

Regulatory Errors with Endogenous Agendas

Daniel Carpenter Harvard University
Michael M. Ting Columbia University

How do a regulator's decisions depend on the characteristics and strategies of its external clients? We develop a theory of approval regulation in which an uninformed regulator may veto the submission of a better-informed firm. The firm can perform publicly observable experiments to generate product information prior to submission. We find that when experimentation is short, Type I errors (approving bad products) are more likely for products submitted by firms with lower experimentation costs (larger firms), while Type II errors (rejecting good products) should be concentrated among smaller firms. These comparative statics are reversed when experimentation is long. We perform a statistical analysis on FDA approvals of new pharmaceutical products using two different measures of Type I error. We find consistent support for the counterintuitive hypothesis that, under particular conditions, errors are decreasing in the size of the firm submitting the product.

Bureaucratic and regulatory organizations commit numerous errors—some trivial, some catastrophic. The standard metaphor used to understand these mistakes is that of “Type I” and “Type II” errors, borrowed from statistical decision theory. An agent commits a Type I error (or an “error of commission”) when she rejects an agreed-upon null (default) hypothesis that is true. She commits a Type II error (or an “error of omission”) when she affirms a null hypothesis that is false.

So understood, bureaucratic and regulatory errors involve crucial questions of politics, policy, and institutions. Welfare agencies may fail to issue a check to a deserving beneficiary (Type II) or pay someone who is not qualified (Type I). Space agencies may launch rockets that are doomed to destruction (I) or ground launches that would have gone off smoothly (II). Family service agencies may dissolve reasonably healthy families (I) or fail to remove children from abusive homes (II). Safety regulators may approve products that should never have been marketed (I) or reject or delay products that are socially beneficial

(II). A grant-making agency may reject a good project (II) while funding a mediocre one (I).

A substantial body of scholarly work has examined bureaucratic decision making and error. Much of its focus has been on how features of a case or of agency procedure lead to error or on the balance between Type I and Type II errors. Yet the problems that occupy bureaucrats and regulators only rarely come to them exogenously. Instead, the bureau's *agenda* may consist of cases that are actively pursued by private entities or other public actors through a process of submission (explicit or implicit). Citizens may bring complaints or charges to a prosecutor or an enforcement agency. Firms may submit license or new product applications. Researchers may submit patents or new research ideas embedded within grant applications.

This “bureaucratic agenda” factor—the extent to which bureaucratic choices (and hence errors) are shaped by external actors, whose submissions are themselves shaped by anticipation of what the regulator will do with them—has not been studied formally. The submission process is a crucial determinant of bureaucratic error

Daniel Carpenter is professor of government, Harvard University, Center for Government and International Studies, Harvard University, Cambridge, MA 02138 (dcarpenter@gov.harvard.edu). Michael M. Ting is associate professor of political science and public affairs, Columbia University, IAB Floor 7, 420 W. 118 St., New York, NY 10027 (mmt2033@columbia.edu).

We thank David Cutler, Tim Deering, James Fearon, Sven Feldmann, John Ferejohn, Tom Hammond, Lewis Kornhauser, Keith Krehbiel, Alison Lawton, Albert Ma, Caroline McGregor, David Meltzer, Ariel Pakes, Allison Sovey, Sonja Tong, Andrea Venneri, Natasha Zharinova, seminar participants at the University of Michigan, Stanford University, NYU Law School, Boston University, Harvard University, the University of Chicago, and Northwestern University, and panel participants at the 2001 annual meeting of the American Political Science Association and 2002 annual meeting of the Midwest Political Science Association for useful comments and discussions. Carpenter acknowledges the National Science Foundation (SES-0076452), the Robert Wood Johnson Foundation Scholars in Health Policy Program, and the RWJ Investigator Awards in Health Policy Research Program. Ting acknowledges the National Science Foundation (SES-0519082) for financial support. All errors, arguments, and interpretations are ours.

American Journal of Political Science, Vol. 51, No. 4, October 2007, Pp. 835–852

©2007, Midwest Political Science Association

ISSN 0092-5853

because it shapes the “pool” of cases considered by an agency. This pool can reduce the probability of error or can heighten it. Where, for instance, the submission pool is composed of cases that have a high Type I error probability, more errors of commission may result even without changes to administrative structure. When the bureaucratic agenda is composed of, for example, firm submissions that the regulator can approve or deny, the agenda will depend heavily on the firm’s expectations of the regulator’s strategy. Hence just as the pool of firm submissions shapes the regulator’s behavior, there is a feedback between the regulator’s behavior and the submissions. This may in turn affect the research and development strategy of the firm.

Our aim in this article is to consider—theoretically with a formal model and empirically with data on U.S. Food and Drug Administration (FDA) drug approval decisions—the implications of endogenous submissions for the study of bureaucratic error. The theoretical model is a game of incomplete information between a firm and a regulator. Both players are interested in the initially unknown quality of a product under development by the firm. Uncertainty over quality is mutual, with the firm initially better informed than the regulator. While both players wish to release high-quality products onto the marketplace, the regulator has a higher standard. This divergence in preferences may reflect the extent of regulatory caution. A marginally effective drug might be worthwhile to a pharmaceutical company because of its expected profitability, but not to the FDA. A regulator with a standard lower than the firm’s need not impose any hurdles in the approval process.

The game begins with a product development phase, in each period of which a firm may submit a costly application, perform an experiment to gather more information, or abandon the product. Each experiment is costly and takes the form of a publicly observable Bernoulli trial. If the firm submits, the regulator must accept or reject the proposal. The firm thus conditions its submission strategy on its private information and the endogenously acquired experimental information. Experiments therefore serve two purposes. As in standard “burned money” games, they are costly signals of private information. But they also provide public information that is used to determine whether a product is approved, or even submitted. Both players need experimentation to determine the product’s viability, and the regulator has an incentive to discourage the development of *ex ante* “bad” products.

In equilibrium, firms with the most favorable private information and initial experimental results tend to submit earliest. Those with less favorable results either withdraw or continue experimenting in the hopes of

improving their record. The equilibrium logic of the game is in some cases therefore the reverse of that of standard costly signaling games: “high” types incur low costs (i.e., burn less money), while lower types must perform more costly experiments in order to put themselves in the position to submit. Interestingly, the quality of approved products is often exactly the regulator’s reservation value. This happens because the regulator can only rarely separate between those types that meet its requirements and those that do not. She can deter the “worst” products from being submitted, but cannot screen out products that are just below standard.

Theoretically, our model shows that the firm’s costs for experimentation and submission affect the credibility of the products submitted to the regulator and hence the likelihood of the product’s acceptance. When the cost of submission is higher, so is the credibility of the submission, as the regulator judges the firm’s willingness to submit to be indicative of high quality. A high acceptance probability in turn generates a high probability of Type I error. But the relationship between the cost of experimentation and credibility is more complicated and depends on the experimental history of the product. To see this, consider the following logic. As costs rise, the regulator may reject more products when R & D is short because she knows that a firm with low-quality products will be tempted not to experiment. The high rejection rate lowers the probability of Type I error. The regulator may then infer when R & D is long that only high-quality products could remain in the population and accept all submissions. Alternatively, if experimental costs are low then she may infer negative information simply from the absence of a previous submission.

The model therefore predicts that errors will depend on the amount of experimentation conducted as well as firm characteristics, but not in a simple manner. A key result for Type I errors is a reversal of comparative statics on firm size (among the set of larger firms), depending on the length of R & D. When experimentation is short, the regulator will accept disproportionately more bad products from firms with low experimentation costs—these are likely to be the largest and oldest firms. However, when experimentation is long, acceptance of bad products is more likely for submitters whose experimentation costs are high (smaller, newer firms). Type I errors will rise with the cost of submission.

Since Type II errors are generated by rejections rather than acceptances, the comparative statics for Type II errors are exactly the reverse of those for Type I errors. Additionally, the model also predicts Type II errors that are generated through delay, as the inability of a firm to submit an acceptable product early can temporarily hurt

consumers. Such errors of delay will be more likely when the regulator has pessimistic prior beliefs about the quality of the proposal.

We assess these empirical predictions of the model in an analysis of FDA decision making. FDA approval of new drug products represents a highly revealing and consequential case in which to study the commission of regulatory error. In many cases, the mistakes of the agency can be cleanly observed. The process by which ideas and products are submitted is also cleanly observed. Finally, the process is repeated with sufficient frequency and diversity to allow aggregation of rich data on regulatory decision making and the incidence of several types of observable error. We examine FDA approval decisions for new molecular entities (NMEs) developed and submitted over the past 25 years. Using several different measures of Type I error, we assess the empirical correlation between regulatory error and the variables of interest predicted by our model. A crucial feature of our empirical analyses is the use of progressive sample restrictions through which we attempt to recreate the conditions for comparative statics obtained in our model. While necessarily observational, our results provide consistent support for our hypotheses on Type I errors.

Before elaborating our model, we devote our first section to explaining the importance of regulatory error and the limitations of previous attempts to study it. The second section presents the basic approval regulation model. The third section derives equilibria of the game and examines a number of testable predictions for regulatory policy, and the following sections present data for and estimations of some of these predictions. We then discuss the results and conclude.

Regulatory Error in Theory and Practice

Regulatory error merits theoretical analysis for at least two reasons. First, the errors can have drastic economic and policy consequences. Perhaps the most prominent example of a Type I error in pharmaceutical regulation became public in September 2004, when Merck announced the voluntary worldwide withdrawal of Vioxx (rofecoxib), a COX-2 inhibitor for the treatment of osteoarthritis and other indications. The stated reason for the withdrawal was that a Phase III trial of rofecoxib for patients with colon polyps had evinced a doubling of the rate of adverse cardiovascular events (stroke and heart attack) relative to the control group. The FDA soon published a controversial memorandum from one of its

epidemiologists, Dr. David Graham of the Office of Drug Safety, suggesting that use of rofecoxib as opposed to other COX-2 inhibitors was responsible for 27,000 to 55,000 deaths from 1999 to 2003.¹ As an example of a Type II error, critics of the FDA have long pointed to the delayed availability of so-called beta-blockers as a telling example of the costs of U.S. pharmaceutical regulation (Hilts 2003; Wardell and Lasagna 1975). By blocking beta-adrenergic receptors to which adrenaline binds in heart cells, beta-blockers can reduce effective cardiovascular “workload.” The late introduction of alprenolol as a treatment for essential hypertension and angina was estimated by Wardell to have resulted in over 10,000 deaths per year in the United States.²

Second, there are typically players outside of the firm-regulator “game” who are also affected by the trade-offs between Type I and Type II errors (e.g., Quirk 1980). Their payoffs are central to understanding the interest group politics surrounding regulatory policy. In the context of drug approvals, for example, disease advocacy groups may favor quicker approvals and care directly about Type II errors, while consumer safety groups may favor slower approvals and care directly about Type I errors.

