An Economic Theory of Statistical Testing[*]

Aleksey Tetenov[†]

August 21, 2016

Abstract

This paper models the use of statistical hypothesis testing in regulatory approval. A privately informed agent proposes an innovation. Its approval is beneficial to the proponent, but potentially detrimental to the regulator. The proponent can conduct a costly clinical trial to persuade the regulator. I show that the regulator can screen out all ex-ante undesirable proponents by committing to use a simple statistical test. Its level is the ratio of the trial cost to the proponent's benefit from approval. In application to new drug approval, this level is around 15% for an average Phase III clinical trial.

The practice of statistical hypothesis testing has been widely criticized across the many fields that use it. Examples of such criticism are Cohen (1994), Johnson (1999), Ziliak and McCloskey (2008), and Wasserstein and Lazar (2016). While conventional test levels of 5% and 1% are widely agreed upon, these numbers lack any substantive motivation. In spite of their arbitrary choice, they affect thousands of influential decisions. The hypothesis testing criterion has an unusual lexicographic structure: first, ensure that the

probability of Type I errors does not exceed a given conventional level under the null hypothesis, then do the best you can to reduce the probability of Type II errors. It is difficult to motivate this structure by considering statistical decision problems. Instead, both Bayesian and frequentist criteria in some way seek to balance the probabilities of both Type I and Type II errors.

This paper focuses on one of the most important contexts in which hypothesis testing is used: regulatory approval of innovations like new pharmaceuticals and medical devices. It is prescribed, for example, in the international guidelines for drug approval (International Committee on Harmonisation, 1999). Hypothesis testing is also used less formally when evidence in favor of some proposed policy is judged on the basis of being "statistically significant." Viewed as decision rules for choosing between two alternative treatments, hypothesis tests with standard significance levels are strongly biased in favor of the status quo. When an immediate choice between two options must be made, Simon (1945) argued that it would be more sensible simply to choose the one favored by the available evidence, even if only by a small margin. This decision rule is essentially a one-sided test with 50% level. A recent line of literature applying the minimax regret criterion to treatment choice problems reaches similar conclusions (Manski, 2004; Manski and Tetenov, 2007; Schlag, 2007; Hirano and Porter, 2009; Stoye, 2009). Decision rules similar to hypothesis tests could be motivated by loss aversion under average risk (Hirano and Porter, 2009) or minimax regret (Tetenov, 2012) criteria, but obtaining tests with conventional levels in this framework requires extreme degrees of loss aversion (Tetenov, 2012).

I study regulatory hypothesis tests as a strategy in a game against economically motivated proponents, rather than in a game against nature. The regulator's statistical decision rule creates incentives that can affect the pool of innovations that are proposed and tested. This important consideration is absent in the analysis of games against nature. In many applications of statistical testing, proponents of innovations have private information about their quality that is not accessible to the regulator. Proponents often stand to benefit from the approval decision even if the innovation is inferior from the regulator's perspective.

While the supply of positive innovations is limited, it may be reasonable to consider the pool of potential bad innovations to be almost unbounded. Pharmaceutical development process, for example, generates thousands of candidate compounds that can plausibly be effective for a given condition. Similarly, one could propose many plausible educational reforms or poverty-reducing programs. An "innovator" may know ex ante that the proposal was chosen with no substantive basis, but reliance on statistical evidence entails that with some probability the evidence will turn out to be strong enough for the proposal to be accepted. The regulator could incur great losses if her statistical tests do not deter such frivolous "innovators" from trying their luck.

I restrict attention to the problem of statistical testing of innovations when the only tool available to the regulator is to accept or reject the innovation based on credible data coming from a *costly* trial with a known statistical structure. The regulator can deter the proponents of inferior innovations from seeking approval by *committing* to a statistical test. A sufficiently strict test ensures that the probability of Type I errors (acceptance of bad innovations) is smaller than the ratio between the proponent's cost of collecting the evidence (e.g., clinical trials) and the proponent's benefit from the regulator's acceptance. I show that hypothesis tests at this level are optimal for a regulator having *maximin utility*[1] with ambiguity regarding the quality of potential proposals. This framework provides a decision-theoretic rationale for using one-sided hypothesis testing in statistical treatment choice problems and a rule for choosing the test level.

I show that the same solution is obtained in the limit case if the regulator places a prior distribution over the quality of potential proposals. As the probability assigned to inferior innovators increases, the Bayesian regulator's optimal policy converges to the same simple one-sided hypothesis test rule that deters the testing of all inferior proposals.

Statistical testing is harder to analyze if the proponents themselves are imperfectly informed about the quality of their innovations. I show that the maximin property of the proposed hypothesis tests extends to a more general setting in which proponents may have arbitrary prior beliefs about a real-valued parameter θ capturing the innovation's value

---

[1]Maximin refers to maximizing the minimum expected payoff. The criterion is usually called *minimax* in Statistics, referring to minimizing maximum expected loss.

to the regulator. This result holds if the proponent's payoff from approval is concave and increasing in θ for a class of distributions of the trial's outcome (which includes normal and exponential distributions).

The model in this paper crudely resembles the regulatory approval of new drugs after large Phase III clinical trials. I use the available aggregate data on the costs of clinical trials and on the profitability of approved drugs to gauge what statistical test levels would be sufficient to deter frivolous proponents in this domain. This level is around 15% for a drug with average profits and a trial with average costs.

In this paper the proponents have an incentive to reveal some of their private information about the innovation through their decision to invest in a costly trial. This leads to non-degenerate approval decisions even if the regulator is infinitely pessimistic about the pool of potential proposals. In contrast, both parties share the same ex ante beliefs about the innovation in the Bayesian persuasion model of Kamenica and Gentzkow (2011), which studies the proponent's choice of an optimal signal structure, and in the sequential information acquisition model of Henry and Ottaviani (2015), which studies different allocations of decision-making power between the two parties. Manski (2015) advocates using randomized regulatory approval for deterring unwanted applications (as well as for diversification and learning). In contrast to the present paper, randomization of approval proposed by Manski does not depend on an informative statistical signal.

This paper does not consider a number of other incentive aspects involved in the conduct of clinical trials and experiments that have been recently analyzed by economists. Chassang et al. (2012) propose incentive mechanisms to disentangle the effects of the treatment from the effects of effort exerted by trial participants to utilize the treatment. Di Tillio et al. (2016) in a very simple model consider the effects of selective sampling or selective reporting of trial results when such manipulation by the proponent is possible.

The paper proceeds as follows: in the next section, I discuss why the hypothesis testing criterion has been difficult to rationalize using statistical decision theory. Section 2 outlines a simple environment that motivates hypothesis test rules as an optimal strategy for a regulator with maximin utility. Section 3 adds the assumption that the data satisfies

the monotone likelihood ratio property, under which decision procedures have a simple threshold form. In this setting, I show that the decisions of a Bayesian regulator converge to a simple hypothesis test rule as the regulator's beliefs become more pessimistic about the pool of potential proposals. Section 4 extends the problem to allow proponents of innovations to be uncertain about their effectiveness prior to collecting the evidence. Section 5 deals with strategic choice of the cost and precision of evidence. Section 6 uses data on the costs of clinical trials and the profitability of approved drugs to gauge what deterrent test levels may be appropriate for Phase III clinical trials.

1   Hypothesis testing is difficult to rationalize

Though widely used and intuitively appealing to many researchers, the classical hypothesis testing criterion is difficult to motivate as a solution to an explicit decision problem. The criterion is succinctly summarized by Lehmann and Romano (2005, p. 57): "It is customary therefore to assign a bound to the probability of incorrectly rejecting H when it is true and to attempt to minimize the other probability subject to this condition." When testing is used to choose between two alternative treatments, which is the focus of this paper, the hypothesis testing criterion raises two questions. First, why is constraining the maximum probability of Type I errors lexicographically more important than minimizing the chance of Type II errors? While one type of errors could be more important than the other, typical decision criteria call for weighted consideration of both, rather than a constraint on just one type. Second, why only the probability of Type I errors is considered and not their magnitude? It would seem that the probability of mistakenly choosing the wrong treatment should be more important if there is a large difference in the effectiveness of the two, and less important if they are almost equally effective. This paper proposes an answer to both of these questions.

Statistical decision problems have been analyzed as *games against nature* starting with Wald (1950) and Savage (1954). First, nature "chooses" an unknown parameter value $\theta \in \Theta$. In this paper, θ will refer to the net effect of a new treatment, with the

net effect of the status quo treatment normalized to zero. Then, statistical data $X \in \mathcal{X}$ is randomly drawn from a probability distribution $F(X; \theta)$ that depends on θ. The statistician then makes a decision $\delta(X)$, possibly randomized, based on observed data. In case of hypothesis testing, the decision is binary: $\delta(X) = 1$ if the alternative treatment is accepted and $\delta(X) = 0$ if the status quo is chosen. The performance of a decision rule under each possible parameter value $\theta \in \Theta$ could be summarized by the average loss

$$(1) \qquad E_{F(X;\theta)}[L(\delta(X), \theta)].$$

Usually, no decision rule minimizes the average loss simultaneously for all values of θ. The choice of δ then depends on the criterion used to deal with the ambiguity regarding the value of θ. The statistician could place a subjective prior distribution μ on Θ and minimize subjective expected loss $E_{\mu(\theta)}[E_{F(X;\theta)}[L(\delta(X), \theta)]]$. Alternatively, the statistician could look for a decision rule that performs uniformly well over Θ, for example, minimizing $\sup_{\theta \in \Theta} E_{F(X;\theta)}[L(\delta(X), \theta)]$ (the maximin criterion). Stoye (2011) provides an extensive overview of various uniform decision criteria and their axiomatic properties.

