REGULATORY ERRORS UNDER TWO-SIDED UNCERTAINTY: OR, THE POLITICAL ECONOMY OF VIOXX[1]

Daniel Carpenter
Department of Government
Harvard University

Michael M. Ting
Department of Political Science and SIPA
Columbia University

January 19, 2005

[1]We thank James Fearon, John Ferejohn, Tom Hammond, Lewis Kornhauser, Alison Lawton, Ariel Pakes, Sonja Tong, Natasha Zharinova, seminar participants at the University of Michigan and NYU Law School, 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 on an earlier draft. 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. All errors, arguments and interpretations are ours.

Abstract

How do the errors of regulators – approving bad products, or rejecting good ones – depend upon the submission strategies and characteristics of submitting private entities or firms? We develop a model of approval regulation in which both firm and regulator are uncertain about the underlying quality of a product, but where the firm is better informed than the regulator. The model predicts that the commission of regulatory error depends heavily upon the induced submission strategies of firms, and in particular on the amount of time that firms take to experiment with their products before submitting them. Specifically, our analysis predicts that when experimentation is short and experimental costs are high, Type I errors (approving bad products) should be disproportionately associated with products submitted by firms with lower experimentation costs (larger firms), and Type II errors (rejecting good products) should be concentrated among products submitted by smaller firms. However, when experimentation is long and experimental costs are low, Type I errors should be concentrated among small firms, while Type II errors will be concentrated among larger firms. Under all conditions Type I errors should be increasing in the cost of regulatory submission, and the inverse should be true for Type II errors. We test hypotheses from the model in a statistical analysis of regulatory errors committed by the FDA in the approval of new pharmaceutical products. Using two different sets of measures for Type I errors, we find consistent support for the predictions of the model. We also find modest evidence that recent predictable reductions in approval times have been associated with a *reduced* rate of Type I error by the FDA, as our model would predict. The model also sheds new light on more (in)famous cases of regulatory error, including the FDA's delay in approving beta-blockers in the 1970s and the recent Vioxx episode.

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 ("commission") when she rejects a null (default) hypothesis when it is true. She commits a Type II error ("omission") when she affirms a null hypothesis that is in fact 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 has not qualified for benefits (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 an excellent project (II) while funding a mediocre one (I).[1]

Despite the salience of bureaucratic and regulatory error, there remain numerous questions for which students of bureaucracy and regulation lack general insight.[2]

- Why do some of the many decisions made by agencies result in error and others not? What circumstances of a case or features of bureaucratic procedure lead to error?
- What causes agencies to commit certain kinds errors rather than others (*i.e.*, Type I versus Type II)?
- Do errors depend in part upon the portfolio of cases that the agency receives? Does this portfolio depend in part upon the agency's own decision making and the way that private entities react to it? In other words, might the relationship between agency decision making and firm submissions be simultaneous, with important consequences for the determination of error?

The problems that occupy bureaucrats and regulators only rarely come to them exogenously. Instead, cases are actively brought to the agency 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 enforcement agency. Firms may submit license or new product applications. Researchers may submit patents or new research ideas embedded within grant applications. Our aim in this paper is to consider – theoretically via analysis of a game-theoretic model and empirically via analysis

[1] All of these examples presume agreement on the null hypothesis, and that the truth can be known *ex post*.

[2] Ours is far from the first treatment of these questions. Below we discuss two sets of theories – redundancy and bounded rationality – that have attempted to shed light upon these problems.

of U.S. Food and Drug Administration (FDA) drug approval decisions – the implications of the submission process and its endogeneity for the study of bureaucratic error.

The submission process is a crucial determinant of bureaucratic error because it shapes the "pool" of cases considered by an agency. The "selection" of cases considered by an agency can reduce the probability of error or can heighten it. Where for instance bureaucratic decision making leads the submission pool to be composed of cases that have a high Type I error probability, more errors of commission may result even without changes to administrative structure.