Previous studies of administrative and regulatory error have generally fallen under three categories. An extensive literature on optimal stopping problems considers the trade-off faced by a single decision maker between accepting an experimental product immediately and waiting for additional (and costly) information (e.g., Dixit and Pindyck 1994; Kamien and Schwartz 1972; Moscarini and Smith 2001; Reinganum 1982). In the regulatory approval context, Carpenter (2004) considers a single decision maker facing an exogenous stream of experimental results. In this environment, the model predicts no asymptotic Type II errors: good products may be delayed, but in the long run will always be accepted eventually.

¹Memorandum from David J. Graham, MD, MPH, Associate Director for Science, Office of Drug Safety to Paul Seligman, MD, MPH, Acting Director, Office of Drug Safety entitled, “Risk of Acute Myocardial Infarction and Sudden Cardiac Death in Patients Treated with COX-2 Selective and Non-Selective NSAIDs,” September 30, 2004; at <http://www.fda.gov/cder/drug/infopage/vioxx/vioxxgraham.pdf> (accessed May 3, 2007).

²While plausible, Wardell and Lasagna’s claims were made without a rigorous statistical methodology, and we are aware of no replication to date which has confirmed or disconfirmed them. Hilts’ account, which estimates fewer deaths and places less blame on the FDA, is also quite threadbare in its particulars. Statement of William Wardell, M.D., Ph.D., at Hearings on the Food and Drug Administration’s New Drug Approval Process, Submitted to the Science, Research and Technology Subcommittee of the Committee on Science and Technology, United States House of Representatives, June 19, 1979.

A second line of argument has focused upon administrative structure and the benefits and limits of redundancy (Bendor 1985; Heimann 1997; Landau 1969; Ting 2003). These efforts focused on the effects of parallel processing (multiple procedurally independent agents) and serial processing (multiple predecision steps). The former structure reduces Type II errors, but increases Type I errors, while the latter does the reverse. Overall, however, redundant and seemingly inefficient administrative arrangements could be useful *ex ante* in minimizing organizational error, particularly for agencies governing risky technologies.

A final set of theories that can be usefully applied to organizational error invokes bounded rationality. Because decision making imposes computational costs, boundedly rational bureaucrats may satisfice rather than optimize (Simon 1968). They may employ shortcuts in learning, fail to collect all relevant information, or rely heavily upon frames in rendering decisions. These behaviors may collectively give rise to organization-wide maladies (e.g., Allison 1971; Bendor and Kumar 2005). Unfortunately, few formal models have used these approaches both to study agency failure as well as to distinguish between outcomes under fully and boundedly rational decision making.

These literatures mark the limits of our analysis. While they suggest natural extensions to our work, our model will produce systematic variations in errors even without their features. Instead, we focus on a common shortcoming of all of these explanations. To our knowledge, all formal analyses of Type I and Type II error have taken the agenda of the policymaker as exogenous. Thus the set of products (drugs, foods, energy technologies) or ideas (military or law enforcement doctrines) considered by the policymaker is unaffected by the anticipation of administrative decisions. Yet some of the most important influences on errors may come from the manner in which the policymaker induces “applicants” to advance or abandon their proposals. A fuller account of agency errors, then, demands an analysis of how bureaucratic agendas are shaped. The ensuing model comprises a first formalized attempt to do this.

The Model

Game Structure

Informational Environment and Players. There are two players: a firm (F) and a regulator (R). Both players are imperfectly informed about a parameter x of a product, which may be thought of as its “quality” or profitability if brought to market. We assume that x follows a Beta distribution, or $x \sim \beta(\theta, n)$, where θ and n are integers

with $1 < \theta < n$. The first parameter of the distribution, θ , is F’s *type*, that is, the estimated product quality prior to any research. Since F has only one product, we use “firm” and “product” interchangeably. The type takes one of two possible values, $\theta \in \{\underline{\theta}, \bar{\theta}\}$ where $\underline{\theta} = \bar{\theta} - 1$ and the probability that $\theta = \bar{\theta}$ is $p \in (0, 1)$. We will refer to $\underline{\theta}$ and $\bar{\theta}$ as the “low” and “high” types, respectively, where the former corresponds to a lower expected quality. The type is known only to F, though R knows p . The second parameter, n , is common knowledge.

A crucial feature of our model, then, is that no player is certain of the true product quality. Moreover, information is asymmetric because the firm’s initial estimate of x is more precise than the regulator’s. In this respect, the Beta distribution of quality is attractive because it offers a natural interpretation of a set of n Bernoulli trials, of which θ resulted in success and $n - \theta$ in failure.³ Additionally, it is flexible enough to accommodate a wide variety of “shapes” of the density function, as determined by θ and n . Given θ and n , the Beta distribution implies that x has a prior mean θ/n and prior variance $\theta(n - \theta)/(n^2(n + 1))$. The uncertainty over x can be reduced partially through observable experiments.

Sequence of Play: Development, Then Regulation. The game has up to four periods ($t = 1, 2, 3, 4$), denoted by subscripts. It begins in a *development* phase with up to three periods and possibly ends with a *regulatory* phase of one period.

In the development phase, F chooses an action $f_t \in \{S, W, E\}$ at $t = 1, 2$. S denotes a submission for approval, which ends the development phase and commences the regulatory phase the next period. W denotes a withdrawal from consideration, ending the game. Finally, E denotes an experiment to gather more data. An experiment is a single Bernoulli trial, which produces a publicly observable result $e_t \in \{0, 1\}$ corresponding to failure or success, respectively. For convenience we let $e_0 = 0$. Each experiment continues the development phase. F cannot experiment past the second period and thus must either submit or withdraw: $f_3 \in \{S, W\}$.

At the beginning of the regulatory phase, R knows F’s actions and experimental results and uses these to form expectations about θ and in turn x . Based on this, she makes a review decision $r \in \{A, R\}$, where A and R denote acceptance and rejection of the firm’s submission, respectively.

Information and Belief Updating. A key component of the model is the way in which experimental results affect the players’ expectations of product quality. (Without experiments, the model reduces to a simple signaling game

³The reputation games of Calvert (1987) and Alt, Calvert, and Humes (1988) use a similar technology.

with F as the sender.) The Beta distribution provides simple ways to calculate both the probability of an experimental success as well as the updated estimate of x . Beginning with a prior distribution of $\beta(\theta, n)$, the probability that the next experiment succeeds (i.e., $e_1 = 1$) is simply θ/n . An experimental record of n' experiments producing θ' successes generates a posterior distribution of $\beta(\theta + \theta', n + n')$, with mean $\frac{\theta + \theta'}{n + n'}$ and variance $\frac{(\theta + \theta')(n + n' - \theta - \theta')}{(n + n')^2(n + n' + 1)}$.

Since the parameters affecting the estimate of x change with experimentation, it will be useful to distinguish notationally between F's "initial" type θ and the numerical value of the first parameter of the Beta distribution by letting $m = \bar{\theta}$.

Utilities. F receives x if the product is approved, and zero for rejection or withdrawal. Each experiment costs c_e , and a submission costs c_s , where:

$$c_e \in \left(0, \frac{m(m-1)}{(n+1)(n+2)(n+m-1)}\right] \quad (1)$$

$$c_s \in \left(0, \frac{m}{n+2}\right]. \quad (2)$$

These assumptions ensure that the low type is not prevented from experimentation or submission by exogenous costs alone (though it may choose not to do so in equilibrium). They also substantially simplify the analysis by eliminating some trivial equilibria.

R receives $x - k$ for an approved product, and zero otherwise. The parameter k is therefore the divergence between the preferences of R and F. It can represent a "certainty equivalent" that a risk-averse citizen or legislature, fearing a product safety disaster, would demand from the firm in order for its product to be marketed. To rule out a number of trivial cases, let k satisfy:

$$k \in \left(\frac{m}{n}, \frac{m+1}{n+2}\right). \quad (3)$$

The lower bound guarantees that some experimentation is necessary to generate a product satisfactory to the regulator. The upper bound forces R to take experimental histories seriously. The high type becomes acceptable to R after one experimental success, while the low type needs two successes. Thus rejecting "early" (period 2) submissions is not a dominant strategy, and R may sometimes wish to accept a low-type product.

Equilibrium

We characterize mixed strategy Perfect Bayesian Equilibria (PBE) that satisfy a minor refinement. Let H_t represent the set of all possible experimental histories (or experimental results) prior to time t ; thus, $H_1 \equiv \emptyset$, $H_2 \equiv \{0, 1\}$, and $H_3 \equiv \{0, 1\} \times \{0, 1\}$. We use h_t to denote generic

elements of H_t . The equilibrium consists of the following three elements.

1. F's strategy is the triple (ϕ_1, ϕ_2, ϕ_3) , where $\phi_t : \{\theta, \bar{\theta}\} \times H_t \rightarrow \Delta(\{S, W, E\})$ for $t = 1, 2$, and $\phi_3 : \{\theta, \bar{\theta}\} \times H_3 \rightarrow \Delta(\{S, W\})$ map types and experimental histories to probability distributions over submitting, withdrawing, and experimenting (where feasible).
2. R's strategy ρ maps experimental histories, conditional upon a submission, into a probability of rejection. (Formally, $\rho : \cup_t H_t \rightarrow [0, 1]$).
3. R has beliefs μ mapping the experimental and submission history into a probability that $\theta = \bar{\theta}$. Formally, $\mu : \cup_t H_t \times \{\emptyset, S\} \rightarrow [0, 1]$. These beliefs must be consistent with Bayes' Rule along the equilibrium path of play. Out of equilibrium, $\mu(h_t, \cdot) = p$.⁴

Our analysis will utilize two other pieces of notation. First, we decompose $\phi_t(\theta, h_t)$ into probabilities of submitting, $\sigma(\theta, h_t)$, withdrawing, $\omega(\theta, h_t)$, and experimenting, $\eta(\theta, h_t)$. Second, we denote the expected quality of a period t submission (given beliefs $\mu(\cdot)$) by:

$$\begin{aligned} \bar{x}(h_t) &= E[x | h_t, f_t = S] \\ &= \frac{\mu(h_t, S) + m - 1 + \sum_{i=0}^{t-1} e_i}{n + t - 1}. \end{aligned} \quad (4)$$

This expression restates the posterior mean of a Bernoulli process with Beta-distributed priors. The numerator gives the number of "successes" and the denominator gives the number of "trials," and both incorporate R's prior information about x and the experimental history.

To simplify our presentation, we ignore "knife-edge" equilibria, which are not robust to small perturbations in parameter values. While we do not prove the result due to space considerations, the remaining equilibria are the only ones in which the counterintuitive strategy of withdrawing good products in the last period is not used.⁵ As the subsequent development shows, the predicted equilibria are unique for most, but not all, parameter values. This nonuniqueness does not affect the tested hypotheses of the model. Therefore, rather than imposing another refinement, we remain neutral on which equilibrium obtains in these cases.