Discussions of hypothesis testing sometimes invoke a 1–K loss function that penalizes all Type I errors by K points and all Type II errors by 1 point,

$$(2) \qquad L(\delta, \theta) = \delta \cdot K \cdot I(\theta \in \Theta_0) + (1 - \delta) \cdot I(\theta \notin \Theta_0).$$

Under some assumptions, one-sided hypothesis tests with size $\alpha = 1/(K + 1)$ coincide with minimax decision rules under the 1–K loss. For example, tests with 5% significance level minimize maximum 1–K loss that places 19 times more weight on a Type I error than on a Type II error. However, loss function (2) does not provide a good rationalization for the use of classical hypothesis testing criterion in treatment choice. First, while decision rules that arise from minimax and Bayes criteria sometimes coincide with hypothesis test rules, none of the criteria are equivalent to the lexicographic hypothesis testing criterion found in statistics and econometrics textbooks. Second, the Type I/II error loss ratios corresponding to typical test levels (K=19 for 5% and K=99 for 1%) seem too large for

many problems. Furthermore, the 1–K loss function ignores the substantive magnitude of committed errors, assigning the same penalty for mistakenly approving treatments that are only infinitesimally worse as for ones that are greatly inferior.

A number of recent papers consider treatment choice as a statistical decision problem using the regret loss function

(3) $$L(\delta, \theta) = \delta \cdot |\theta| \cdot I(\theta < 0) + (1 - \delta) \cdot |\theta| \cdot I(\theta > 0),$$

which penalizes both Type I and Type II errors proportionately to the magnitude of the treatment effect $|\theta|$. The optimal decision rules under both minimax and average risk (flat prior) decision criteria are essentially *empirical success rules* (Manski, 2004, 2005) that prescribe choosing the treatment that appears to be more successful in trials, whether by a small or by a wide margin (see also Schlag, 2007; Hirano and Porter, 2009; Stoye, 2009). These decision rules are comparable to one-sided 50% hypothesis tests, rather than to the conventional 5% or 1% tests.

The regret loss function (3) could be modified to place asymmetric weight on Type I and Type II errors of the same magnitude:

(4) $$L(\delta, \theta) = \delta \cdot K \cdot |\theta| \cdot I(\theta < 0) + (1 - \delta) \cdot |\theta| \cdot I(\theta > 0).$$

Minimax and average risk decision rules with this asymmetric loss function bear similarity to hypothesis test rules because innovations are approved when the estimate of θ exceeds some multiple of its standard error (Hirano and Porter, 2009; Tetenov, 2012). However, to obtain decision rules comparable to tests at conventional levels, the asymmetry factor K has to be much greater than with the 1-K loss function (2). The difference is due to the interaction between the magnitude of errors and their probability. One-sided 5% tests are minimax optimal for K=102, while 1% tests are optimal for K=970 (Tetenov, 2012). In contrast, a moderate loss aversion coefficient of K=3 would lead to a one-sided 34% test. While loss aversion may seem like a plausible explanation for the asymmetry of conventional hypothesis tests, it cannot easily rationalize conventional test levels or

the hypothesis testing criterion itself.

2 Screening proponents through hypothesis testing

This section describes a regulatory approval game from which the classical hypothesis testing criterion (with problem-specific significance levels) emerges as an optimal strategy for a regulator with maximin utility playing against a perfectly informed proponent. This result does not require any restrictions on the family of distributions generating the trial data. Further sections impose the *monotone likelihood ratio property*, which allows for a richer theoretical analysis at the expense of statistical generality.

There are two parties in the game: the *proponent* of an innovation and the *regulator*. The regulator has to decide whether to accept or reject the proposed innovation based on costly statistical evidence produced by the proponent. The quality of the proponent's innovation, which determines the parties' payoffs, is $\theta \in \Theta$. The proponent knows θ, while the regulator does not (section 4 extends the findings to the case in which the proponent is uncertain about the value of θ). Monetary transfers between the parties are not possible.

If the regulator accepts the innovation, the proponent gets a payoff of $b(\theta)$ and the regulator gets a payoff of $v(\theta)$. The payoffs to both parties are zero if the regulator rejects the innovation. The function $b(\theta)$ is known to the regulator. When the proponent knows θ with certainty, we could assume that $b(\theta) \geq 0$ for all $\theta \in \Theta$, since for other proponents participation is certainly unprofitable. I assume throughout the paper that both parties are risk-neutral with respect to objectively known probabilities.

Let $\Theta_0 \equiv \{\theta : v(\theta) < 0\}$ denote the set of innovations that are detrimental for the regulator (the *null hypothesis*) and let $\Theta_1 \equiv \{\theta : v(\theta) \geq 0\}$ be the set of innovations valuable for the regulator (the *alternative hypothesis*).

To convince the regulator that $\theta \in \Theta_1$, the proponent could conduct a trial of the innovation that costs $c > 0$ and generates a sufficient statistic $X \in \mathcal{X}$ with probability distribution $F(X; \theta)$. The conditional data generating process $F(X; \theta)$ is known to both

parties. The trial cannot be manipulated by the proponent and only its full results could be provided to the regulator. The full trial cost c is sunk before any data is realized and c is observed by the regulator. The regulator commits to using a *statistical decision rule* $\delta : \mathcal{X} \to [0, 1]$, where $\delta(X)$ denotes the probability with which the regulator will accept the innovation if the outcome of the trial equals X.

To summarize, the timing of the game is as follows. The regulator commits to an approval decision rule δ as a function of the trial statistic X (and of the publicly known characteristics $b(\theta)$ and c of the proponent). The proponent, knowing his type θ, chooses whether to invest c in collecting the data. If the proponent chooses to conduct the trial, nature draws X according to the distribution $F(X; \theta)$. The proponent chooses whether to request approval and the regulator grants it with probability $\delta(X)$. The parties' final payoffs are $(b(\theta) - c, v(\theta))$ if the proponent conducts a trial and the regulator approves, $(-c, 0)$ if the proponent conducts a trial and the regulator does not approve, and $(0, 0)$ otherwise.

The ex ante probability that an innovation with parameter θ will be accepted following the trial equals

$$(5) \qquad \beta_\delta(\theta) \equiv \int_{\mathcal{X}} \delta(X) dF(X; \theta).$$

In statistical terminology, $\beta_\delta(\theta)$ is the test's "power function," since $\delta(X) = 1$ denotes rejection of the null hypothesis by the regulator. It is optimal for a risk-neutral proponent to invest in a trial if its expected payoff is positive, that is, if $\beta_\delta(\theta) > c/b(\theta)$. Proponents of type θ who face approval probability $\beta_\delta(\theta) < c/b(\theta)$ are deterred by the regulator's statistical decision rule from conducting a trial. Proponents could make either decision if $\beta_\delta(\theta) = c/b(\theta)$. To simplify exposition, I assume that indifferent proponents invest in a trial, but this is not significant for the paper's analysis.

Since the regulator moves first and commits, her choice of a statistical decision rule δ could be analyzed as a single agent decision problem, taking into account the proponent's best response to δ. The regulator's expected payoff when facing a proponent of type θ

equals

$$(6) \qquad V(\delta, \theta) \equiv \begin{cases} v(\theta)\beta_\delta(\theta) & \text{if} \quad \beta_\delta(\theta) \geq c/b(\theta), \\ 0 & \text{if} \quad \beta_\delta(\theta) < c/b(\theta). \end{cases}$$

To achieve the maximum feasible payoff of zero for θ such that $v(\theta) < 0$, it is sufficient for the regulator to pick any decision rule with $\beta_\delta(\theta) < c/b(\theta)$. To achieve the same objective in a game against nature (in which the decision to generate the trial data is not strategic), the regulator would have to pick a decision rule with $\beta_\delta(\theta) = 0$. Only degenerate decision rules that reject innovations for all realizations of the data are maximin in a game against nature, which led Manski (2004) to question the usefulness of the maximin decision criterion. In contrast, strategic entry decision by the proponent allows the regulator to employ non-degenerate decision rules even under the pessimistic maximin criterion.

The following proposition characterizes optimal strategies for a regulator with maximin preferences.

Proposition 1. *Decision rule δ^* is maximin-optimal for the regulator, i.e.,*

$$\delta^* \in \arg\max_\delta \min_\theta V(\delta, \theta)$$

if and only if

$$(7) \qquad \beta_{\delta^*}(\theta) < \frac{c}{b(\theta)} \text{ for all } \theta \in \Theta_0.$$

Proof. The maximin payoff for the regulator cannot be greater than zero, since zero is the highest possible payoff for $\theta \in \Theta_0$. Then the maximin payoff has to equal zero, because a degenerate decision rule $\delta_0(x) = 0 \ \forall x \in \mathcal{X}$, which always rejects the innovation, yields the regulator a payoff of zero playing against any proponent type.

If δ^* satisfies (7), then the regulator's payoff equals zero for all $\theta \in \Theta_0$ by (6). The regulator's payoff is always non-negative for $\theta \in \Theta_1$, hence δ^* is maximin.

