., 2009. Adaptive partial drug approval: a health policy proposal. The Economists' Voice 6 (4) (Article 9).
Marinoni, G., 2012. Market Access Lessons from Benlysta's Failure to Wow European Payers [cited 2015 October 5]; Available from http://blog.ihs.com/market-access-lessons-from-benlystas-failure-to-wow-european-payers.
Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory, Vol. 1. Oxford university press, New York.
Mascarenhas, J., Hoffman, R., 2013. A comprehensive review and analysis of the effect of ruxolitinib therapy on the survival of patients with myelofibrosis. Blood 121 (24), 4832–4837.
McKenna, C., et al., 2015. Unifying research and reimbursement decisions: case studies demonstrating the sequence of assessment and judgments required. Value Health 18 (6), 865–875.
Medicare Payment Advisory Commission, Report to the Congress: Medicare Payment Policy. 2015: Washington, DC.
Michiels, S., et al., 2016. Progression-free survival as surrogate endpoint for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer. Ann. Oncol.
National Institute for Health and Care Excellence, 2013. Guide to the Methods of Technology Appraisal., pp. 2013 [cited 2016 November 7]; Available from: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781.
National Institutes of Health, 2015. NIH Joins Public-private Partnership to Fund Research on Autism Biomarkers.
Niewoehner, D.E., Collins, D., M.L.E.f.t.D.o, 2000. Veterans affairs cooperative study group, relation of FEV1 to clinical outcomes during exacerbations of chronic obstructive pulmonary disease. Am. J. Respir. Crit. Care Med. 161 (4), 1201–1205.
Niezen, M., et al., 2007. Conditional reimbursement within the Dutch drug policy. Health Policy 84 (1), 39–50.
Nordhaus, W.D., 1969. An economic theory of technological change. Am. Econ. Rev. 59 (2), 18–28.
Paoletti, X., et al., 2013. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: a meta-analysis. J. Natl. Cancer Inst. 105 (21), 1667–1670.
Paul, G.D., David, M., 2014. Validating forced expiratory volume in one second As a surrogate endpoint for death or hospitalization In randomized controlled trials of chronic obstructive pulmonary disease or cystic fibrosis, in D43. COPD: FACTORS ASSOCIATED WITH THE IMPLEMENTATION OF GOLD AND OTHER ORGANIZATIONS' RECOMMENDATIONS. Am. Thorac. Soc., A5968.
Prasad, V., et al., 2015. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-Level meta-analyses. JAMA Intern. Med. 175 (8), 1389–1398.
Reinke, T., 2016. Drug Companies Ask Plans To Just Say 'Yes' to Risk [cited 2016 August 18]; Available from: http://www.managedcaremag.com/archives/2016/4/drug-companies-ask-plans-just-say-yes-risk.
Sanchez, Y., et al., 2012. The option value of innovative treatments in the context of chronic myeloid leukemia. Am. J. Manag. Care 18 (11 Suppl), S265–71.
Sanders, G.D., et al., 2016. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. JAMA 316 (10), 1093–1103.
Shafrin, J., et al., 2016. The value of surrogate endpoints for predicting real-world survival across five cancer types. Curr. Med. Res. Opin. 32 (4), 731–739.
Smith, J.P., 2015. Application number: 125559Orig1s000 summary review. Center Drug Eval. Res.
Soares, H.P., et al., 2005. Evaluation of new treatments in radiation oncology: are they better than standard treatments? JAMA 293 (8), 970–978.
The Biomarkers Consortium, 2011. About Us [cited 2016 August 28]; Available from: http://www.biomarkersconsortium.org/about.php.
US Food and Drug Administration, 2011a. Ruxolitinib Phosphate Summary Review 2016, Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/202192Orig1s000SumR.pdf.
US Food and Drug Administration, 2011b. Proposal to Withdraw Approval for Breast Cancer Indication for AVASTIN (bevacizumab): Decision of the Commisioner [cited 2015 October 8]; Available from: http://www.fda.gov/downloads/NewsEvents/Newsroom/UCM280546.pdf.
US Food and Drug Administration, 2011c. FDA Commissioner Removes Breast Cancer Indication from Avastin Label [cited 2016 August 21]; Available from: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm279485.htm.
US Food and Drug Administration, 2015. Novel Drugs Approved Using Surrogate Endpoints. 2016, Available from: http://www.fda.gov/downloads/NewsEvents/Testimony/UCM445375.pdf.
US Food and Drug Administration, 2016. Oncologic Drugs Advisory Committee (ODAC) Meeting: Tuesday 2016. Center for Drug Evaluation and Research.
US Government Accountability Office, New Drug Approval: FDA Needs to Enhance Its Oversight of Drugs Approved on the Basis of Surrogate Endpoints. 2009: Washington, DC.
Verstovsek, S., et al., 2012a. A double-Blind: placebo-Controlled trial of ruxolitinib for myelofibrosis. New Engl. J. Med. 366 (9), 799–807.
Verstovsek, S., et al., 2012b. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. Blood 120 (6), 1202–1209.
Vestbo, J., et al., 2008. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). Eur. Respir. J. 31 (4), 869–873.
Viscusi, W.K., Zeckhauser, R.J., 2015. Regulating ambiguous risks: the less than rational regulation of pharmaceuticals. J. Legal Stud. 44 (S2), S387–S422.