⁴R therefore believes that each type is equally likely to submit out of equilibrium and so (3) implies that $\bar{x}(h_t) < k$ for any $h_t \notin \{1, (1, 1)\}$. These beliefs are relatively innocuous. Any sufficiently "pessimistic" beliefs would essentially support the equilibria identified here. If beliefs were more "optimistic," then R would accept all submissions for a larger set of histories. This would induce all types to submit at those histories. It is easily shown that no equilibrium can be sustained in this manner.

⁵The proofs of Propositions 1 and 2 rule out a number of candidate equilibria.

Main Results: Experimentation and Regulatory Agendas

We now turn to the main model. There are two types of equilibria—the *Early Submission Equilibrium* (ESE) and the *Late Submission Equilibrium* (LSE)—depending on how players react to a successful period 1 experiment.

General Features

We begin by deriving strategies that are shared by both the ESE and LSE. Throughout the game, each player's strategy can be stated in the following general terms. R's decision problem occurs when F submits. Clearly, R accepts a submission if its expected quality is greater than k , rejects if it is less, and is indifferent (and so may mix) otherwise. Thus:

$$\rho^*(h_t) = \begin{cases} 0 & \text{if } \bar{x}(h_t) > k \\ \in [0, 1] & \text{if } \bar{x}(h_t) = k \\ 1 & \text{if } \bar{x}(h_t) < k. \end{cases} \quad (5)$$

F's choice will depend on his assessment of the value of experimentation. Let $v(\theta, h_t)$ denote type- θ 's continuation value from experimentation, conditional upon experimental history h_t . Clearly, $v(\theta, h_3) = 0 \forall \theta$. He prefers submission over experimentation in period t if:

$$(1 - \rho^*(h_t)) \left[\frac{\theta + \sum_{i=0}^{t-1} e_i}{n + t - 1} \right] - c_s > v(\theta, h_t). \quad (6)$$

Finally, F prefers submission or experimentation over withdrawal in period t if the expected payoff from either is nonnegative.

The derivation of equilibrium strategies is simplified by a number of cases in which R does not need to distinguish between F's types. First, there are histories for which both types are either acceptable or unacceptable to R. By (3), R would reject any submission with no successes in the experimental history, and therefore neither type of F submits under any such history. If no successes have been achieved by $t = 3$, then the inability to experiment further forces a withdrawal. Likewise, R would approve any submission with two successes, and thus both types of F submit for $h_2 = (1, 1)$. Thus we have:

$$\rho^*(h_t) = \begin{cases} 1 & \text{if } h_t \in \{\emptyset, 0, (0, 0)\} \\ 0 & \text{if } h_3 = (1, 1) \end{cases} \quad (7)$$

$$\sigma^*(\theta, h_t) = \begin{cases} 0 & \text{if } h_t \in \{\emptyset, 0, (0, 0)\} \\ 1 & \text{if } h_3 = (1, 1) \end{cases} \quad (8)$$

$$\omega^*(\theta, (0, 0)) = 1. \quad (9)$$

Since these expressions hold in any equilibrium, they are omitted from the following discussion.

Second, when p is sufficiently high, there are histories for which the "average" quality of a submission when both firm types experiment and submit exceeds k , causing R to accept all submissions. This may occur for $h_2 = 1$ and $h_3 = (0, 1)$ or $(1, 0)$. It will therefore be useful to define the expected period 2 quality conditional upon $h_2 = 1$, given that both types experiment in period 1:

$$\tilde{x}_1 = \frac{m}{n+1} \left(1 + \frac{p}{m-1+p} \right) \quad (10)$$

Likewise, we define the expected period 3 quality given that $h_3 = (0, 1)$ or $(1, 0)$ and both types experiment and submit:

$$\tilde{x}_2 = \frac{m}{n+2} \left(1 + \frac{p(n-m)}{(n-m+1)(m-1)+p(n-2m+1)} \right) \quad (11)$$

The parameters \tilde{x}_1 and \tilde{x}_2 are important to R because they determine whether her acceptance strategy needs to distinguish between types. For example, if $\tilde{x}_1 \geq k$ and both types experiment and submit upon a success, then R accepts a submission after a single experimental success with probability one.

Two Equilibria

Early Submission Equilibrium. An ESE is an equilibrium in which F submits "early" (i.e., in period 2) with positive probability. Clearly, an early submission requires that F's first experiment be successful. There are two possibilities induced by the history $h_2 = 1$. These depend on F's quality threshold, k , relative to \tilde{x}_1 , the expected period 2 quality conditional upon both types experimenting successfully in period 1.

If $k \leq \tilde{x}_1$, then the prior quality distribution is high. R is thus willing in expectation to accept the set of all successful first-period experimenters, which includes low types. Since additional experiments are costly and cannot improve the odds of acceptance, both types submit and are accepted with certainty.

If $k > \tilde{x}_1$, then R would not accept the set of all first-period successes. However, by (3), she wishes to accept the high type and can choose a rejection strategy to deter the low type from always submitting. Early submission requires that an initially successful high type prefer

submission to continued experimentation. This occurs under the following condition, which we label *Early Submission* (ES):

$$c_s - c_e > \frac{m(m+1)}{(n+1)(n+2)}. \quad (12)$$

This condition is counterintuitive because it suggests that F submits early when c_e is low, even though F would appear to have an incentive to wait for a possibly better product. In equilibrium, this logic works only for the low type, which R wishes to discourage from submitting early. Deterring these submissions is more easily achieved if c_e is low, and hence the probability of early acceptance is decreasing in c_e . Thus for c_e sufficiently low, the high type prefers submission to the risk of a bad second experimental result.

If instead F's first experiment ends in failure, then by (8) she cannot submit. The low type may then either continue experimenting or withdraw. The incentive to continue experimentation is maximized when $\bar{x}_2 \geq k$; i.e., the ex ante expected quality is high enough that a history of (0, 1) is sufficient for R to accept, given that both types experiment.

A withdrawal will occur when c_s and c_e are high enough to outweigh any possible gain from approval. This occurs under the following condition, which we label *Early Withdrawal* (EW):

$$\frac{m-1}{n+1} \left(\frac{m}{n+2} - c_s \right) - c_e \leq 0. \quad (13)$$

When EW holds, the information asymmetry between players disappears because R infers that only a high type could remain after a failure. Note that by (1) and (2), EW implies ES.

The first result uses these observations to characterize the ESE strategies. Its derivation may be found in the appendix. Figure 1 graphically depicts F's equilibrium strategy.

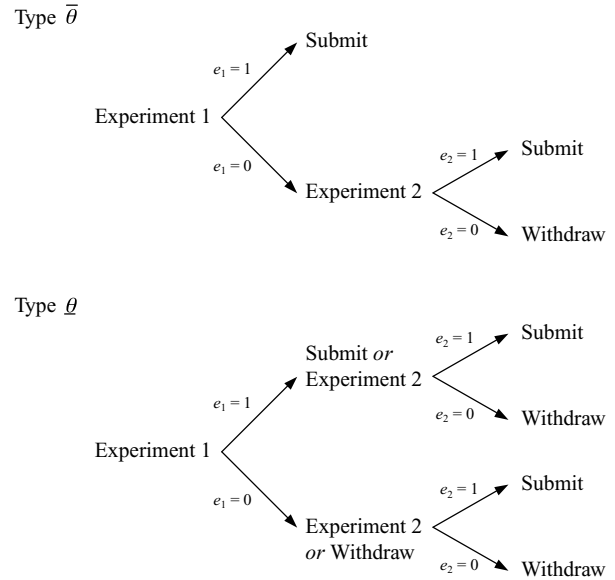
Proposition 1. (*Early Submission Equilibrium*) If $k \leq \bar{x}_1$ or ES holds, then there exists an equilibrium given by (7)–(9) and:

$$\text{For } F(\text{type } \bar{\theta}) : \eta^*(\bar{\theta}, \emptyset) = \eta^*(\bar{\theta}, 0) = 1, \sigma^*(\bar{\theta}, 1) = 1, \\ \sigma^*(\bar{\theta}, (0, 1)) = 1.$$

$$\text{For } F(\text{type } \underline{\theta}) : \eta^*(\underline{\theta}, 0)$$

$$= \begin{cases} 0 & \text{if EW holds} \\ 1 & \text{if EW does not} \\ & \text{hold and } k \leq \bar{x}_2 \\ \frac{p(n-m)m(m+1-k(n+2))}{(1-p)(n-m+1)(m-1)(k(n+2)-m)} & \text{otherwise,} \end{cases}$$

FIGURE 1 Firm Strategy in Early Submission Equilibrium.



Notes: With a type $\bar{\theta}$ product, F submits after one success. With a type $\underline{\theta}$ product, F may be forced to withdraw after one failure.

$$\omega^*(\underline{\theta}, 0) = 1 - \eta^*(\underline{\theta}, 0), \eta^*(\underline{\theta}, \emptyset) = 1, \sigma^*(\underline{\theta}, 1)$$

$$= \begin{cases} 1 & \text{if } k \leq \bar{x}_1 \\ \frac{pm(m+1-k(n+1))}{(1-p)(m-1)(k(n+1)-m)} & \text{if } k > \bar{x}_1, \end{cases}$$

$$\eta^*(\underline{\theta}, 1) = 1 - \sigma^*(\underline{\theta}, 1), \omega^*(\underline{\theta}, (1, 0)) = 1,$$

$$\sigma^*(\underline{\theta}, (0, 1)) = 1.$$

$$\text{For } R: \rho^*(1)$$

$$= \begin{cases} 0 & \text{if } k \leq \bar{x}_1 \\ \frac{n-m+1}{n+2} + \frac{n+1}{m}c_e - \frac{n+1-m}{m}c_s & \text{if } k > \bar{x}_1, \end{cases}$$

$$\rho^*(1, 0) = 1, \rho^*(0, 1)$$

$$= \begin{cases} 0 & \text{if EW holds or } k \leq \bar{x}_2 \\ 1 - \frac{n+2}{m} \left(\frac{n+1}{m-1}c_e + c_s \right) & \text{otherwise.} \end{cases}$$

Late Submission Equilibrium. An LSE is an equilibrium in which there are no submissions at $t = 2$. This may occur either because the expected quality of successful experimenters at $t = 2$ does not warrant acceptance (i.e., $k > \bar{x}_1$), or because an initially successful experimenter would prefer to gather more information (i.e., ES is violated). The existence of the LSE is assured when the ESE does not exist and furthermore does not depend on

the values of c_s or c_e . The LSE requires only that $k > \tilde{x}_1$, while the ESE requires that either $k \leq \tilde{x}_1$ or ES hold. Thus when $k \leq \tilde{x}_1$, the model uniquely predicts the ESE. And when $k > \tilde{x}_1$ and ES does not hold, the model uniquely predicts the LSE. If $k > \tilde{x}_1$ and ES holds, then both equilibria exist.