Our game begins with a product development phase, in each period of which a firm may submit an application at a cost, 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, then regulator the chooses whether to accept or reject the proposal. The firm also has private information about the quality of the product, and thus conditions its strategy on both it and its endogenously acquired experimental information.

Experiments thus 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. In our analysis, neither firm nor regulator knows the absolute truth. Uncertainty is two-sided – the firm does not fully know the quality or hazards of its product, but is better informed than the regulator – and both players learn from experimentation. Thus the firm itself needs 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 submit tend to submit earliest. Those with less favorable results may continue experimenting in the hopes of improving their record, or perhaps withdraw. 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 typically 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.

The model makes predictions about both types of error. Type I errors depend on the amount of experimentation conducted (which determines the amount of information available to players), as well as firm characteristics and regulatory procedures. When experimentation is short, the regulator will accept disproportionately more bad products from firms with low experimentation costs—these

are likely to be larger and older firms. However, the reverse is true when experimentation is long; then acceptance of bad products is more likely for submitters whose experimentation costs are high (smaller, newer firms). Additionally, Type I errors will rise with the cost of submission. The results for Type II errors are the reverse of those for Type I errors. Thus, when experimentation is short (respectively, long), the regulator will reject disproportionately more good products submitted from small (respectively, large) firms. Finally, Type II errors are decreasing in the cost of submission.

Additionally, a second class of Type II errors is generated through delay, as the inability of a firm to submit an acceptable product early can temporarily hurt consumers. Such "errors of delay" are not directly committed by the regulator but are nonetheless of important interest for policymakers. Our model predicts that errors of delay will be higher when the regulator has pessimistic prior beliefs about the quality of the proposal, as well as for small firms and high submission costs.

We assess these empirical predictions of the model in an analysis of FDA decisionmaking. 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 rather 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 decisionmaking and the incidence of several types of observable error. We examine FDA approval decisions for new molecular entities (NMEs) developed and submitted to the FDA 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. Section 3 derives the two equilibria of the game. Section 4 examines a number of testable predictions and implications for regulatory policy. Sections 5 and 6 present some data for and preliminary estimations of these predictions. Section 7 presents a number of case studies. Section 8 discusses the results and concludes.

1. 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 in many approval regulation settings. The most

publicized cases have occurred when a pharmaceutical entry regulator has approved bad products, such as when European regulators approved the sedative *Kevadon* (thalidomide). Thalidomide caused thousands of birth deformities in Europe, and the role of the FDA in keeping it from the U.S. market was widely celebrated as a regulatory success.[3] In recent years, critics of the FDA have focused upon its rejection or delay of valuable products, pointing to the costs of procedural conservatism in holding up everything from therapies for AIDS to beta-blockers for heart disease.

Second, there are typically players outside of the firm-regulator "game" who are also affected by the tradeoffs between Type I and Type II errors. 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 care directly about Type II errors (they might like drugs approved more quickly), while consumer safety groups may care directly about Type I errors (they might wish for drugs to be approved more slowly).

Previous studies of administrative and regulatory error have generally fallen under three categories. There exists an extensive decision-theoretic literature on optimal stopping problems, in which a single decision-maker trades off between accepting an experimental product immediately, and waiting for additional (and costly) information (*e.g.*, Kamien and Schwartz 1972, Reinganum 1982, Dixit and Pindyck 1994, Moscarini and Smith 2001). 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.

A second line of argument has focused upon administrative structure and the benefits and limits of redundancy (Landau 1969, Bendor 1985, Heimann 1997). These efforts focused on the effects of parallel processing (*e.g.*, adding extra regulators whose decisions are procedurally independent of existing agents' choices) and serial processing (adding extra steps before a decision must be made). The former structure can reduce Type II errors, but increases Type I errors, while the latter does the reverse. Overall, however, apparently redundant and seemingly inefficient administrative arrangements could be useful *ex ante* in minimizing organizational error, particularly for agencies governing risky technologies.[4]

A final set of theories that can be usefully applied to organizational error invoke bounded rationality. Because decision making imposes costs upon computation and the calculus of choice, boundedly rational bureaucrats may satisfice rather than optimize (Simon 1968). They may employ shortcuts in learning, fail to collect all of the information that an optimal decision would require,

[3]For an overview see Quirk (1980).