If δ^* does not satisfy (7), then $\beta_{\delta^*}(\tilde\theta) \geq c/b(\tilde\theta)$ for some $\tilde\theta \in \Theta_0$ and the regulator's

expected payoff equals $v(\tilde{\theta})\beta_{\delta^*}(\tilde{\theta}) < 0$, hence δ^* is not maximin. $\qquad\qquad\square$

If the proponent's payoff does not depend on θ (i.e., $b(\theta) = b$), then condition (7) in Proposition 1 simplifies to the standard condition for controlling the level of a hypothesis test:

$$(8) \qquad\qquad\qquad \beta_{\delta^*}(\theta) < \frac{c}{b} \text{ for all } \theta \in \Theta_0,$$

with the requisite test level determined by the economic parameters b and c.

There are generally many alternative maximin decision rules, including the degenerate rule $\delta_0(x) = 0$. Some of these rules could be *inadmissible* (weakly or strongly dominated by other decision rules). More structure needs to be placed on the data distribution $F(X; \theta)$ to determine which decision rules are admissible. Further sections impose the monotone likelihood ratio property on F, which yields a very strong characterization of admissible decision rules.

Maximin is a conservative decision criterion for choice under ambiguity. In this case, the regulator's ambiguity about the distribution of θ among potential proponents. The next section compares maximin decision rules to optimal decision rules for regulators who do not face ambiguity and could place a subjective prior distribution on the potential proponent type θ. I show that the maximin decision rule could be seen as a limit case when the prior probability of the null hypothesis ($\theta \in \Theta_0$) converges to one.

Maximin decision rules are optimal (or nearly optimal) if the regulator believes that the supply of inferior innovations (proponents with $\theta \in \Theta_0$) is almost bottomless compared to the supply of valuable ones. In many contexts, a large number of potential proposals with $v(\theta) < 0$ could be generated almost effortlessly and ultra-pessimistic beliefs are reasonable. The relevant distribution of proponent types is the distribution of *potential* proponents, which is difficult to observe in practice. If some deterrent policy is already in place (which is the case in all fields using some form of hypothesis testing), the distribution of deterred proponents is completely hidden from the view of the regulator. This makes it impossible for the regulator to base her beliefs about the distribution of

11

potential proponent types on empirical evidence and provides an additional reason to consider robust decision criteria like maximin.

3 Proponents with precise knowledge of θ

To better understand the properties of hypothesis testing, in this section I place additional structure on the payoffs and on the statistical properties of the data-generating process.

Assumption P1 (payoffs).

The quality of an innovation is indexed by a real-valued parameter $\theta \in \Theta \subseteq \mathbb{R}$ and $\Theta = (\theta_l; \theta_u)$, $-\infty \leq \theta_l < 0$, $0 < \theta_u \leq \infty$. θ fully captures the innovation's value to the regulator: $v(\theta) = \theta$. The proponent's benefit from an acceptance decision $b(\theta) > 0$ is a continuous, non-decreasing function of θ.

Assumption S1 (data).

There is a continuously distributed sufficient statistic $X \in \mathcal{X} \subseteq \mathbb{R}$ of the trial data with density $f(x; \theta) > 0$ and c.d.f. $F(x; \theta) = P_\theta(X \leq x)$. As a function of θ, $F(x; \theta)$ is continuously differentiable and strictly decreasing for all $x \in \mathcal{X}$. $f(x; \theta)$ possesses the Monotone Likelihood Ratio property:

$$(9) \qquad x_1 > x_2, \theta_1 > \theta_2 \Rightarrow \frac{f(x_1; \theta_1)}{f(x_1; \theta_2)} \geq \frac{f(x_2; \theta_1)}{f(x_2; \theta_2)}.$$

$f(x; \theta)$ is bounded for each θ, and $\lim_{\theta' \to \theta} \int_{\mathcal{X}} |f(x, \theta') - f(x, \theta)| dx = 0$ for all $\theta \in \Theta$.

A normally distributed statistic $X \sim \mathcal{N}(\theta, \sigma^2)$ with a fixed variance σ^2 and $F(x; \theta) = \Phi((x - \theta)/\sigma)$ is a leading example satisfying Assumption S1.

Under assumptions P1 and S1, it is sufficient for the regulator to consider *monotone* (threshold) decision rules[2]:

$$(10) \qquad \delta_T(X) \equiv \mathrm{I}[X \geq T], \ T \in \mathbb{R}.$$

---

[2]The class of monotone decision rules also allows for randomization at the threshold value T, but it is not necessary to consider it here for continuously distributed X.

Lemma 3 in Karlin and Rubin (1956) establishes that for any decision rule $\delta : \mathcal{X} \to [0,1]$ there exists a unique monotone decision rule δ_T that yields higher approval probability for all good innovations and lower approval probability for bad ones,

(11)
$$\beta_{\delta_T}(\theta) \leq \beta_\delta(\theta) \text{ for all } \theta \leq 0 \text{ and}$$
$$\beta_{\delta_T}(\theta) \geq \beta_\delta(\theta) \text{ for all } \theta \geq 0.$$

The proponent's decision to conduct a trial is then higher under δ_T for positive θ and lower for negative θ,

(12)
$$\mathrm{I}\left[\beta_{\delta_T}(\theta) \cdot b(\theta) \geq c\right] \leq \mathrm{I}\left[\beta_\delta(\theta) \cdot b(\theta) \geq c\right] \text{ for all } \theta \leq 0 \text{ and}$$
$$\mathrm{I}\left[\beta_{\delta_T}(\theta) \cdot b(\theta) \geq c\right] \geq \mathrm{I}\left[\beta_\delta(\theta) \cdot b(\theta) \geq c\right] \text{ for all } \theta \geq 0.$$

The regulator's payoff (6) from the threshold decision rule δ_T is therefore at least as great as the payoff from δ for all θ.

The acceptance probability $\beta_{\delta_T}(\theta) = 1 - F(T; \theta)$ of a monotone decision rule is continuously decreasing in the threshold T for every θ. There exists, then, a monotone decision rule

(13)
$$\delta^* \equiv \mathrm{I}[X \geq T^*],$$
$$T^* \equiv F^{-1}\left(1 - \frac{c}{b(0)}; 0\right),$$

whose acceptance probability at $\theta = 0$ equals

(14)
$$\beta_{\delta^*}(0) = \frac{c}{b(0)}.$$

The threshold T^* is unique because the c.d.f. $F(x; 0)$ is strictly increasing in x. This decision rule is the same as a one-sided hypothesis test of $H_0 : \theta \leq 0$ with size $c/b(0)$. For the normal case $X \sim \mathcal{N}(\theta, \sigma^2)$,

(15)
$$\delta^*(X) = \mathrm{I}\left[X \geq \sigma \Phi^{-1}\left(1 - \frac{c}{b(0)}\right)\right].$$

Under decision rule δ^*, all innovations with $\theta < 0$ face acceptance probability that is too low to make trials profitable for their proponents, whereas the acceptance probability for innovations with $\theta > 0$ is sufficiently high. The following proposition shows that δ^* is a maximin and admissible decision rule for the regulator. Monotone decision rules with higher acceptance probability at $\theta = 0$ are not maximin because they make it profitable for proponents with some $\theta < 0$ to conduct trials and the regulator faces a negative expected payoff against this type of proponents. Decision rules with lower acceptance probability at $\theta = 0$ are also maximin, but they are inadmissible because they yield the same payoff as δ^* for $\theta \leq 0$, but lower payoff than δ^* for $\theta > 0$.

Proposition 2. *Monotone decision rule δ^* satisfying (14) is a maximin and admissible decision rule for the regulator under assumptions P1 and S1.*

Proof. The maximin payoff for the regulator equals zero. It cannot be lower because degenerate decision rule $\delta_0(x) = 0$ yields $V(\delta_0, \theta) = 0$ for all θ. It cannot be higher because $V(\delta, \theta) \leq 0$ for any δ and any $\theta < 0$.

When the regulator commits to decision rule δ^*, it is not optimal for potential proponents with $\theta < 0$ to conduct trials because

$$(16) \qquad\qquad \beta_{\delta^*}(\theta)b(\theta) < \beta_{\delta^*}(0)b(0) = c.$$

This inequality follows from strict monotonicity of $\beta_{\delta^*}(\theta)$ and weak monotonicity of $b(\theta)$. The regulator's payoff under δ^* from facing proponents with $\theta < 0$ then equals the maximin value of zero. For proponents with $\theta \geq 0$, the regulator's payoff is nonnegative for any decision rule, hence δ^* is maximin.

Decision rule δ^* is admissible if there exists no decision rule δ' that yields the regulator a strictly higher payoff for at least one value of θ and weakly higher payoffs for all values of θ. If any decision rule dominates δ^*, there must also be a monotone decision rule δ' that dominates δ^*.

If $\beta_{\delta'}(0)b(0) > c$, then $\beta_{\delta'}(\tilde{\theta})b(\tilde{\theta}) > c$ for some $\tilde{\theta} < 0$ because $\beta_{\delta'}(\theta)$ is continuous in θ (since $|\beta_{\delta'}(\theta') - \beta_{\delta'}(\theta)| \leq \int |f(x, \theta') - f(x, \theta)| dx \to 0$ for $\theta' \to \theta$). This makes it

profitable for proponents of type $\tilde{\theta}$ to invest in trials. The regulator's expected payoff is then negative at $\tilde{\theta}$, hence δ' cannot dominate δ^*.