Analogously to Proposition 1, the next result characterizes the LSE strategies. Much of the proof is identical to that of Proposition 1, as the two equilibria are identical for histories starting with $h_2 = 0$. The derivation can be found in the supplementary appendix.⁶

Proposition 2. (*Late Submission Equilibrium*) *If $k > \tilde{x}_1$, then there exists an equilibrium given by (7)–(9) and:*

$$\text{For } F(\text{type } \bar{\theta}) : \eta^*(\bar{\theta}, \emptyset) = \eta^*(\bar{\theta}, 0) = \eta^*(\bar{\theta}, 1) = 1,$$

$$\sigma^*(\bar{\theta}, (0, 1)) = \sigma^*(\bar{\theta}, (1, 0)) = 1.$$

$$\text{For } F(\text{type } \underline{\theta}) : \eta^*(\underline{\theta}, 0)$$

$$= \begin{cases} 0 & \text{if EW holds} \\ 1 & \text{if EW does not} \\ \frac{p(n-m)m(m+1-k(n+2))}{(1-p)(n-m+1)(m-1)(k(n+2)-m)} & \text{hold and } k \leq \tilde{x}_2 \\ & \text{otherwise,} \end{cases}$$

$$\omega^*(\underline{\theta}, 0) = 1 - \eta^*(\underline{\theta}, 0), \eta^*(\underline{\theta}, \emptyset) = \eta^*(\underline{\theta}, 1) = 1,$$

$$\sigma^*(\underline{\theta}, (0, 1)) = 1, \sigma^*(\underline{\theta}, (1, 0))$$

$$= \frac{p(n-m)m(m+1-k(n+2))}{(1-p)(n-m+1)(m-1)(k(n+2)-m)}.$$

$$\text{For } R: \rho^*(1) = 1, \rho^*(1, 0) = \begin{cases} 0 \\ 1 - \frac{n+2}{m}c_s, \rho^*(0, 1) \end{cases}$$

$$= \begin{cases} 0 & \text{if EW holds or } k \leq \tilde{x}_2 \\ 1 - \frac{n+2}{m} \left(\frac{n+1}{m-1}c_e + c_s \right) & \text{otherwise.} \end{cases}$$

These two equilibria share several noteworthy features. First, costly activity by F can serve multiple purposes, thus distinguishing our model from standard costly signaling models. Here experimentation and submission are both signals of type, but experimentation also endogenously generates additional public information. This information, in addition to type, determines submission strategies.

Second, at histories $h_2 = 1$ and $h_3 = (0, 1)$, R may face the problem of distinguishing between types. The expected quality of submitted and accepted products is often R's reservation value of k . This makes R indifferent between rejection and acceptance. R therefore chooses rejection probabilities that screen out the low type by making

him indifferent between experimentation or submission and withdrawal. R thereby benefits from its gatekeeping power, as $k > \frac{m}{n}$ by assumption. However, the expected quality may exceed k if the proportion p of high types is large or low types withdraw early from consideration. Thus the approval regulation process allows R to “skim” the best products from a population that is ex ante unacceptable. A corollary is that, because of asymmetric information, R makes both Type I and Type II errors. Errors occur because submissions are pooled and R mixes between acceptance and rejection. This result contrasts with a decision-theoretic world in which R knows θ and can experiment on her own. In this world, no errors would ensue.

Predictions: Type I and Type II Errors

We now examine the model's predictions about the likelihood of ex ante regulatory errors.⁷ Recall that R commits a Type I error by approving a product with expected quality below k . Let $\psi(h_t)$ denote the equilibrium probability of a Type I error at each history h_t in which F submits with positive probability. For all such histories except $(1, 1)$, $\psi(h_t)$ is the same as the probability of accepting a low type, and so $\psi(h_t) = (1 - \rho^*(h_t))(1 - \mu(h_t, S))$. This expression makes clear the fact that Type I errors are correlated with acceptance rates.

We are primarily interested in comparative statics on c_e and c_s . There are three histories $(1, (1, 0))$ and $(0, 1))$ for which $\psi(h_t)$ varies with these parameters. Calculating $\mu(h_t, S)$ is facilitated by the fact that $\bar{x}(h_t) = k$ whenever the rejection probability is interior, as it is for these histories. (When the rejection probability is not interior, $\mu(h_t, S)$ must be calculated directly from F's equilibrium strategies.) Thus we may use (4) to obtain:

$$\mu(h_t, S) = k(n+t-1) - m + 1 - \sum_{i=1}^t e_{i-1}. \quad (14)$$

Substituting $\rho^*(h_t)$ from Propositions 1 and 2, we can then easily calculate the error probabilities that vary with costs at each history conditional upon submission:

$$\psi(1) = \left(\frac{m+1}{n+2} - \frac{n+1}{m}c_e + \frac{n+1-m}{m}c_s \right) \times (1 + m - k(n+1)) \text{ if } k > \tilde{x}_1 \text{ and ES holds.}$$

⁶The supplementary appendix may be accessed at http://people.hmdc.harvard.edu/~dcarpent/regerror_appendix.pdf.

⁷We restrict attention to errors from the standpoint of the regulator's preferences and do not consider the optimality of those preferences (e.g., Peltzman 1973, 1976). The focus on ex ante errors implicitly assumes that the regulator fares worse reputationally if it approves a product that it would not have wanted to had θ been known. Obviously, ex ante and ex post (i.e., occurring after the realization of x) errors are correlated.

$$\begin{aligned}\psi(1, 0) &= \frac{n+2}{m} c_s (1+m-k(n+2)) \text{ if } k > \tilde{x}_2. \\ \psi(0, 1) &= \frac{n+2}{m} \left(\frac{n+1}{m-1} c_e + c_s \right) (1+m-k(n+2)) \\ &\quad \text{if } k > \tilde{x}_2 \text{ and EW does not hold.}\end{aligned}$$

Note that all histories except $h_3 = (0, 1)$ are uniquely associated with an equilibrium type, while $(0, 1)$ may occur under either equilibrium and yields the same error probability.

Type I errors are increasing in c_s because of its role as a one-time signal of quality (implying higher acceptance rates). Likewise, c_e has a similar effect late in the development phase. In these cases, F's choice to bear costs conveys product quality and thus increases acceptance probabilities. This in turn raises the probability of Type I errors.

In some histories, however, the dynamics of experimentation can confound this logic. For example, if $h_3 = (0, 1)$ and EW holds, then only the high type submits and all submissions are accepted. The error rate is then zero and obviously does not depend on costs. Even more interestingly, for early submissions the effect of c_e on $\psi(h_t)$ can be the opposite of that for late submissions. Since c_e reduces the incentive of low types to continue experimentation, high values reduce R's acceptance probability, thus reducing the likelihood of an unwanted acceptance.

Because the specific history of clinical trials can be difficult to obtain, the tested comparative statics predictions are based on the length of the presubmission experimental history. Let $\bar{\psi}_\tau(\cdot)$ denote the probability of a Type I error conditional upon submission after τ periods of experimentation. Aggregating over possible values of p , it is straightforward to obtain the following comparative statics:

- *Short experimentation* ($\tau = 1$):

$$\begin{cases} \frac{\partial \bar{\psi}_1}{\partial c_e} < 0, \frac{\partial \bar{\psi}_1}{\partial c_s} > 0 & \text{if ES holds} \\ \frac{\partial \bar{\psi}_1}{\partial c_e} = 0, \frac{\partial \bar{\psi}_1}{\partial c_s} = 0 & \text{otherwise.} \end{cases}$$

- *Long experimentation* ($\tau = 2$):⁸

$$\begin{cases} \frac{\partial \bar{\psi}_2}{\partial c_e} = 0, \frac{\partial \bar{\psi}_2}{\partial c_s} > 0 & \text{if EW holds} \\ \frac{\partial \bar{\psi}_2}{\partial c_e} > 0, \frac{\partial \bar{\psi}_2}{\partial c_s} > 0 & \text{otherwise.} \end{cases}$$

Note finally that for all histories, Type I errors are decreasing in k and weakly increasing in c_s , and so at the margin raising the quality standard and reducing submission hurdles will decrease Type I errors. While it is not a

⁸The prediction that $\frac{\partial \bar{\psi}_2}{\partial c_s} > 0$ requires only that players do not *always* play the ESE when both equilibrium types exist. If they do play the ESE, then the history $(1, 0)$ cannot occur if ES holds. Since EW implies ES, $\psi(1, 0)$ does not enter the calculation of $\bar{\psi}_2$ when EW holds, and thus $\frac{\partial \bar{\psi}_2}{\partial c_s} = 0$.

direct prediction of the model, our analysis suggests that consumer protection groups such as Public Citizen, who place more weight on Type I errors, would support increasing regulatory standards and reducing submission costs.

While our empirical results do not focus on them, a similar analysis can be performed on Type II errors. R commits a Type II error by rejecting a product with expected quality above k . Given h_t and a submission, the equilibrium probability of a Type II error is simply $\rho^*(h_t) \mu(h_t, S)$, which for $h_3 \neq (1, 1)$ is the probability of rejecting the high type. Again, there are four histories—1, $(1, 0)$, $(0, 1)$, and $(1, 1)$ —for which submissions occur with positive probability. Clearly, no Type II errors occur whenever all submissions are accepted. For the other histories, the error probabilities conditional upon submission are easily calculated using (14) and Propositions 1 and 2.

Carrying out the exercise analogous to that for Type I errors, it is straightforward to show that the comparative statics on Type II errors mirror exactly those for Type I errors. Perhaps unsurprisingly, the same factors that increase acceptance rates (and Type I errors) reduce Type II errors. All values of c_e or c_s are therefore Pareto optimal. For all histories, Type II errors are increasing in k and weakly decreasing in c_s . Accordingly, our analysis suggests that disease advocates, by placing more weight on Type II errors, would likely support lower quality standards and higher submission costs.

Another kind of Type II error can occur if the submission of a product that R would accept is delayed. Due in part to the demand for treatments for HIV and other life-threatening illnesses in the 1990s, patient advocates have pushed extensively for procedures that would reduce approval times, and hence delay-induced errors.⁹ Our model provides a simple way of predicting submission delays. After one experiment, only the high type's product is acceptable to R. In the ESE, these products are submitted, while in the LSE they are not. An "unwanted" submission delay therefore occurs in the LSE with probability pm/n . The LSE requires that $k > \tilde{x}_1$, which implies a low value of p . Thus, low ex ante confidence in the product inhibits early acceptances.¹⁰

A final class of Type II errors occurs if F stops development of a potentially acceptable product. This does not occur in our model because abandonments only oc-

⁹In response, regulations passed in 1992 and 1997 allowed the use of "surrogate markers"—which are less conclusive than the traditionally used "clinical endpoints"—to establish the efficacy of a new drug. See Section 112 of the FDA Modernization Act of 1997 (21 USC 351), and the Accelerated Approval Rule (21 CFR 314, 601).