[4]More recently, Ting (2003) presents a strategic model in which the benefits of redundancy may be attenuated by intra-organizational shirking, as agents might free-ride off the efforts of others.

or rely heavily upon frames in rendering decisions. These properties of organizational learning (or failure to learn) can give rise to organization-wide maladies (*e.g.*, Allison 1971). Unfortunately, few formal models have used bounded rationality to study agency failure.[5] Further, little effort has been made thus far in distinguishing between agency error under fully rational as opposed to boundedly rational decision-making.

These literatures define the limitations of our analysis. That is, we do not make use of the statistical decision theory techniques found in optimal stopping theories, in part because of the complexity introduced by strategic behavior. We also do not examine variations in administrative structure in the manner of redundancy theory. Attention is restricted to variation *within* the same organization. Finally, the model does not use boundedly rational actors. As our model will show, it is possible even without these features for an organization to produce systematic variations in errors. The existing literature, however, does suggest natural extensions of our theoretical and empirical work.

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 in policy settings have taken the agenda of the policymaker as exogenous. This implies that the set of products (drugs, foods, energy technologies) or ideas (new military or law enforcement doctrines) considered by the policymaker is fixed and 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 their proposals or abandon them before decisions are made. 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.

2. The Model

2.1 Game Structure

Environment and Players. There are two players: a (F)irm and a (R)egulator. Both are imperfectly informed about a parameter x of a product, which may be thought of as the revenue expected from bringing the product to market. We assume that $x \sim \beta(\theta, n)$, where $\theta, n \in \mathbf{Z}_+$, and $1 < \theta < n$. The first parameter of the distribution, $\theta \in \{\underline{\theta}, \overline{\theta}\}$ where $\underline{\theta} = \overline{\theta} - 1$, is F's type, which is private information. We will refer to $\underline{\theta}$ and $\overline{\theta}$ as the "low" and "high" types, respectively. The second parameter, n, is common knowledge. Let $p \in (0, 1)$ represent R's prior belief of the high type ($\theta = \overline{\theta}$).

[5] A recent exception is Bendor and Kumar (2005), who analyze an "adaptive" model of redundancy in which pharmaceutical regulation is the motivating example.

The Beta distribution of firm priors has two attractive features for our purposes. First, it admits a natural interpretation as a set of n Bernoulli trials, of which θ resulted in success and $n-\theta$ in failure.[6] Given θ and n, the Beta distribution implies a prior mean θ/n and prior variance $\theta(n-\theta)/(n(n+1))$. The uncertainty over x can be resolved partially through observable experiments. Second, it is flexible enough to accommodate a wide variety of "shapes" of the density function, as determined by θ and n.

Sequence. The game has up to four periods, divided into a *development* phase with up to three periods and possibly a *regulatory* phase of one period. Periods are denoted by subscripts, and generic periods are denoted t. The phases are distinguished by the mover: only F moves in the development phase, and only R moves in the regulatory phase.

The game begins in the development phase. In it, 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 performed 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. We adopt the convention that $e_0 = 0$. If $f_t = E$, then the development phase continues. F cannot experiment past the second period and thus $f_3 \in \{S, W\}$.

At the beginning of the regulatory phase, R knows F's actions and experimental results (but not x or θ). Based on this, she makes a review decision $r \in \{A, R\}$, where A and R denote acceptance and rejection, respectively.

Information. Without experimentation, the model reduces to a simple signaling game (with F as the sender). With experimentation, both players update their expectations of x. The assumption of Beta-distributed priors makes the calculation of posterior beliefs very simple when both parameters of the distribution are known. For example, beginning with a prior of $\beta(m, n)$, two experiments producing e_1+e_2 successes generate a posterior on x that is distributed according to $\beta(\theta+e_1+e_2, n+2)$. Accordingly, $E[x \mid e_1, e_2] = \frac{\theta+e_1+e_2}{n+2}$ and $Var[x \mid e_1, e_2] = \frac{(\theta+e_1+e_2)(n+2-\theta-e_1-e_2)}{(n+2)(n+3)}$.