If $\beta_{\delta'}(0)b(0) < c$, then $\beta_{\delta'}(\tilde{\theta})b(\tilde{\theta}) < c$ for some $\tilde{\theta} > 0$, hence it is unprofitable for proponents of type $\tilde{\theta}$ to conduct trials and the regulator's expected payoff at $\tilde{\theta}$ is zero. Under δ^*, instead, it is profitable for proponents with $\tilde{\theta} > 0$ to conduct trials, yielding a strictly positive expected payoff $V(\delta^*, \tilde{\theta}) > 0$ to the regulator. Thus δ' cannot dominate δ^*. $\qquad\square$

3.1 Decision rules for Bayesian regulators

I show below how hypothesis test rules with test level $c/b(0)$ relate to decision rules that are optimal for a Bayesian regulator who has a subjective prior distribution over the types of potential proponents. The first finding is that the test with level $c/b(0)$ always sets a higher threshold of evidence than a Bayesian regulator. Second, this test rule is a limit of decision rules adopted by Bayesian regulators who assume higher and higher proportion of potential proponents with inferior innovations.

Let $Q(\theta)$ denote the regulator's prior distribution over the potential proponent's type $\theta \in \Theta \subseteq \mathbb{R}$. For simplicity, assume that Q is continuous and its density function is $q(\theta)$. The regulator's expected payoff from using decision rule δ then equals

$$(17) \qquad V_Q(\delta) \equiv \int_\Theta V(\delta, \theta) dQ(\theta) = \int_\Theta \theta \beta_\delta(\theta) \cdot \mathrm{I}\left[\beta_\delta(\theta)b(\theta) \geq c\right] dQ(\theta).$$

It is sufficient to consider maximizing (17) over the set of threshold decision rules (10). Since $\beta_{\delta_T}(\theta)b(\theta)$ is increasing in θ, the subset of Θ on which $\beta_{\delta_T}(\theta)b(\theta) \geq c$ is an interval $[\bar{\theta}(T), \theta_u)$, where $\bar{\theta}(T)$ is the proponent's participation threshold. The regulator faces a one-dimensional problem of finding the threshold of the optimal decision rule

$$(18) \qquad \begin{aligned} \delta_Q &\equiv \mathrm{I}[X \geq T_Q], \\ T_Q &\equiv \arg\max_T V_Q(\delta_T) = \arg\max_T \int_{\bar{\theta}(T)}^{\theta_u} \theta \beta_{\delta_T}(\theta) dQ(\theta). \end{aligned}$$

Proposition 3 shows that it is optimal for a Bayesian regulator to set the evidence

15

threshold T_Q lower than for the hypothesis test (13). While some proponents with inferior innovations ($\theta < 0$) will then find it optimal to conduct trials and will gain approval with a positive probability, the loss from approving them is offset by a higher probability of approving valuable innovations ($\theta > 0$).

Proposition 3. *If Assumptions P1 and S1 hold, the regulator's prior Q places a positive probability $Q(\theta > 0) > 0$ on good potential proposals and has a bounded density $q(\theta) \leq \bar{q}$, then the optimal decision rule (18) has a lower threshold than the hypothesis test rule (13): $T_Q < T^*$.*

Proof. The hypothesis test threshold T^* is constructed so that $\beta_{\delta_{T^*}}(0)b(0) = c$, therefore $\bar{\theta}(T^*) = 0$ (only proponents with $\theta \geq 0$ find it profitable to conduct a trial). The participation threshold $\bar{\theta}(T)$ is an increasing differentiable function of T by the implicit function theorem.

The regulator's optimal decision rule cannot have $T_Q > T^*$ because for any $T > T^*$,

$$
V_Q(\delta_T) = \int_{\bar{\theta}(T)}^{\theta_u} \theta \beta_{\delta_T}(\theta) dQ(\theta) \leq \int_0^{\theta_u} \theta \beta_{\delta_T}(\theta) dQ(\theta) < \int_0^{\theta_u} \theta \beta_{\delta_{T^*}}(\theta) dQ(\theta) = V_Q(\delta_{T^*}),
$$

since $\theta \beta_{\delta_T} \geq 0$ for $\theta \geq 0$ and $\beta_{\delta_T}(\theta)$ is strictly decreasing in T for all θ.

For $T < T^*$, the regulator's expected payoff equals

$$
(19) \qquad V_Q(\delta_T) = \int_{\bar{\theta}(T)}^0 \theta \beta_{\delta_T}(\theta) dQ(\theta) + \int_0^{\theta_u} \theta \beta_{\delta_T}(\theta) dQ(\theta).
$$

The first term in (19) is negative and is bounded by

$$
\left| \int_{\bar{\theta}(T)}^0 \theta \beta_{\delta_T}(\theta) q(\theta) d\theta \right| \leq \bar{q} \left| \int_{\bar{\theta}(T)}^0 \theta d\theta \right| = \bar{q} \frac{(\bar{\theta}(T))^2}{2}
$$

because $\beta_{\delta_T} \in [0, 1]$ and $q(\theta) \in [0, \bar{q}]$. It follows that the derivative of the first term of (19) (the effect on losses from approving inferior innovations) with respect to $\bar{\theta}(T)$ at $\bar{\theta}(T) = \bar{\theta}(T^*) = 0$ equals zero. Since $\bar{\theta}(T)$ is differentiable in T,

$$
\frac{d}{dT} \int_{\bar{\theta}(T)}^0 \theta \beta_{\delta_T}(\theta) dQ(\theta) \bigg|_{T=T^*} = 0.
$$

On the other hand, since $\beta_{\delta_T}(\theta)$ is strictly decreasing in T and Q places a positive measure on $\theta > 0$, the derivative of the second term is strictly negative

$$\frac{d}{dT} \int_0^{\theta_u} \theta \beta_{\delta_T}(\theta) dQ(\theta) = \int_0^{\theta_u} \theta \frac{d\beta_{\delta_T}(\theta)}{dT} dQ(\theta) < 0,$$

hence the threshold T_Q maximizing $V_Q(\delta_T)$ is smaller than T^*. $\qquad\square$

The next proposition shows that the hypothesis test rule with level $c/b(0)$ could be interpreted as an approximation of Bayesian decision rules that would be taken by regulators sufficiently pessimistic about the pool of potential proponents seeking approval for their innovations. If the regulator places higher and higher prior probability on $\theta < 0$, the threshold of the optimal decision rule converges to the hypothesis test threshold T^*.

Proposition 4. *Let assumptions P1 and S1 hold. Let Q be a probability measure with density $q(\theta)$ and a finite mean. Assume that $Q((0, \theta_u)) > 0$ and $q(\theta) > 0$ on the interval $[-\varepsilon, 0]$ for some $\varepsilon > 0$. Let $\{a_n\}$ be an increasing sequence of numbers $a_n \in (0, 1)$, $a_n \to 1$, denoting the prior weight on $\theta < 0$. Define a sequence of probability measures Q_n with densities*

$$q_n(\theta) \equiv \left(\frac{a_n \mathrm{I}[\theta < 0]}{Q((\theta_l, 0))} + \frac{(1 - a_n) \mathrm{I}[\theta \geq 0]}{Q([0, \theta_u))} \right) q(\theta)$$

and a sequence of decision rule thresholds $T_n \equiv \arg\max_T V_{Q_n}(\delta_T)$ that are optimal for regulators with priors Q_n. Then $T_n \to T^*$.

Proof. It follows from Proposition 3 that $T_n < T^*$ for all n. Fix any $\bar{T} < T^*$. For any δ_T with $T \leq \bar{T}$,

$$
\begin{aligned}
V_{Q_n}(\delta_T) &= \int_{\bar{\theta}(T)}^0 \theta \beta_{\delta_T}(\theta) q_n(\theta) d\theta + \int_0^{\theta_u} \theta \beta_{\delta_T}(\theta) q_n(\theta) d\theta \\
&= \frac{a_n}{Q((\theta_l, 0))} \int_{\bar{\theta}(T)}^0 \theta \beta_{\delta_T}(\theta) q(\theta) d\theta + \frac{1 - a_n}{Q([0, \theta_u))} \int_0^{\theta_u} \theta \beta_{\delta_T}(\theta) q(\theta) d\theta \\
&\leq \frac{a_n}{Q((\theta_l, 0))} \int_{\bar{\theta}(\bar{T})}^0 \theta \beta_{\delta_T}(\theta) q(\theta) d\theta + \frac{1 - a_n}{Q([0, \theta_u))} \int_0^{\theta_u} \theta \beta_{\delta_T}(\theta) q(\theta) d\theta \\
&\leq \frac{a_n}{Q((\theta_l, 0))} \int_{\bar{\theta}(\bar{T})}^0 \theta \beta_{\delta_{\bar{T}}}(\theta) q(\theta) d\theta + \frac{1 - a_n}{Q([0, \theta_u))} \int_0^{\theta_u} \theta q(\theta) d\theta.
\end{aligned}
$$

(20)

The first inequality holds because the first integrand is negative over $\theta < 0$ and the participation threshold $\bar{\theta}(T)$ is increasing in T. The second inequality applies because $\beta_{\delta_{\bar{T}}}(\theta) \leq \beta_{\delta_T}(\theta) \leq 1$.