¹⁰If players expected the ESE to be played when both equilibrium types exist, then the LSE also becomes less likely when ES holds (i.e., low experimentation and high submission costs).

cur when the low type withdraws after one failure. By assumption, such products could never be acceptable to R. However, if the model were extended to allow more experimentation, or cases where $k < m/(n + 2)$, then a low firm type might be induced to abandon a potentially good product after an initial experimental failure.

Data

Our statistical data for testing the model's predictions are drawn from a larger database of 32,216 molecular entities that have been developed to some stage of product maturity over the past 30 years under the system of pharmaceutical regulation in the United States. In part because U.S. regulatory authorities track drugs from preclinical stages to postapproval experience on the market, U.S. pharmaceutical development and regulation are some of the most heavily documented political-economic activities in the world, offering large amounts of fine-grained data for analysis.¹¹ We have collected these data from a variety of sources, including Freedom of Information Act requests to the FDA, FDA archives, *FDA Annual Reports* from its drug reviewing divisions, *F-D-C (Food, Drug, and Cosmetic) Reports*, and the trade database PharmaprojectsTM. Where our data concern recent products, we rely more heavily upon trade reporters and databases. Where our data concern older drugs, we rely much more heavily upon FDA sources, checked against the data of Dranove and Meltzer (1994) and PharmaprojectsTM.

Of the 32,216 molecular entities in total, 15,282 are "preclinical" drugs that have not yet been tested on humans. Of the remaining 16,934, we have reliable clinical development data on 16,723, of which 16.7%, or 2,789, were eventually submitted to the FDA. The others have been abandoned or are currently in a limbo R & D status. Our analyses here sample only the approved drugs, though it is worth remarking that a richer estimator would consider the quasi-selection effect by which some drugs get submitted and others get abandoned.

Measuring Type I Errors. We confine our statistical analysis to Type I errors or "bad" approvals. The reason for doing so concerns the complexity of Type II errors. Coding Type II errors as drugs that are rejected in the United States but accepted and prescribed with relative success elsewhere would yield a very small set that precludes statistical analysis with regressors. The reason is that most drugs that are approved in overseas markets are

eventually approved in the United States. It would also be almost impossible to know what drugs were wrongly rejected everywhere, as this would require information that was presumably unavailable to regulators and medical authorities. Hence Type II errors are better quantitatively measured in terms of delay, not absolute rejection, though this task presents considerable difficulties as well (Carpenter 2002, 2004; Dranove and Meltzer 1994).

Even restricting our analysis in this way, measurement of Type I error presents some difficult questions. How would we know, after the fact, that the FDA had made a bad approval decision? There is no way of knowing this with certainty, but several observables would seem to be correlated with such events.

First, a Type I error is more likely to have occurred when, after the approval of a new molecule, the FDA significantly revises the drug's labeling. That is, the FDA must attach evidence of new contraindications, or new side effects, that are serious in some way. This was precisely the intuition of the U.S. General Accounting Office (GAO) when it produced a report on "postapproval risks" of FDA-approved drugs (U.S. GAO 1990). This report determined whether there had been a "significant" labeling revision for the drug, and if so, what those labeling revisions were. In part to ensure that the subjectivity of coding Type I errors is placed elsewhere, we adopt straightforwardly the GAO assessments for the dependent variable *GAO Lines*. *GAO Lines* is the number of lines of text describing significant labeling revisions in the report and is scored 0 for all unrevised drugs. This variable amounts to an implicit assumption that the Type I error was worse where more ex post revisions were added to the labeling, as detected and reported by the GAO. One drawback of these data is that we can only focus on NMEs approved from 1976 to 1985, and then only on the 198 drugs that the GAO selected for study. We are missing firm data on two of the drugs in the report, making our effective sample size 196 molecular entities.

Second, if international regulatory decisions are at least somewhat independent of the FDA's, then we could infer Type I errors from approval and recall decisions for overseas markets for the same molecule, after approval by the FDA. Our second set of measures uses international comparisons in precisely this way. We define two indicators of a Type I error (many others are possible). We first code whether a drug was removed from at least one foreign market in a "highly developed" (HD) country (*Withdrawn*). One can adjust the cutoff here by specifying the number of withdrawals in HD countries (2, 5, 10, or more) necessary for the FDA's approval to qualify as a Type I error. Next, we also examine those HD markets where the drug never entered. If the drug failed to enter five or more markets in highly developed countries

¹¹ Scholars have only recently begun to analyze pharmaceutical regulation and development using the richest data available, so the extensiveness of documentation has not been accompanied by extensiveness of study.

(*No Entry*), then it is likely that overseas regulators were less enthusiastic about the drug, and that they wished to avoid a Type I error that the FDA may have committed in approving the molecule.

Measuring Experimentation Cost (c_e). There are two dimensions along which c_e can be thought to vary. The first, which we use here, is by firm (sponsor). One can think of experimentation cost as inversely related to firm size and experience. Larger firms should face lower experimentation costs, in part because they have (probably) already developed products with similar R & D processes in the past, in part because R & D obeys economies of scale. We measure size and experience jointly, by calculating the number of previous investigational new drugs (INDs) developed at the date of submission of the new drug application (NDA) for the molecular entity in question to the FDA. We label this variable *Firm INDs* and the log of this variable *Log Firm INDs*. This variable has the advantage of changing over time, by firm, allowing for firm-specific effects in our estimations. This measure is also less sensitive to shocks than size measures based upon sales or revenues, which are highly dependent upon patent expiry and other “denominator” effects.¹²

Measuring Submission Cost (c_s). It is more difficult to produce a finely grained measure of submission cost that would differ from experimentation cost. One tempting possibility is to consider onerous aspects of the regulatory process as a sort of submission cost. For instance, the FDA imposes hundreds of stylistic and procedural requirements upon new drug applications, and if the agency judges that a firm has not met these requirements, it can issue a “refusal-to-file” (RTF) judgment. One testable implication is that, where these requirements become more stringent, empirically observed acceptance probabilities will rise, *ceteris paribus*. However, changes in these regulatory submission requirements are difficult to measure other than by indicator (dummy) variables indexing when the requirements changed, and changes in these requirements are so common as to render this infeasible.

The larger problem here is that any firm-level variable measuring submission cost is likely to be highly correlated with firm-level covariates of experimentation cost. Since the estimated costs of clinical development are much greater than the estimated costs of regulatory submission (Dranove and Meltzer 1994; Wardell and Lasagna 1975),

we think it is preferable to use firm size to measure experimentation cost and not submission cost.

We focus on a more blunt measure, namely the influence of the Prescription Drug User Fee Act (PDUFA) of 1992 and its expected effects upon drug approval times (Olson 2000). The law requires firms to submit a user fee with each new drug application to the FDA (in FY 2004, \$573,500 for a new drug application requiring analysis of clinical data; *Federal Register* 68(148), August 1, 2003, pp. 45, 249–45, 252). The PDUFA dummy variable is scored 1 for products submitted after 1992.

While PDUFA may have raised nominal submission costs by requiring all sponsors to pay for each new NDA, in fact the law was widely predicted to reduce firm costs by accelerating FDA drug review. As part of the deal that secured the PDUFA legislation, it was agreed that the proceeds from user fee payments would go to hire more reviewers at the FDA’s Center for Drug Evaluation and Research. It was understood by all that the explicit goal (and probable effect) of this legislation would be to reduce FDA review times for new drug applications. As one report described the legislation in 1992:

In return for collecting what is expected to be about \$300 million over the next five years from companies that want their drugs reviewed, Dr. David Kessler, the Commissioner of Food and Drugs, agreed to hire 600 new examiners to speed drug approval . . . The plan is considered politically feasible because . . . F.D.A. estimates that companies may earn an average of \$10 million for each additional month they have a drug on the market.¹³

As part of the 1992 agreement, the FDA agreed to review 90% of all “standard” new molecular entities within 12 months or less by 1997, and agreed to review 90% of “priority” drugs within six months or less. In part because of the review time goals and in part because of the enhanced resources, approval times did fall appreciably in the years following PDUFA’s enactment. Importantly, while FDA review times fell following PDUFA’s enactment, clinical development (experimentation) times did not.¹⁴ Hence the user-fee regime seems to have affected submission costs much more than pre-NDA experimentation costs.

¹² Another plausible source of variation in c_e is by disease. Diseases that have a higher hospitalization burden are typically more costly per patient in clinical trials, in part because the trial will often have to pay the cost of hospitalization for treatment and placebo arms. By asking how costly a clinical trial is for patients with a given disease it may be possible to “back out” reasonable estimates of c_e from aggregated grants or clinical trials data. We do not attempt this approach here.

¹³ See Philip J. Hils, “Senate Passes Bill to Charge Makers for Drug Approval,” *New York Times* October 8, 1992.

¹⁴ See U.S. General Accounting Office, *FDA Review Time for Drugs Has Decreased in Recent Years*, GAO/PEMD-96-1 (Washington: GAO, October 1995), and Carpenter et al. (2003).

The remaining question is whether variation in approval times properly measures variations in submission cost. Here it is important to remember that, when a drug is submitted to the FDA, three phases of clinical trials have been completed.¹⁵ Hence the direct experimentation costs have already been expended. A reduction in FDA approval time then clearly reduces a drug's *development* cost but does not by definition influence the cost of experimenting with the drug. Hence a reasonable inference is that the user-fee regime has reduced the submission-cost component of drug development cost, and not the experimentation-cost component.

Since the user fees are quite small (ranging from \$240,000 to about \$1,000,000) relative to the capitalized benefits of earlier approval (Carpenter 2004; Carpenter et al. 2003), the implicit benefits of acceleration outweigh the explicit increase in user-fee cost, and so the onset of PDUFA should be conceived as having reduced c_s .

While a firm-level measure of submission cost would be superior, and while our measure is admittedly blunt, we believe that any firm-level measure of submission cost would pick up variation in experimentation cost. We acknowledge the weaknesses of our measure, as other time-varying factors could be influencing the FDA's error rate while not necessarily measuring submission cost.