Since the parameters affecting $E[x]$ change with experimentation, it will be useful to distinguish between F's initial type (denoted by θ) and the numerical value of the first parameter of the Beta distribution as determined by experimentation. We therefore denote by $m = \overline{\theta}$ the numerical value of the high type's first parameter.

Utilities. F receives x if the product is approved, and zero for rejection. Each experiment costs

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

[6] See the two-period models of Calvert (1987) and Alt, Calvert and Humes (1988) for a similar technology in the context of a reputation game.

while submission costs

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

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

To rule out a number of trivial cases, we assume that k satisfies:

$$\frac{m+1}{n+2} > k > \frac{m}{n}. \tag{3}$$

The first part of (3) ensures that a high type may become acceptable to R after one round of experimentation, thus eliminating the dominant strategy of rejecting all "early" (*i.e.*, period 2) submissions. It also guarantees that the low type may become acceptable to R after two successful experiments, so that R's problem is not merely one of "separating" the two types. The second part states that k is higher than both types' *ex ante* expected product quality, so that neither will submit and be accepted in period 1. Thus some experimentation is necessary to generate a product satisfactory to the regulator, but the two players can disagree over the desirability of marginal products.

2.2 Equilibrium

We characterize Perfect Bayesian Equilibria (PBE) of this game that satisfy one additional refinement. Let H_t represent the set of experimental histories 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 has three elements.

1. F's strategy is a set $\{\phi_t\}_{t=1}^3$, where $\phi_t : \{\underline{\theta}, \overline{\theta}\} \times H_t \to \Delta(\{S, W, E\})$ for $t = 1, 2$, and $\phi_3 : \{\underline{\theta}, \overline{\theta}\} \times H_3 \to \Delta(\{S, W\})$ map types and experimental histories to probability distributions over submitting, withdrawing, and experimenting (where feasible).

2. R's strategy is the mapping $\rho : \bigcup_t H_t \to [0,1]$ from experimental histories, conditional upon a submission, into a probability of rejection.

3. R has beliefs $\mu : \bigcup_t H_t \times \{\emptyset, S\} \to [0,1]$ mapping the experimental and submission history into a probability that $\theta = \overline{\theta}$. These beliefs must be consistent with Bayes' Rule along the equilibrium path of play.[7]

[7] Along with h_t, this posterior distribution in turn induces a distribution over possible values of x.

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, $\eta(\theta, h_t)$. Letting $\eta(\theta, h_3) = 0$, it is clear that $\sigma(\theta, h_t)+\omega(\theta, h_t)+\eta(\theta, h_t) = 1$ for all t.

Second, denote the expected quality of a period-t submission (given beliefs $\mu(\cdot)$) by:

$$\bar{x}(h_t) = E[x \mid h_t, f_t = S] = \frac{\mu(h_t, S) + m - 1 + \sum_{i=1}^{t} e_{i-1}}{n + t - 1}. \tag{4}$$

In some cases, it will be possible for submissions to occur out of equilibrium. To complete the PBE, we assume that in these cases, R's beliefs about θ are given by $\mu(\emptyset, \emptyset)$. This is equivalent to R ignoring the experimental history of these submissions. By (3), this implies that $\bar{x}(h_t) < k$ for all such h_t, and thus R rejects the submission.[8] These beliefs are relatively innocuous. To see why, suppose instead that R's beliefs were "optimistic," in the sense that $\bar{x}(h_t) > k$. R's subsequent acceptance of all submissions would then induce all types to submit. It is easily shown that no equilibrium can be sustained in this manner.

As the subsequent development shows, the PBE concept does not always isolate a unique equilibrium. We therefore refine the set of equilibria with the following *submission equilibrium* criterion: where possible, we select the equilibrium in which submissions occur with positive probability at $t = 2$. Because no submissions can occur at $t = 1$, the refinement favors equilibria with "early" submissions; that is, it rules out equilibria in which R does not accept submissions at $t = 2$, and F therefore does not submit. This restriction is sensible in the context of our objectives, as the ruled-out equilibria have submissions only in the final period and therefore lack many of the interesting dynamics of the retained equilibria.