The first integral in (20) is strictly negative because $q(\theta) > 0$ on $[-\varepsilon, 0]$. The second integral is positive and finite because Q has a finite mean. Their weighted sum is negative for sufficiently large n because $a_n \to 1$. Hence, for sufficiently large n, $V_{Q_n}(\delta_T) < 0 \leq V_{Q_n}(\delta_{T^*})$ for every $T \leq \bar{T}$, implying that $T_n > \bar{T}$. Since this is true for any $\bar{T} < T^*$, it follows that $T_n \to T^*$. $\qquad\square$

4 Proponents imperfectly informed about the value of their innovations

Proponents may have some information about the value of their innovations, but not enough to know θ with certainty, as the paper assumed thus far. In this section, I show that hypothesis tests with level $c/b(0)$ are still maximin and admissible for the regulator if the proponent's payoff function $b(\theta)$ is concave. The proponent's type in this case will be denoted by π - the probability distribution that the proponent places on the quality θ of his innovation prior to conducting a trial. The proponent's beliefs about θ may be based on any evidence collected by the proponent before the trial. It is important, though, that this evidence should be uninformative about the distribution of the trial outcome $P(X|\theta)$, except through the proponent's beliefs about θ. The proponent's type π could be any probability distribution on $\Theta \subset \mathbb{R}$ with a finite mean. Denote the set of all such distributions by Δ and the subset of degenerate distributions, representing certainty about θ, by Δ_0.

The regulator needs to take into account that proponents may not want to seek approval for their innovations upon observing the outcome X of the trial. This happens if the proponent's posterior $\pi(\theta|X)$ is sufficiently negative. Denote the proponent's optimal decision on whether to seek approval upon observing the trial outcome by the function[3]

---

[3] I assume for simplicity that the proponent chooses to seek approval if he is indifferent. Modifying this assumption does not affect the results.

$\eta_\pi(X) \equiv \mathrm{I}\left[\int_\Theta b(\theta)d\pi(\theta|X) \geq 0\right]$. When the regulator commits to a decision rule δ, the proponent's expected probability of approval prior to conducting the trial equals

$$(21) \qquad \beta_{\delta,\pi}(\theta) \equiv \int_{\mathcal{X}} \delta(X)\eta_\pi(X)dF(X;\theta).$$

It is profitable for the proponent to conduct a trial if

$$(22) \qquad \int_\Theta b(\theta)\beta_{\delta,\pi}(\theta)d\pi(\theta) - c \geq 0.$$

Since $\beta_{\delta,\pi}(\theta)$ is nonlinear in θ, the proponent's payoff depends in a nontrivial fashion on his prior π and a closed-form solution to the proponent's decision problem seems unlikely.

Threshold hypothesis test rule (13) with test level $c/b(0)$ remains maximin and admissible for the regulator in this setting with the following additional assumptions on the proponent's payoff function and on the distribution $F(X;\theta)$.

Assumption P2 (payoffs).
The proponent's payoff $b(\theta)$ is a continuous, non-decreasing, weakly concave function of θ. It may take negative values, but $b(0) > 0$.

Assumption S2 (data).
The ratio $\left(-\frac{dF(T;\theta)}{d\theta}\right) \Big/ (1 - F(T;\theta))$ is non-increasing in θ for all $T \in \mathbb{R}$.

Examples of distributions that satisfy Assumptions S1 and S2 include the family of normal distributions $X \sim \mathcal{N}(\theta, \sigma^2)$ with known variance σ^2 and the family of exponential distributions with means $\mu_0 + \theta$, where μ_0 is known.

The following proposition shows that conducting the trial is only optimal for proponents with beliefs π for which the regulator's expected payoff, evaluated using the proponent's beliefs π, is non-negative.

Proposition 5. *Suppose that Assumptions P1, P2, S1 and S2 hold, and that the regulator commits to the hypothesis test rule δ^* in (13). If the proponent's beliefs π imply a negative*

expected payoff to the regulator following the trial, that is, if

$$(23) \qquad \int_\Theta \theta \beta_{\delta^*,\pi}(\theta) d\pi(\theta) < 0,$$

then it is unprofitable for a proponent of type π to conduct the trial,

$$(24) \qquad \int_\Theta b(\theta) \beta_{\delta^*,\pi}(\theta) d\pi(\theta) - c < 0.$$

Proof. See Appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

It follows from Proposition 5 that the regulator faces a non-negative expected payoff from any potential proponent type π, hence the test rule δ^* is maximin for the regulator. The following proposition also establishes that the hypothesis test rule δ^* is admissible.

Proposition 6. *Under Assumptions P1, P2, S1 and S2, the hypothesis test rule δ^* in (13) is maximin and admissible for the regulator with respect to $\pi \in \Delta$.*

Proof. For δ^* to be inadmissible, there must be a different monotone decision rule δ that yields strictly higher payoff to the regulator for some type π and does not yield lower payoff for any type π. Any other monotone decision rule $\delta \neq \delta^*$ either has lower power $\beta_\delta(\bar\theta) < \beta_{\delta^*}(\bar\theta)$ for all $\bar\theta > 0$ or higher size $\beta_\delta(0) > \beta_{\delta^*}(0) = c/b(0)$ by the Neyman-Pearson lemma. In the first case, it yields the regulator a lower payoff than δ^* for any $\pi \in \Delta_0$ that places probability one on some $\bar\theta > 0$. In the second case, the regulator's payoff is negative for proponent types that place probability one on some $\bar\theta < 0$ sufficiently close to zero.

It follows from Proposition 5 that the regulator's payoff from δ^* is non-negative for any $\pi \in \Delta$. No decision rule could yield a higher minimum, since all decision rules yield payoff of zero for π_0 that places probability $\pi_0(\theta = 0) = 1$, hence δ^* is maximin. $\qquad$ □

One-sided hypothesis test rules with level $c/b(0)$ are attractive if the regulator cannot place more precise restrictions on the distribution of potential proponent types, at least if the proponent's benefit from approval $b(\theta)$ is concave in θ.

The results of Propositions 5 and 6 do not hold if the proponent's benefit $b(\theta)$ is not concave. The following example illustrates that proponents with non-concave $b(\theta)$ may find it profitable to test innovations that yield a negative payoff to the regulator. It may be possible, however, to deter all undesirable proponents with a test rule stricter than δ^*.

Example 7. *Suppose that $b(-1) = 0.1$, $b(0) = 1$, and $b(1) = 10$, thus $b(\theta)$ is not concave. Let X be normally distributed with $F(x; \theta) = \Phi(x - \theta)$ and let $c = 1/2$. The hypothesis test rule (13) must have size $1/2$ and equals $\delta^*(X) = \mathrm{I}[X \geq 0]$. The probability of approval as a function of θ equals $\beta_{\delta^*}(\theta) = 1 - F(0; \theta) = 1 - \Phi(-\theta) = \Phi(\theta)$.*

If the proponent's prior beliefs place probabilities $\pi(\theta = -1) = 0.9$ and $\pi(\theta = 1) = 0.1$, then the expected payoff to the regulator is negative if the trial is conducted,

$$\sum_{\theta \in \{-1, 1\}} \theta \beta_{\delta^*}(\theta) \pi(\theta) = -1 \cdot \Phi(-1) \cdot 0.9 + 1 \cdot \Phi(1) \cdot 0.1 \approx -0.059.$$

However, conducting the trial is profitable for the proponent, whose expected payoff is

$$\sum_{\theta \in \{-1, 1\}} b(\theta) \beta_{\delta^*}(\theta) \pi(\theta) - c = 0.1 \cdot \Phi(-1) \cdot 0.9 + 10 \cdot \Phi(1) \cdot 0.1 - 0.5 \approx 0.356.$$

5　Endogenous choice of testing cost and precision

So far, the cost c of conducting the trial and the data distribution $F(X; \theta)$ (e.g., the sample size) were treated as exogenous for both parties and were not determined within the game. If proponents could freely choose c and F, hypothesis test rules with the proposed level $c/b(0)$ remain an effective deterrent for proponents with $\theta < 0$, as long as the test level is based on their chosen trial cost and data distribution. However, it may be optimal for the regulator to make the test stricter for some choices of (c, F) in order to induce the proponent to choose another trial design (c', F') that would yield a higher expected payoff to the regulator. The regulator's concerns when all proponents are perfectly informed about θ are different from the case when proponents may be imperfectly informed.

If all proponents are ex ante certain about the value of θ and could choose how much to

spend on conducting the trial, the regulator would prefer to push them towards choosing $c = b(0)$. This effectively replaces statistical signaling with pure monetary signaling of the proponent's type.

Example 8. *Suppose that $\pi \in \Delta_0$ and $b(\theta)$ is strictly increasing in θ. Let $\delta^*(X; c, F)$ be a test rule that satisfies condition (7) for each choice of (c, F). If the proponent's choice set of (c, F) includes at least one choice with $c = b(0)$, then the decision rule*

$$\delta(X; c, F) = \mathrm{I}[c = b(0)]$$

dominates δ^*. All proponents with $\theta > 0$ still find it profitable to conduct trials and the approval rate for them equals 1, which yields the highest possible payoff for the regulator. On the other hand, conducting the trial remains unprofitable for all proponents with $\theta < 0$ because $b(\theta) < c = b(0)$.