Measuring Early and Late Submission Regimes. Because many of the model's comparative statics depend on the length of experimentation, we often restrict the sample to only those products with sufficiently long or short development times, and to only those products submitted by sufficiently large or small companies. Beginning with development time restrictions, for each drug, the variable *IND Time* measures the number of months it held IND status. To get at drugs that have been developed for a greater or lesser time, we compute quantile information for the *IND Time* distribution and use these quantiles as cutoffs. For example, if the mean of the *IND Time* is 72 months, one way of differentiating "long experimentation" from "short experimentation" is to consider all drugs for which *IND Time* < 72 as "short" and all drugs for which *IND Time* > 72 as "long." As a robustness check and as a way of revealing theoretically relevant patterns in the data, we replicated our analyses across numerous cutoff points at different quantiles.¹⁶ We conduct an analogous set of "regime" restrictions for the *Log Firm INDs* variable.

¹⁵In rare instances such as under "accelerated approval" rules, drugs are submitted after Phase II clinical trials.

¹⁶Not all of these analyses are shown here. A fuller set of results is available from the authors upon request.

TABLE 1 Summary Statistics

Variable	Mean	SD	N
<i>GAO Lines</i>	5.60	8.80	196
<i>Log Firm INDs</i>	1.91	0.96	196
<i>Hospitalization Length</i>	2.84	3.84	196
<i>Withdrawn</i>	0.061	0.241	1071
<i>No Entry</i>	0.029	0.168	1071
<i>Log Firm INDs</i>	2.86	1.53	1071
<i>Hospitalization Length</i>	2.14	3.77	1071

Controls. For all analyses, we control for firm fixed effects.¹⁷ Where relevant, we also control for a brute time trend by including the year in which the NDA was submitted (*Submission Year*). The final control we use throughout is a blunt but effective measure of the severity or costliness of the disease targeted by the drug in question. For the primary indication of the drug, we calculate from the U.S. federal government's Health Care Utilization Project data, the average number of days of hospitalization per hospitalization. We call this variable *Hospitalization Length* and assign a value of zero whenever there are no recorded hospitalizations for a disease. This variable covaries positively with the mortality of a disease and its expected cost of treatment. It is also negatively and significantly associated with FDA approval times for NMEs (Carpenter 2002).

Summary. Descriptive statistics for the most important variables appear in Table 1. The first half of the table details the GAO data. Exactly half of in-sample NMEs approved from 1976 to 1985 were subject to "serious" postapproval labeling revisions and therefore had nonzero *GAO Lines* values. The maximally revised labeling had 48 lines.

The second half of the table covers the international comparison data. Here the frequency of Type I error as coded by international comparison is much smaller. Only 6.1% of NMEs approved by the FDA from 1984 to the present were approved but then withdrawn in one or more highly developed markets (*Withdrawn*). The fraction of entities approved by the FDA but that failed to enter five or more highly developed foreign markets is 2.9% for this period (*No Entry*). Notice that because our international comparisons data is of more recent vintage, the firm size variable has a sample distribution with higher mean and higher variance than the sample distribution for drugs studied by the GAO.

¹⁷We also control for disease category fixed effects, e.g., cardiovascular drugs, oncologic drugs, drugs for the central nervous system, etc. These results are not shown here but yield substantially similar results and are available from the authors upon request.

TABLE 2 Predicted Effects on Type I Error

	Low IND Time		High IND Time	
	Firm Size	Sub. Cost	Firm Size	Sub. Cost
Low <i>Firm INDs</i> (high c_e)	0	0	0	+
High <i>Firm INDs</i> (low c_e)	+	+	–	+

Estimation and Results

We test the comparative statics of the model in a multivariate setting, using a mix of robust general linear models and maximum likelihood estimators. Some of the hypotheses are counterintuitive and depend on subsamples. To differentiate subsamples from one another, we use cutoffs given by the sample means of (logged) previous firm submissions and development time (in months; see Table 1). We do this in order to construct an artificial sample similarity with empirical regimes with low experimental cost and longer experimentation. Additionally, our measure of firm size, *Firm INDs*, is inversely related to c_e . Table 2 summarizes the predicted coefficient signs by subsample.

GAO Data

We begin with analysis of the GAO data. In Table 3, we regress *GAO Lines* on the relevant regressors using negative binomial regression.¹⁸ By Table 2, we have four predictions. First, for drugs that are submitted relatively early and by low- c_e firms, we predict that Type I error is increasing in *Firm INDs*. In specification (1) of Table 3, we present estimates derived from two truncated samples restricted to larger firms and drugs with shorter development times. There is some support for this hypothesis. The coefficient on *Firm INDs* is generally positive but attains statistical significance at the $p < 0.05$ level only in the smaller subsample where investigational development time is 72 months or less.¹⁹ The coefficient estimates in the sample with IND time less than 96 months is not statistically differentiable from zero. If the larger of these two estimates is used, a one standard deviation hike in previous INDs (39.4 INDs previously submitted) is associated with a 2.8-line increase in the length of GAO listing.

Our second prediction is that, for products characterized by longer experimentation times and submitted

by larger firms, Type I errors should be decreasing in c_e . By the results of the third section, long experimentation and low c_e should be associated with a *reversal* of the comparative static on *Firm INDs*. In specification (2) of Table 3, we present estimates derived from truncated samples in which analysis is restricted to larger firms and drugs with longer development times. As predicted, the coefficient estimate on *Firm INDs* is negative, and this estimate is statistically significant. Here the marginal effects are smaller, about one-half-line increase in the GAO listing for the larger subsample (with development time greater than 72 months), and only a one-eighth-line increase in GAO listings for the smaller subsample.²⁰

The final two predictions correspond to the high c_e empirical regimes, where the model predicts that firm size should have no effect. Specifications (3) and (4) of Table 3 show that we cannot reject the null of zero, which modestly supports the hypotheses of zero association.

Although we do not report them here, these results are generally replicated if we use generalized linear regressions with heteroskedasticity corrections for these variables. In particular, regressing *Log GAO Lines* (where $\text{LogGAOLines} = \ln(1 + \text{GAOLines})$) upon *Log Firm INDs* yields results that are substantively identical to those of the maximum likelihood models.

International Data

We now turn to the sample of drugs approved over the past 25 years ($N = 1071$) and employ our measures of Type I error as coded via international comparison. Again we use the relevant sample mean *Log Firm INDs* (approximately 3) as a cutoff in splitting subsamples from one another. Aside from the larger sample, the primary benefit of this dataset is that it allows us to test for the effects of the PDUFA dummy, *PDUFA*. We predict that PDUFA should decrease Type I errors under both short and long experimentation. We again include firm fixed effects. This results in a material reduction in sample size for these models, mainly because the fraction of ones in the sample dependent variable is so small.²¹ Hence any time that a firm has never had a foreign withdrawal in our dataset, or anytime that a firm has never been denied

¹⁸We also analyzed the data using a binary dependent variable, that is, examining whether or not the drug was listed at all in the GAO report. This produces results similar to those reported here.

¹⁹This is due in part to the inclusion of fixed effects, which raise the standard error associated with *Log Firm INDs*.

²⁰It merits note that in these samples the development time data are left-skewed such that movements that are several multiples or more of the standard deviation are common.

²¹We have also estimated extreme-value regression models (which embed the assumption of asymmetric link functions) for these samples. The extreme-value regression estimates do not change substantively; they remain statistically differentiable from zero and retain their signs. Estimated marginal effects are also similar to those retrieved from probit regressions. See the supplementary appendix for the results.

TABLE 3 Negative Binomial Regression Analyses of GAO Coded Error

Dependent Variable: <i>GAO Lines</i>								
	(1)		(2)		(3)		(4)	
<i>Firm INDs</i>	0.173 (0.071)	0.019 (0.032)	−0.026 (0.014)	−0.059 (0.023)	0.085 (0.362)	0.440 (0.287)	0.036 (0.125)	−0.246 (0.149)
<i>Hospital. Length</i>	−0.185 (0.086)	−0.145 (0.039)	−0.016 (0.046)	−0.023 (0.079)	−0.004 (0.095)	0.070 (0.082)	−0.160 (0.049)	−0.211 (0.080)
<i>Submission Year</i>	−0.054 (0.934)	0.206 (0.106)	0.030 (0.010)	0.075 (0.024)	−0.164 (0.171)	−0.131 (0.079)	−0.003 (0.009)	−0.008 (0.008)
Constant	107.03 (184.50)	−406.60 (210.27)	−56.44 (18.82)	−145.23 (47.95)	326.79 (339.27)	259.93 (156.77)	6.704 (18.85)	17.57 (17.23)
<i>N</i>	18	34	66	50	19	31	93	81
<i>Log Firm INDs</i>	>2	>2	>2	>2	<2	<2	<2	<2
<i>IND Time (mos.)</i>	<72	<96	>72	>96	<72	<96	>72	>96

Note: All models have firm fixed effects and heteroskedasticity-consistent standard errors.

TABLE 4 Probit and GLS Regression Analyses of Internationally Coded Error

Dependent Variable: <i>Withdrawn</i>								
	(1)		(2)		(3)		(4)	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
<i>Log Firm INDs</i>	0.016 (0.043)	0.371 (0.453)	−0.042 (0.023)	−0.362 (0.179)	0.019 (0.018)	0.497 (0.327)	−0.004 (0.018)	−0.041 (0.155)
<i>Hospital. Length</i>	−0.006 (0.004)	−0.066 (0.063)	−0.005 (0.003)	−0.068 (0.043)	−0.004 (0.003)	—	−0.004 (0.003)	−0.037 (0.027)
<i>PDUFA</i>	−0.131 (0.053)	—	−0.036 (0.044)	—	0.027 (0.031)	0.560 (0.564)	−0.031 (0.026)	−0.405 (0.435)
Constant	0.096 (0.155)	−2.226 (1.842)	0.233 (0.089)	−0.037 (0.845)	−0.008 (0.027)	−2.985 (0.776)	0.060 (0.027)	−1.551 (0.238)
<i>N</i>	151	45	286	205	132	78	457	360
<i>Log Firm INDs</i>	>3	>3	>3	>3	<3	<3	<3	<3
<i>IND Time (mos.)</i>	<120	<120	>120	>120	<120	<120	>120	>120

Note: All models have firm fixed effects and heteroskedasticity-consistent standard errors.

entry into five or more foreign markets, the inclusion of an indicator variable for this firm produces a perfect prediction, and the variable is dropped from the probit analysis.

We first test our hypotheses by estimating GLS and probit regression models. Table 4 presents linear probability and probit regression results for the dependent variable *Withdrawn*. Here the results are quite interesting for the *Log Firm INDs* variable. In specification (1), where the sample is restricted to drugs produced by large firms and excludes drugs with long IND times, the coefficient estimates are positive but not statistically significant. However, in specification (2), where the sample is restricted to large firms and long IND times, we

observe negative coefficient estimates, as predicted, and these estimates are statistically differentiable from zero.²² Marginal effects computations from the statistical model suggest that, in the “late-submission equilibria” empirical regimes where we observe statistically significant associations, a one standard deviation increase in *Log Firm INDs*

²²A direct statistical comparison of these coefficient estimates, or a joint test, would not be a valid test of the model, since one cannot compare comparative statics across equilibria in a Bayesian game. We can only test whether the coefficient estimates are positive and nonzero in one regime, and negative and nonzero in the other regime. In many of these subsamples collinearity induces the dropping of *PDUFA*, but when it is included it is never statistically distinguishable from zero.