Even if $b(\theta)$ is not strictly increasing, but $F(X; \theta)$ has the MLR property, the regulator could achieve almost the same effect by using decision rules that depend only slightly on the statistical signal. Suppose that (c^*, F^*) is a trial design with $c^* < b(0)$ and $\delta^*(X; c^*, F^*)$ satisfies (13). If trial designs (c', F^*) with the same statistical signal and higher costs $c' \in (c^*, b(0))$ are also available, then consider decision rules

$$
\begin{aligned}
&\delta_{c'}(X; c, F) = \frac{c' - c^*}{b(0) - c^*} + \frac{b(0) - c'}{b(0) - c^*}\delta^*(X; c^*, F^*) \quad &&\text{if } (c, F) = (c', F^*), \\
&\delta_{c'}(X; c, F) = 0 &&\text{if } (c, F) \neq (c', F^*).
\end{aligned}
$$
(25)

For all $\theta < 0$, the approval rate under $\delta_{c'}$ is strictly lower than

$$\frac{c' - c^*}{b(0) - c^*} + \frac{b(0) - c'}{b(0) - c^*}\beta_{\delta^*(\cdot; c^*, F^*)}(0) < \frac{c' - c^*}{b(0) - c^*} + \frac{b(0) - c'}{b(0) - c^*} \cdot \frac{c^*}{b(0)} = \frac{c'}{b(0)},$$

which makes the trial unprofitable for proponents with $\theta < 0$. But for all $\theta > 0$ the approval rate under $\delta_{c'}$ converges to 1 as $c' \to b(0)$.

If proponents are not perfectly informed about θ and could choose trial design (c, F), the result of Proposition 5 still holds, hence $\delta^*(X; c, F)$ is a maximin decision rule if $b(\theta)$

22

is concave. It is less clear whether it is admissible, i.e., whether there may be a decision rule that dominates the hypothesis test rule $\delta^*(X; c, F)$ when the proponent could choose the trial design.

The regulator's incentives to demand costlier trials (which yield higher approval rates for $\theta > 0$) are offset by the risk of making trials too costly for imperfectly informed proponents that the regulator would prefer to conduct trials. The following proposition illustrates this intuition in a setting where the data is normally distributed and proponents could choose between two trial designs, one of which yields more precise evidence at a higher cost. In this setting, there is no decision rule that dominates letting the proponent choose the trial design and then applying test rule (15) to the data.

Proposition 9. *Suppose that $X \sim \mathcal{N}(\theta, \sigma^2)$, i.e. $F_\sigma(x; \theta) \equiv \Phi((x - \theta)/\sigma)$, $\pi \in \Delta$ and $b(\theta) = b$ for all θ. If the proponent could choose between two trial designs (c_1, σ_1) and (c_2, σ_2) with $b/2 > c_1 > c_2$ and $\sigma_1 < \sigma_2$, then the decision rule*

$$(26) \qquad \delta^* (X; c_i, \sigma_i) = \mathrm{I} \left[X > \sigma_i \Phi^{-1}(c_i/b) \right]$$

is admissible.

Proof. See Appendix. □

6 Statistical Testing for Drug Approval

The approval of new drugs after Phase III clinical trials has many similarities to the environment of this paper. This section gauges what test levels could result from applying this paper's analysis to drug approval. Phase III trials are large-sample double-blind trials in which the proposed drug is compared to a placebo or an existing treatment. They are the last stage of trials before the regulator reviews a drug for approval and statistical hypothesis testing of the drug's effectiveness is an important part of the approval decision. The costs of the trials are borne by the companies and now constitute 36% of their R&D expenses (PhRMA, 2013). Deterrence of frivolous proposals should be a serious concern

because the pool of new compounds that pharmaceutical companies could potentially propose is very large, with 5,000–10,000 compounds entering the R&D pipeline per each approved drug (PhRMA, 2013).

If approval is based on a one-sided hypothesis test against the status quo treatment, then this paper suggests that its significance level should be $c/b(0)$. The trial cost c, in this case, is the present value of sunk costs of the Phase III clinical trial. The benefit $b(0)$ is the present value of the expected profit from the regulatory approval of the drug if the drug's true treatment effect is zero. Both parameters vary a lot from drug to drug (Grabowski et al., 2002; DiMasi et al., 2003), but the relevant data for individual drugs are not readily available.

I derive the deterrent test level for a "representative drug" with average Phase III clinical trial costs $\bar{c}$ and expected profits $\overline{b(0)}$ equal to the average profits of approved drugs. This overestimates $\overline{b(0)}$ if approved drugs have positive effects and hence profits higher than $b(0)$. Given the simplicity of the formula, it is very easy to see how this test level varies depending on the values of clinical trial costs and expected profits.

The estimate of $\bar{c}$ comes from R&D cost estimates of DiMasi et al. (2003), who collected detailed confidential cost data from pharmaceutical firms for a sample of drugs first tested in 1983–1994. They report \$119.2 million (in 2000 dollars) as the average cost of a Phase III clinical trial, including trials that did not lead to approval. The trial costs are spread over an average of 30.5 months and the estimate discounts the costs at 11% rate to the time of approval (estimated to be 18.2 months after the end of Phase III trials). The rate of 11% is the real cost of capital estimated by DiMasi et al. (2003) for the pharmaceutical industry during the study's time period.

I use average pre-approval R&D costs divided by the number of approved drugs to estimate the average profits of approved drugs. These should be equal if firms face zero expected profits. DiMasi et al. (2003) estimate the average R&D costs to be \$802 million (in 2000 dollars). This estimate discounts the costs of preclinical research and all stages of clinical trials to the time of drug approval. It includes the costs of developing drug candidates that were abandoned or were not approved. Grabowski et al. (2002) collected

sales data for drugs analyzed by DiMasi et al. (2003) that were introduced in 1990–1994. They estimate the difference between the average present value of profits from future sales and the average present value of R&D costs for these drugs to be less than 10%.

Combining these estimates yields the deterrent level for testing a "representative drug" equal to

$$(27) \qquad \alpha = \frac{\$119.2 \text{ million}}{\$802 \text{ million}} = 14.9\%.$$

Both the distribution of sales and the distribution of clinical trial costs are very dispersed, so this is a point from a large range of deterrent test levels For example, Grabowski et al. (2002) report that sales of drugs in the top decile of the distribution are almost 5.5 times higher than the mean. The deterrent test level for an average top-decile drug with average clinical trial costs would then be 2.7%.

These estimates pertain to a very simplified scenario in which drug approval depends on results of a single hypothesis test of the drug's overall treatment effect. The tested treatment effect ought to capture all the relevant dimensions of patient outcomes, including side effects, inconveniences, and cost differences. The actual drug approval process is more complex. If it were a matter of such a single test, though, the estimates above suggest that conventional test levels of 5% and 1% would be too strict in some cases. Higher test levels could still offer sufficient deterrence against all frivolous applications and increase the chance of approval for effective drugs.

7   Conclusion

The paper presented a new theory for the use of hypothesis test rules in regulatory approval and for choosing the levels of these tests. The probability of Type I errors for innovations with negative effects has to be contained to deter potential proponents from flooding the regulator with bad proposals. The proposed test level is determined by the ratio of the proponent's testing costs over his expected benefits from the proposal's approval by the regulator. In this setting, hypothesis test rules turn out to be admissible

and maximin decision rules for the regulator with respect to the proponent's ex ante beliefs about the quality of the innovation. They can also be seen as a limit of test rules adopted by Bayesian regulators who are sufficiently pessimistic about the proportion of frivolous proposals. This paper illustrates that some statistical inference procedures are, in effect, policy tools and should take into account their incentive effects.

Appendix

Proof of Proposition 5. The proponent's decision rule for requesting approval following a trial always has the threshold form $\eta_\pi(X) = \mathrm{I}\left[X \geq T_P\right]$, $T_P \geq -\infty$, due to the MLR property of $F(X; \theta)$. Then the joint decision rule (probability that both the regulator and the proponent want to approve the innovation upon observing X) is also a threshold function of X,

$$\tilde{\delta}(X) \equiv \delta^*(X)\eta_\pi(X) = \mathrm{I}\left[X \geq \max(T^*, T_P)\right],$$

equivalent to the outcome of a one-sided hypothesis test rule with threshold $\tilde{T} = \max(T^*, T_P)$, which is a test with size $\beta_{\tilde{\delta}}(0) = \tilde{c}/b(0)$ for some value $\tilde{c} \leq c$, and acceptance probability $\beta_{\tilde{\delta}}(\theta) = \beta_{\delta^*, \pi}(\theta)$.

The weakly concave function $b(\theta)$ can be bounded above by a linear function

$$\tilde{b}(\theta) \equiv b(0) + \gamma\theta \geq b(\theta).$$

passing through $b(0)$. Since $b(\theta)$ is non-decreasing, $\gamma \geq 0$.