TABLE 5 Binary Analyses of Internationally Coded Error (Effects of PDUFA)

	Dep. Var.: <i>Withdrawn</i>		Dep. Var.: <i>No Entry</i>	
	Probit	LPM	Probit	LPM
<i>Log Firm INDs</i>	0.103 (0.096)	0.007 (0.008)	0.139 (0.157)	0.005 (0.005)
<i>Hospitalization Length</i>	−0.049 (0.021)	−0.004 (0.002)	−0.127 (0.048)	−0.004 (0.001)
$\Delta E(\text{Approval Time})$	0.170 (0.126)	—	0.674 (0.195)	—
$\Delta E(\text{Approval Time})^*$ (*instrumented by <i>PDUFA</i>)	—	0.009 (0.009)	—	0.020 (0.006)
Constant	44.26 (23.05)	4.90 (2.59)	−10.85 (34.66)	−1.27 (1.51)
<i>N</i>	956	1026	851	1026

Note: All models have firm fixed effects and heteroskedasticity-consistent standard errors.

is associated with a 9.0 percentage point decline in Type I errors (this corresponds to a 97% decline in the sample rate of error). Finally, in the third and fourth empirical regimes (high c_e , or low INDs), where the model predicts no variate relationship of experimentation cost and error, we cannot reject the null of zero.

The estimates in Table 4 also allow for a partial test of our hypotheses on submission costs. If PDUFA lowered submission costs, it may have also induced a lower Type I error rate at the margin. Across the various subsamples defined by the submission and withdrawal conditions, there is inconsistent support for this hypothesis from Table 4. Specification (1) yields the predicted statistically significant negative coefficient estimate. Specifications (2) and (4) have the correct sign, but we cannot reject the null hypothesis of no relationship. Finally, the model predicts no relationship in specification (3), and we cannot reject that hypothesis.

Better traction on this hypothesis is had from examining the “full sample” for effects of secular changes in the cost of submission. Because c_s has a nonnegative comparative static in all cases, we can combine the cases and test for a positive relationship between submission cost and Type I error (equivalently, a negative relationship between PDUFA and Type I error). Table 5 shows the results of four regressions. In the first two specifications, the dependent variable is the same as in Table 4, *Withdrawn*, while the last two use *No Entry*. All models include a measure of submission cost as well as a time trend, which is the year in which the molecule in question commenced development. Some plausible evidence for the validity of these indicators comes in the fact that the measured covariates are of the same sign across specifications, significantly so for the hospitalization and time trend covariates. Notice

that once the full sample is examined without theoretically relevant conditions and restrictions, simple linear inclusion of *Log Firm INDs* does not add to the model’s predictive power and does not yield coefficient estimates statistically differentiable from zero.

In the probit regressions (first and third columns), submission cost is measured as the (differenced) series in median approval time, calculated for the submission year of the molecule in question as $\Delta E(\text{ApprovalTime})$. In toto, this represents firms’ generally unconditioned expectations about the time- or delay-cost of their submission. In the linear probability models, we instrument for $\Delta E(\text{ApprovalTime})$ with *PDUFA*.²³ The idea here is that PDUFA represents a shift in expected approval times that was anticipatable by submitting firms. Here we find more robust associations between the instrumented change in median approval time and Type I error rate.

Rather strikingly, the estimates suggest that reductions in approval time may be associated with less Type I error, not more. All four estimates are positive, but they are statistically significant only in the *No Entry* regressions. This is counterintuitive in light of recent debates and arguments which suggest that PDUFA may have raised the occurrence of Type I errors (see Avorn 2004). Since our data offer no measure of severity and our model assumes homogeneity of Type I error, we cannot make statements here about PDUFA’s welfare effects. Additionally, these

²³This was done as follows. For each drug in the more recent sample, we calculated the (annual) change in mean approval time for all drugs submitted in the same year. We then regressed this variable upon *PDUFA* and retrieved the predicted values of the change in approval time. This variable enters the linear probability model regressions as a right-hand-side covariate; it is unwarranted to use this instrumented variable as a linear regressor in the probit models. We do not present the corrected standard errors in the table here.

relationships are hypothesized to hold at the margin, and so PDUFA may have had effects on other institutional parameters that influence error but are not captured by our model.

Summary and Caveats

We find appreciable support for our primary hypotheses, though the pattern of support differs based upon the hypothesis examined. For our more intuitive hypothesis—that where experimentation cost is low and development times are shorter, Type I regulatory error is increasing in firm size or experience—we find mixed support, with positive coefficient estimates that are statistically significant only in some subsamples. For our more counterintuitive hypothesis—that where experimentation cost is low and development is longer, increases in firm experience are accompanied by reductions in Type I error—we find consistent support across datasets and across specifications. This inversion of statistical associations occurs across two datasets and in each of two measures of Type I error in each dataset.²⁴ The results of Tables 3 and 4 lend solid support to the predicted reversal of the comparative static on c_e across the low- c_e subsamples.

We can test for effects of submission cost only brutally, and only in our second set of analyses on more recent data. There, the coefficient estimates for $\Delta E(\text{ApprovalTime})$ in Table 5 broadly support the hypothesis that by reducing the cost of submission, PDUFA lowered the probability of Type I errors. The coefficient for expected approval time is positive in all estimations and significant in two, as well as in a variety of specifications not reported here. If it can be safely interpreted that PDUFA implicitly reduced submission costs, then we have evidence that Type I error rates are decreasing in c_e . However, all of our causal leverage here is nonlinear over-time variation, so anything else that might have occurred in the past 10 years, that covaries with PDUFA but not with a linear trend, and that also covaries with Type I error rates, would confound our inferences. Among these, better technology for detecting drug hazards *ex ante* is one plausible candidate.

We conclude with two observations. First, we do not attempt to attach any theoretical meaning to the estimated effect of *Hospitalization Length*. However, it should be noted that *Hospitalization Length* is only weakly associated with error as coded via international comparisons, but sometimes negatively and robustly associated with GAO-coded errors. Second, one weakness with *No Entry* is that it may be determined as much by firms' decisions to avoid market entry as by regulators' decisions to deny

entry. If this is true, and if a significant component of *No Entry* is affected by firm-level variables that are not responsive to anticipation of regulatory strategy (e.g., portfolio factors in the company's pipeline), then it is possible that our inferences are subject to error from confounding variables.

Discussion

Regulatory decision making and error are frequently shaped by an endogenous agenda, which is determined by a firm's anticipation of the regulator's behavior. In this context, an important determinant of regulatory approval is the credibility of the firm's submission. The theory produces several counterintuitive predictions about the interaction between costs and the dynamics of research and development choices. To date, these predictions have not been generated by existing approaches to agency failure, either structural or cognitive. Empirical tests of some predictions about Type I error are supported by data on pharmaceutical relabeling, safety-based withdrawals, and product introductions.

These results may be useful for guiding both theoretical and empirical research on a large class of regulation and political economy problems. Two of these merit remark here. First, government grant making and procurement represent an appealing policy arena in which to apply the model. In the United States alone, the National Science Foundation and National Institutes of Health award over \$23 billion in grants and contracts annually, while the Department of Defense spends over \$84 billion on procurement annually (figures taken from the *Budget Request of the President of the United States*, FY 2007). Over \$100 billion per year, then, is spent according to mechanisms that resemble the game laid out here. Our model would predict similar results based on the length and costs of development, as well as the characteristics of the research team or company. Second, while our empirical analysis concerns the regulation of pharmaceuticals in the United States, there are dozens of national and federation-based regulators of drugs, medical devices, vaccines, and other health-related goods throughout the world that play roles similar to that of the FDA. In principle, at least, the same sort of data could be gathered for these regulators as for the FDA.

There remain features of approval regulation that are poorly captured by our model. First, we could incorporate a richer regulatory phase by allowing the regulator to audit, request additional testing, monitor the firm post-approval, or decide according to more realistic decision rules (such as the voting rules used by FDA review teams). Second, the development phase may better reflect

²⁴Somewhat stronger results are typically obtained if firm fixed effects are removed in all models.

empirical reality by incorporating mandated minimum pre-submission experiments (e.g., Phase I and Phase II clinical trials) or “fast track” procedures. Finally, the theory could incorporate actors such as a legislature, competing firms, or interest groups, who may differ in both their objectives and their abilities to detect and prevent errors. We underscore the difficulty of adding these to the present model while retaining analytic clarity, however. As it stands, the approval regulation framework provides a tractable foundation for studying bureaucratic decision making and error in an environment with endogenous information acquisition and submissions.

Appendix

Proof of Proposition 1. We derive the result by first considering histories where $e_1 = 1$, then histories where $e_1 = 0$. Finally, we check that all types experiment at $t = 1$.

Histories Beginning with $e_1 = 1$. There are two cases. First, if $k \leq \bar{x}_1$, then if $\eta(\theta, \emptyset) = 1 \forall \theta$, and $\sigma(\theta, 1) = 1 \forall \theta$ (i.e., both types experiment and submit if successful), then $\bar{x}(h_t) > k$ and R accepts the submissions. Since $E[E[x | h_3] | h_2 = 1] = E[x | h_2 = 1]$, F can do no better by conducting another costly experiment. R’s out of equilibrium beliefs about θ in the event that F continues experimenting are therefore inconsequential. Thus, $\sigma^*(\theta, 1) = 1 \forall \theta$ and $\rho^*(1) = 0$.