The proponent's expected payoff from conducting a trial then equals

$$
\begin{aligned}
\text{(A1)} \qquad \int_\Theta b(\theta)\beta_{\tilde{\delta}}(\theta)d\pi(\theta) - c &= (\tilde{c} - c) + \int_\Theta [b(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}]d\pi(\theta) \leq \\
&\leq (\tilde{c} - c) + \int_\Theta [\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}]d\pi(\theta).
\end{aligned}
$$

The integrand $[\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}]$ equals zero at $\theta = 0$, is positive for $\theta > 0$ and negative for $\theta < 0$. The same is true for $\theta\beta_{\tilde{\delta}}(\theta)$, hence the ratio

$$r(\theta) \equiv \frac{\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}}{\theta\beta_{\tilde{\delta}}(\theta)}$$

is positive for all $\theta \neq 0$. While it is undefined at $\theta = 0$, it has a well-defined limit from both sides as $\theta \to 0$, which will be denoted by $r(0)$. Let $\beta'_{\tilde{\delta}}(\theta) \equiv \frac{d\beta_{\tilde{\delta}}(\theta)}{d\theta}$. By L'Hopital's

rule,

$$
\begin{aligned}
r(0) &\equiv \lim_{\theta \to 0} \frac{\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}}{\theta \beta_{\tilde{\delta}}(\theta)} = \frac{\lim_{\theta \to 0}[\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}]'}{\lim_{\theta \to 0}[\theta \beta_{\tilde{\delta}}(\theta)]'} = \\
&= \frac{\lim_{\theta \to 0}\left[\gamma \beta_{\tilde{\delta}}(\theta) + [b(0) + \gamma\theta]\,\beta'_{\tilde{\delta}}(\theta)\right]}{\lim_{\theta \to 0}\left[\beta_{\tilde{\delta}}(\theta) + \theta\beta'_{\tilde{\delta}}(\theta)\right]} = \\
&= \frac{\gamma \beta_{\tilde{\delta}}(0) + b(0)\,\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)} = \gamma + b(0)\frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)}.
\end{aligned}
$$

Since $\gamma > 0$, $b(0) > 0$, $\beta_{\tilde{\delta}}(0) > 0$ and $\beta'_{\tilde{\delta}}(0) > 0$, it follows that $r(0) > 0$.

I will now show that $r(\bar{\theta}) > r(0)$ for all $\bar{\theta} < 0$ and that $r(\bar{\theta}) < r(0)$ for all $\bar{\theta} > 0$. Since

$$
\frac{\beta'_{\tilde{\delta}}(\theta)}{\beta_{\tilde{\delta}}(\theta)} = \frac{-\frac{dF(\tilde{T};\theta)}{d\theta}}{1 - F(\tilde{T};\theta)},
$$

it follows from Assumption S2 that $\frac{\beta'_{\tilde{\delta}}(\theta)}{\beta_{\tilde{\delta}}(\theta)}$ is non-increasing in θ.

Take any $\bar{\theta} < 0$. Then for all $\theta \in (\bar{\theta}, 0)$, $\beta_{\tilde{\delta}}(\theta) > \beta_{\tilde{\delta}}(\bar{\theta})$, and hence

$$
\frac{\beta'_{\tilde{\delta}}(\theta)}{\beta_{\tilde{\delta}}(\bar{\theta})} > \frac{\beta'_{\tilde{\delta}}(\theta)}{\beta_{\tilde{\delta}}(\theta)} \geq \frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)}.
$$

Then the difference $\beta_{\tilde{\delta}}(0) - \beta_{\tilde{\delta}}(\bar{\theta})$ could be bounded below by

$$
\beta_{\tilde{\delta}}(0) - \beta_{\tilde{\delta}}(\bar{\theta}) = \int_{\bar{\theta}}^{0} \beta'_{\tilde{\delta}}(\theta)d\theta = \beta_{\tilde{\delta}}(\bar{\theta})\int_{\bar{\theta}}^{0}\frac{\beta'_{\tilde{\delta}}(\theta)}{\beta_{\tilde{\delta}}(\bar{\theta})}d\theta > \beta_{\tilde{\delta}}(\bar{\theta})\int_{\bar{\theta}}^{0}\frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)}d\theta = -\bar{\theta}\beta_{\tilde{\delta}}(\bar{\theta})\frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)},
$$

therefore $\frac{\beta_{\tilde{\delta}}(\bar{\theta}) - \beta_{\tilde{\delta}}(0)}{\bar{\theta}\beta_{\tilde{\delta}}(\bar{\theta})} > \frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)}$. Substituting $\tilde{c} = b(0)\beta_{\tilde{\delta}}(0)$ and $\tilde{b}(\bar{\theta}) = b(0) + \gamma\bar{\theta}$,

$$
\begin{aligned}
r(\bar{\theta}) - r(0) &= \frac{\tilde{b}(\bar{\theta})\beta_{\tilde{\delta}}(\bar{\theta}) - \tilde{c}}{\bar{\theta}\beta_{\tilde{\delta}}(\bar{\theta})} - \gamma - b(0)\frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)} = \\
&= \frac{\left(b(0) + \gamma\bar{\theta}\right)\beta_{\tilde{\delta}}(\bar{\theta}) - b(0)\beta_{\tilde{\delta}}(0)}{\bar{\theta}\beta_{\tilde{\delta}}(\bar{\theta})} - \gamma - b(0)\frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)} = \\
&= b(0)\left[\frac{\beta_{\tilde{\delta}}(\bar{\theta}) - \beta_{\tilde{\delta}}(0)}{\bar{\theta}\beta_{\tilde{\delta}}(\bar{\theta})} - \frac{\beta'_{\tilde{\delta}}(0)}{\beta_{\tilde{\delta}}(0)}\right] > 0.
\end{aligned}
$$

The proof that $r(\bar{\theta}) - r(0) < 0$ for all $\bar{\theta} > 0$ is analogous.

Since $r(\theta) < r(0)$ and $\theta\beta_{\tilde{\delta}}(\theta) > 0$ for all $\theta > 0$,

$$\int_0^{\theta_u} \left[\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}\right] d\pi(\theta) = \int_0^{\theta_u} r(\theta)\theta\beta_{\tilde{\delta}}(\theta)d\pi(\theta) \leq r(0) \int_0^{\theta_u} \theta\beta_{\tilde{\delta}}(\theta)d\pi(\theta).$$

Similarly, since $r(\theta) > r(0)$ and $\theta\beta_{\tilde{\delta}}(\theta) < 0$ for all $\theta < 0$,

$$\int_{\theta_l}^0 \left[\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}\right] d\pi(\theta) = \int_{\theta_l}^0 r(\theta)\theta\beta_{\tilde{\delta}}(\theta)d\pi(\theta) \leq r(0) \int_{\theta_l}^0 \theta\beta_{\tilde{\delta}}(\theta)d\pi(\theta).$$

Adding these two inequalities yields

(A2) $$\int_\Theta \left[\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}\right] d\pi(\theta) \leq r(0) \int_\Theta \theta\beta_{\tilde{\delta}}(\theta)d\pi(\theta).$$

If $\int_\Theta \theta\beta_{\tilde{\delta}}(\theta)d\pi(\theta) < 0$, then it follows from (A1), (A2), and $r(0) > 0$ that also

$$\int_\Theta [b(\theta)\beta_{\tilde{\delta}}(\theta) - c]d\pi(\theta) \leq \int_\Theta [\tilde{b}(\theta)\beta_{\tilde{\delta}}(\theta) - \tilde{c}]d\pi(\theta) < 0.$$

$\square$

Proof of Proposition 9. For δ^* to be inadmissible, there has to exist another decision rule δ' that satisfies condition (7), gives the regulator strictly higher payoffs for at least one proponent type π and does not give lower payoffs to the regulator for any type π.

First, I will show that such a decision rule δ' must be equal to δ^* for (c_1, σ_1). This is established by considering proponents with beliefs that place mass one on $\bar{\theta} > 0$ for $\bar{\theta}$ sufficiently close to zero. At $\theta = 0$, the approval probability under δ^* for the more precise trial has a higher derivative $\beta'_{\delta^*(\cdot;c_1,\sigma_1)}(\theta)\big|_{\theta=0} > \beta'_{\delta^*(\cdot;c_2,\sigma_2)}(\theta)\big|_{\theta=0}$. This implies that for some range of values of $\bar{\theta} > 0$, proponents who believe that $\theta = \bar{\theta}$ find it profitable to choose (c_1, σ_1) under δ^*, since at $\theta = 0$ their payoffs equal 0. By the Neyman-Pearson lemma, any decision rule based on X with variance σ_1^2 different from $\delta^*(\cdot; c_1, \sigma_1)$ either violates (7) or offers lower probability of acceptance at $\bar{\theta}$, hence yields a lower payoff to the regulator. Because of the higher variance of X under the choice of (c_2, σ_2), the probability of acceptance at $\bar{\theta}$ is also lower under any decision rule $\delta'(\cdot; c_2, \sigma_2)$ if the

proponent is forced to choose (c_2, σ_2) instead, yielding a lower payoff to the regulator. Hence, it must be that $\delta'(\cdot; c_1, \sigma_1) = \delta^*(\cdot; c_1, \sigma_1)$.

With $\delta'(\cdot; c_1, \sigma_1) = \delta^*(\cdot; c_1, \sigma_1)$ fixed, I show that any deviations from $\delta^*(\cdot; c_2, \sigma_2)$ will reduce the regulator's payoff for some proponent type π. It will be useful to consider proponents who have two-point priors on θ that place probability q on some $\theta_1 < 0$ and probability $1 - q$ on some $\theta_2 > 0$, which is sufficiently high so that $b\beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_2) - c_2 > b\beta_{\delta^*(\cdot; c_1, \sigma_1)}(\theta_2) - c_1$. Such a value of θ_2 exists because as $\theta \to \infty$, $b\beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_2) - c_2 \to b - c_2$, whereas $b\beta_{\delta^*(\cdot; c_1, \sigma_1)}(\theta_2) - c_1 \to b - c_1 < b - c_2$.