In the second case, $k > \bar{x}_1$. In this case, at $t = 2$ P would reject submissions if $\eta(\theta, \emptyset) = 1 \forall \theta$ and $\sigma^*(\theta, 1) = 1 \forall \theta$. Note, however, that if only type $\bar{\theta}$ submits, then $\bar{x}(h_t) > k$ and $\rho^*(h_t) = 0$. This induces both types to submit, which forces $\bar{x}(h_t) < k$ and creates an obvious contradiction. Thus, type $\bar{\theta}$ must submit with some interior probability when $h_2 = 1$. Since this implies indifference between submission and either continued experimentation or withdrawal, we have:

$$(1 - \rho^*(1)) \frac{m}{n+1} - c_s = v(\bar{\theta}, 1) \\ \Leftrightarrow \rho^*(1) = 1 - \frac{n+1}{m} (v(\bar{\theta}, 1) + c_s). \quad (15)$$

For R to choose an interior probability of rejection, it must be indifferent between acceptance and rejection. As a candidate for an equilibrium, suppose that $\eta^*(\theta, \emptyset) = 1 \forall \theta$ and $\sigma^*(\bar{\theta}, 1) = 1$. Then indifference by R at $t = 2$ implies:

$$k = \frac{p \frac{m}{n} \frac{m+1}{n+1} + (1-p) \frac{m-1}{n} \frac{m}{n+1} \sigma^*(\bar{\theta}, 1)}{p \frac{m}{n} + (1-p) \frac{m-1}{n} \sigma^*(\bar{\theta}, 1)},$$

$$\text{which yields } \sigma^*(\bar{\theta}, 1) = \frac{pm(m+1-k(n+1))}{(1-p)(m-1)(k(n+1)-m)}.$$

Now note that because $\sigma^*(\bar{\theta}, 1) = 1$, the history $h_3 = (1, 0)$ implies that $\mu((1, 0), S) = 0$. Because $k > \frac{m}{n+2}$, this implies $\rho^*(1, 0) = 1$ and $\omega^*(\theta, (1, 0)) = 1$ for all θ . Using (7) and (8), it is then straightforward to derive $v(\bar{\theta}, 1)$ if R continues experimentation:

$$v(\bar{\theta}, 1) = \frac{m}{n+1} \left(\frac{m+1}{n+2} - c_s \right) - c_e. \quad (16)$$

Since the right-hand side of (16) is strictly positive by (1) and (2), type $\bar{\theta}$ prefers submitting or experimenting at $h_2 = 1$ to withdrawing. Thus $\omega^*(\bar{\theta}, 1) = 0$ and $\eta^*(\bar{\theta}, 1) = 1 - \sigma^*(\bar{\theta}, 1)$.

Substituting (16) back into (15), we obtain:

$$\rho^*(1) = \frac{n-m+1}{n+2} + \frac{n+1}{m} c_e - \frac{n-m+1}{m} c_s. \quad (17)$$

We finally verify that type $\bar{\theta}$ in fact prefers submitting to experimenting when $h_2 = 1$:

$$(1 - \rho^*(1)) \frac{m+1}{n+1} - c_s > v(\bar{\theta}, 1) \\ \frac{(m+1)^2}{(n+1)(n+2)} - \frac{m+1}{m} c_e \\ + \left(\frac{1}{m} - \frac{m+1}{n+1} \right) c_s > \frac{m+1}{n+1} \left(\frac{m+2}{n+2} - c_s \right) - c_e \\ c_s - c_e > \frac{m(m+1)}{(n+1)(n+2)},$$

which is ensured by condition ES (12).

Histories Beginning with $e_1 = 0$. There are two cases. First, suppose that both types continue experimentation with positive probability in period 2. If $e_2 = 0$, then clearly $\omega^*(\theta, (0, 0)) = 1$. If $e_2 = 1$, then type $\bar{\theta}$ is acceptable to R (i.e., $\frac{m+1}{n+2} > k$), and type $\bar{\theta}$ is not (i.e., $\frac{m+1}{n+2} < k$).

Note that in equilibrium, type $\bar{\theta}$ must strictly prefer submission to withdrawal at $h_3 = (0, 1)$. Otherwise, $v(\bar{\theta}, (0, 1)) = 0$, but because $c_e > 0$, experimentation at $t = 2$ would imply that $v(\bar{\theta}, 0) < 0$. Thus period 3 submissions yield strictly positive payoffs for both types, and so $\sigma^*(\theta, (0, 1)) = 1$ for all θ . Further, type $\bar{\theta}$ can only be present at $t = 3$ if it weakly prefers experimentation to withdrawal at $t = 2$; hence $v(\bar{\theta}, 0) \geq 0$, or:

$$\frac{m-1}{n+1} \left(\frac{m}{n+2} (1 - \rho^*(0, 1)) - c_s \right) - c_e \geq 0. \quad (18)$$

This implies $\rho^*(0, 1) \leq 1 - \frac{n+2}{m} \left(\frac{n+1}{m-1} c_e + c_s \right)$, and also that type $\bar{\theta}$ experiments at $h_2 = 0$ (i.e., $\eta^*(\bar{\theta}, 0) = 1$ and $v(\bar{\theta}, 0) > 0$). To ensure that $\rho^*(0, 1) \geq 0$, EW (13) must not hold. There are two subcases for $\eta^*(\bar{\theta}, 0)$, depending on $\bar{x}(0, 1)$. If $\bar{x}_2 \geq k$, then $\rho^*(0, 1) = 0$ and $\eta^*(\bar{\theta}, 0) = 1$. Otherwise, at an interior $\rho^*(0, 1)$, $\bar{x}(0, 1) = k$ and thus

$\eta^*(\theta, 0)$ must satisfy:

$$k = \frac{p \left(\frac{n-m}{n} \right) \left(\frac{m}{n+1} \right) \frac{m+1}{n+2} + (1-p) \left(\frac{n-m+1}{n} \right) \left(\frac{m-1}{n+1} \right) \frac{m}{n+2} \eta^*(\theta, 0)}{p \left(\frac{n-m}{n} \right) \left(\frac{m}{n+1} \right) + (1-p) \left(\frac{n-m+1}{n} \right) \left(\frac{m-1}{n+1} \right) \eta^*(\theta, 0)},$$

which yields $\eta^*(\theta, 0) = \frac{p(n-m)m(m+1-k(n+2))}{(1-p)(n-m+1)(m-1)(k(n+2)-m)}$. To maintain this interior experimentation probability, let $\rho^*(0, 1) = \frac{n+2}{n} \left(\frac{n+1}{m-1} c_e + c_s \right)$. Finally, it is clear that $\omega^*(\theta, 0) = 1 - \eta^*(\theta, 0)$.

Second, suppose that only the high type experiments with positive probability in period 2. If EW (13) holds then (18) cannot hold for any $\rho(0, 1) \in (0, 1)$. In this case type θ must withdraw after a failure (i.e., $\omega^*(\theta, 0) = 1$). This implies that if $\eta^*(\theta, 0) > 0$, then at $h_3 = (0, 1)$, $\mu((0, 1), S) = \mu((0, 1), \emptyset) = 1$. Because $k < \frac{m+1}{n+2}$, R's response is to accept any submission: $\rho^*(0, 1) = 0$. Given $\rho^*(0, 1)$ and (1) and (2), type θ receives strictly positive expected utility from experimenting and submitting upon success, and thus we have $\eta^*(\theta, 0) = 1$ and $\sigma^*(\theta, (0, 1)) = 1$.

Finally, we verify that $\eta^*(\theta, \emptyset) = 1 \forall \theta$. It is sufficient to ensure that $\eta^*(\theta, \emptyset) = 1$, which will be the case if $\frac{m-1}{n} v(\theta, 1) - c_e \geq 0$, or equivalently $\frac{m(m-1)}{n+1} \times \left(\frac{m+1}{n+2} - c_s \right) - (n+m-1)c_e \geq 0$. It is easily checked that this is implied by (1) and (2). Thus ES or $k \leq \bar{x}$ is sufficient for existence of the ESE.

Supplementary Material

The following supplementary material is available for this article:

Regulatory Errors with Endogenous Agendas: Supplementary Appendix

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1540-5907.2007.00284.x> (This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

- Allison, G. T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown.
- Alt, J., R. Calvert, and B. Humes. 1988. "Reputation and Hegemonic Stability: A Game-Theoretical Analysis." *American Political Science Review* 82(1): 445–66.
- Avorn, J. 2004. *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. New York: Knopf.
- Bendor, J. 1985. *Parallel Systems: Redundancy in Government*. Berkeley: University of California Press.
- Bendor, J., and S. Kumar. 2005. "The Perfect Is the Enemy of the Best: Adaptive versus Optimal Organizational Reliability." *Journal of Theoretical Politics* 17(1): 5–39.
- Calvert, R. 1987. "Reputation and Legislative Leadership." *Public Choice* 55(1–2): 81–119.
- Carpenter, D. P. 2002. "Groups, the Media, Agency Waiting Costs and FDA Drug Approval." *American Journal of Political Science* 46(3): 490–505.
- Carpenter, D. P. 2004. "Protection without Capture: Dynamic Product Approval by a Politically Responsive, Learning Regulator." *American Political Science Review* 98(4): 613–31.
- Carpenter, D. P., M. Chernew, A. M. Fendrick, and D. Smith. 2003. "Approval Times for New Drugs: Does the Source of Funding for FDA Staff Matter?" *Health Affairs* (Web Exclusive), W3-618–624.
- Dixit, A., and R. Pindyck. 1994. *Investment Under Uncertainty*. Princeton, NJ: Princeton University Press.
- Dranove, D., and D. Meltzer. 1994. "Do Important Drugs Reach the Market Sooner?" *RAND Journal of Economics* 25(3): 402–23.
- Heimann, C. F. L. 1997. *Acceptable Risks: Politics, Policy, and Risky Technologies*. Ann Arbor: University of Michigan Press.
- Hilts, P. 2003. *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation*. New York: Knopf.
- Kamien, M. I., and N. L. Schwartz. 1972. "Timing of Innovations under Rivalry." *Econometrica* 40(1): 43–60.
- Landau, M. 1969. "Redundancy, Rationality, and the Problem of Duplication and Overlap." *Public Administration Review* 29(4): 346–58.
- Moscarini, G., and L. Smith. 2001. "The Optimal Level of Experimentation." *Econometrica* 69(6): 1629–44.
- Olson, M. 2000. "Regulatory Reform and Bureaucratic Responsiveness to Firms: The Impact of User Fees in the FDA." *Journal of Economics and Management Strategy* 9(3): 363–95.
- Peltzman, S. 1973. "An Evaluation of Consumer Protection Regulation: The 1962 Drug Amendments." *Journal of Political Economy* 81(5): 1049–91.
- Peltzman, S. 1976. *Regulation of Pharmaceutical Innovation: The 1962 Amendments*. Washington, DC: American Enterprise Institute.
- Quirk, P. J. 1980. "Food and Drug Administration." In *The Politics of Regulation*, ed. J. Q. Wilson. New York: Basic Books, pp. 191–235.
- Reinganum, J. F. 1982. "A Dynamic Game of R and D: Patent Protection and Competitive Behavior." *Econometrica* 50(3): 671–88.
- Simon, H. A. 1968. *Administrative Behavior*. New York: Free Press.
- Ting, M. M. 2003. "A Strategic Theory of Bureaucratic Redundancy." *American Journal of Political Science* 47(2): 274–92.
- United States General Accounting Office. 1990. "FDA Drug Review—Postapproval Risks 1976–1985." GAO/PEMD-90-15.
- Wardell, W. M., and L. Lasagna. 1975. *Regulation and Drug Development*. Washington, DC: American Enterprise Institute.