Given the regulator's decision rule δ, the expected payoff from conducting a trial (c_i, σ_i) for such proponent types equals

(A3) $$q(b\beta_{\delta(\cdot; c_i, \sigma_i)}(\theta_1) - c_i) + (1 - q)(b\beta_{\delta(\cdot; c_i, \sigma_i)}(\theta_2) - c_i).$$

Under decision rule $\delta^*(\cdot; c_i, \sigma_i)$, $(b\beta_{\delta^*(\cdot; c_i, \sigma_i)}(\theta_1) - c_i) < 0$ and $(b\beta_{\delta^*(\cdot; c_i, \sigma_i)}(\theta_2) - c_i) > 0$. Conducting the trial (c_i, σ_i) is only profitable for proponents with $q \in [0, q_i]$, where

(A4) $$q_i = \frac{(b\beta_{\delta^*(\cdot; c_i, \sigma_i)}(\theta_2) - c_i)}{(b\beta_{\delta^*(\cdot; c_i, \sigma_i)}(\theta_2) - c_i) - (b\beta_{\delta^*(\cdot; c_i, \sigma_i)}(\theta_1) - c_i)}.$$

By the choice of θ_2, $(b\beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_2) - c_2) > (b\beta_{\delta^*(\cdot; c_1, \sigma_1)}(\theta_2) - c_1) > 0$. I will also establish below that $-(b\beta_{\delta^*(\cdot; c_1, \sigma_1)}(\theta_1) - c_1) > -(b\beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_1) - c_2) > 0$. It follows that $q_2 > q_1$. For proponents with $q \in (q_1, q_2)$ it is unprofitable to conduct a trial with design (c_1, σ_1) when faced with the test rule δ^*, but it is profitable to conduct a trial with design (c_2, σ_2).

Suppose that the regulator replaced $\delta^*(\cdot; c_2, \sigma_2)$ with a different decision rule $\delta'(\cdot; c_2, \sigma_2)$. First, it follows from the Neyman-Pearson lemma that either $\beta_{\delta'(\cdot; c_2, \sigma_2)}(0) > \beta_{\delta^*(\cdot; c_2, \sigma_2)}(0)$ or $\beta_{\delta'(\cdot; c_2, \sigma_2)}(\theta_2) < \beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_2)$. If $\beta_{\delta'(\cdot; c_2, \sigma_2)}(0) > \beta_{\delta^*(\cdot; c_2, \sigma_2)}(0)$ then $\delta'(\cdot; c_2, \sigma_2)$ makes the trial profitable for proponents who are certain about θ for some values $\theta < 0$ sufficiently close to 0, therefore δ' cannot dominate δ^*. Thus, it must be that $\beta_{\delta'(\cdot; c_2, \sigma_2)}(\theta_2) < \beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_2)$.

The acceptance probability of δ' at θ_1 could either be $\beta_{\delta'(\cdot; c_2, \sigma_2)}(\theta_1) \leq \beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_1)$ or $\beta_{\delta'(\cdot; c_2, \sigma_2)}(\theta_1) > \beta_{\delta^*(\cdot; c_2, \sigma_2)}(\theta_1)$. In the first case, the proponent's payoff from trial (c_2, σ_2) is

strictly lower than under δ^*. Proponents with some beliefs $q \in (q_1, q_2)$ will no longer find it profitable to conduct the trial (c_2, σ_2) and thus will not conduct any trial. According to Proposition 5, a regulator with decision rule δ^* gets a positive expected payoff from their participation, thus δ' yields a strictly lower payoff for this type of proponent and does not dominate δ^*. In the second case, $\beta_{\delta'(\cdot;c_2,\sigma_2)}(\theta_1) > \beta_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_1)$, there are two possibilities. A proponent with beliefs $q \in (q_1, q_2)$ may stop participating, thus reducing the regulator's payoff to zero. If the proponent still finds it profitable to conduct the trial facing decision rule $\delta'(\cdot; c_2, \sigma_2)$, the regulator will get a lower payoff from this proponent type because δ' has a lower acceptance probability for good innovations with $\theta = \theta_2$ and a higher probability of accepting bad innovations with $\theta = \theta_1$. Therefore, there is no alternative decision rule δ' that dominates $\delta^*(\cdot; c_i, \sigma_i)$.

It remains to be shown that $-(b\beta_{\delta^*(\cdot;c_1,\sigma_1)}(\theta_1) - c_1) > -(b\beta_{\delta^*(\cdot;c_2,\sigma_2)}(\theta_1) - c_2) > 0$. Let $T_i^* = \sigma_i \Phi^{-1}\left(1 - \frac{c_i}{b}\right)$ be the threshold of the decision rule $\delta^*(\cdot; c_i, \sigma_i)$. By the design of $\delta^*(\cdot; c_i, \sigma_i)$, $c_i = b\beta_{\delta^*(\cdot;c_i,\sigma_i)}(0) = b(1 - \Phi(T_i^*/\sigma_i))$. Then

$$b\beta_{\delta^*(\cdot;c_i,\sigma_i)}(\theta) - c_i = b\left(1 - \Phi\left(\frac{T_i^* - \theta}{\sigma_i}\right)\right) - b\left(1 - \Phi\left(\frac{T_i^*}{\sigma_i}\right)\right) = b\int_{(T_i^*-\theta)/\sigma_i}^{T_i^*/\sigma_i} \phi(t)dt.$$

Since $c_2 < c_1 < b/2$, $\frac{T_2^*}{\sigma_2} = \Phi^{-1}\left(1 - \frac{c_2}{b}\right) > \frac{T_1^*}{\sigma_1} = \Phi^{-1}\left(1 - \frac{c_1}{b}\right) > \Phi^{-1}\left(\frac{1}{2}\right) = 0$ and $\left|\frac{\theta}{\sigma_1}\right| > \left|\frac{\theta}{\sigma_2}\right|$. It follows that for any $\theta < 0$,

$$\int_{(T_1^*-\theta)/\sigma_1}^{T_1^*/\sigma_1} \phi(t)dt < \int_{(T_2^*-\theta)/\sigma_2}^{T_2^*/\sigma_2} \phi(t)dt < 0,$$

therefore $b\beta_{\delta^*(\cdot;c_1,\sigma_1)}(\theta) - c_1 < b\beta_{\delta^*(\cdot;c_2,\sigma_2)}(\theta) - c_2 < 0$.

$\square$

References

CHASSANG, S., G. PADRO I MIQUEL, AND E. SNOWBERG (2012): "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 102, 1279–1309.

COHEN, J. (1994): "The earth is round (p < .05)," *American Psychologist*, 49, 997–1003.

DI TILLIO, A., M. OTTAVIANI, AND P. N. SORENSEN (2016): "Persuasion Bias in Science: Can Economics Help?" *Working Paper*.

DIMASI, J. A., R. W. HANSEN, AND H. G. GRABOWSKI (2003): "The price of innovation: new estimates of drug development costs," *Journal of Health Economics*, 22, 151–185.

GRABOWSKI, H., J. VERNON, AND J. DIMASI (2002): "Returns on Research and Development for 1990s New Drug Introductions," *PharmacoEconomics*, 20, 11–29.

HENRY, E. AND M. OTTAVIANI (2015): "The Organization of Persuasion: Wald Deconstructed," *Working Paper*.

HIRANO, K. AND J. R. PORTER (2009): "Asymptotics for Statistical Treatment Rules," *Econometrica*, 77, 1683–1701.

INTERNATIONAL COMMITTEE ON HARMONISATION (1999): "ICH E9 Expert Working Group. Statistical Principles for Clinical Trials: ICH Harmonized Tripartite Guideline," *Statistics in Medicine*, 18, 1905–1942.

JOHNSON, D. H. (1999): "The Insignificance of Statistical Significance Testing," *The Journal of Wildlife Management*, 63, 763–772.

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

KARLIN, S. AND H. RUBIN (1956): "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *The Annals of Mathematical Statistics*, 27, 272–299.

LEHMANN, E. AND J. ROMANO (2005): *Testing Statistical Hypotheses*, Springer Texts in Statistics, Springer, 3 ed.

MANSKI, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

——— (2005): *Social Choice with Partial Knowledge of Treatment Response*, Princeton University Press.

——— (2015): "Randomizing Regulatory Approval for Adaptive Diversification and Deterrence," *The Journal of Legal Studies*, 44, S367–S385.

MANSKI, C. F. AND A. TETENOV (2007): "Admissible treatment rules for a risk-averse planner with experimental data on an innovation," *Journal of Statistical Planning and Inference*, 137, 1998–2010.

PHARMACEUTICAL RESEARCH AND MANUFACTURERS OF AMERICA (2013): *2013 Biopharmaceutical Research Industry Profile*.

SAVAGE, L. J. (1954): *The Foundations of Statistics*, Wiley.

SCHLAG, K. H. (2007): "Eleven – Designing Randomized Experiments Under Minimax Regret," *Working Paper*.

SIMON, H. A. (1945): "Statistical Tests as a Basis for "Yes-No" Choices," *Journal of the American Statistical Association*, 40, 80–84.

STOYE, J. (2009): "Minimax regret treatment choice with finite samples," *Journal of Econometrics*, 151, 70–81.

TETENOV, A. (2012): "Statistical treatment choice based on asymmetric minimax regret criteria," *Journal of Econometrics*, 166, 157–165.

WALD, A. (1950): *Statistical Decision Functions*, Wiley.

WASSERSTEIN, R. L. AND N. A. LAZAR (2016): "The ASA's Statement on p-Values: Context, Process, and Purpose," *The American Statistician*, 70, 129–133.

ZILIAK, S. AND D. MCCLOSKEY (2008): *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, University of Michigan Press.