2.3 A Benchmark

A useful benchmark for the subsequent sections will be a model that is constrained to only one possible period of experimentation.

In equilibrium, R will accept a submission if and only if $\bar{x}(h_t) \geq k$. By assumption (3), no acceptances can take place unless F experiments successfully, and so it is clear that $\rho^*(\emptyset) = 1$, $\rho^*(0) = 1$ and $\sigma^*(\theta, \emptyset) = 0\ \forall\theta$.

There are two cases, which depend on whether the "average" quality of a submission with one successful experiment exceeds k. It will therefore be useful to define:

$$\tilde{x} \;=\; \frac{p\frac{m}{n}\frac{m+1}{n+1} + (1-p)\frac{m-1}{n}\frac{m}{n+1}}{p\frac{m}{n} + (1-p)\frac{m-1}{n}}$$

[8] *Any* "pessimistic" beliefs that induce $\bar{x}(h_t) < k$ would have the same effect. Thus, more complex beliefs that condition on h_t would support the same equilibrium behavior as that identified here.

$$= \frac{m}{n+1}\left(1+\frac{p}{m-1+p}\right) \tag{5}$$

as the expected period 2 quality conditional upon a single experimental success, given that both types experiment in period 1.

In the first case, $k \leq \tilde{x}$. This implies that if both types experiment at $t=1$, and all successful experimenters submitted at $t=2$, then $\bar{x}(1) > k$ and R must accept the submissions ($\rho^*(1)=0$). Under these conditions, F will clearly submit at $t=2$ ($\sigma^*(\theta,1)=1\ \forall\theta$) if it experiments successfully, and withdraw otherwise ($\omega^*(\theta,0)=1\ \forall\theta$). It is also easily verified that both types will experiment in period 1: $\eta^*(\theta,\emptyset)=1\ \forall\theta$.

In the second case, $k > \tilde{x}$. In this case, R would reject period 2 submissions if both types experimented and submitted if successful. Note however that if only high types submit, then $\bar{x}(1) > k$ and R's best response is $\rho^*(1)=0$. This induces all types to submit, which causes $\bar{x}(1) < k$ and creates an obvious contradiction.

Since there cannot exist an equilibrium in which $\rho^*(1)=0$, R must mix between accepting and rejecting submissions. This requires R to be indifferent between acceptance and rejection, and is possible if the low type does not experiment or submit with probability one. (The high type must experiment and submit if successful with probability one, since it receives strictly higher utility from a submission than the low type.) Hence,

$$p\frac{m}{n}\frac{m+1}{n+1}+(1-p)\frac{m-1}{n}\frac{m}{n+1}\eta^*(\underline{\theta},\emptyset)\sigma^*(\underline{\theta},1)=k \tag{6}$$

Note, however, that an interior value of $\sigma^*(\underline{\theta},1)$ would imply that the low type cannot receive a positive payoff at $t=2$, and thus does not experiment at $t=1$. Thus, $\sigma^*(\underline{\theta},1)=1$ and $\eta^*(\underline{\theta},\emptyset)=\frac{n(n+1)k-pm(m+1)}{(1-p)m(m-1)}$.

To determine $\rho^*(1)$, note that $\eta^*(\underline{\theta},\emptyset)$ implies that the low type firm is indifferent between experimentation and withdrawal. This implies that they are indifferent between submission and withdrawal, and thus:

$$\frac{m-1}{n}\left[(1-\rho^*(1))\frac{m}{n+1}-c_s\right]-c_e=0, \tag{7}$$

which yields $\rho^*(1)=1-\frac{n+1}{m}\left[\frac{n}{m-1}c_e+c_s\right]$. Note that this rejection probability is consistent with the claimed experimentation and submission strategies of the high type.

The example here has a number of noteworthy features that will be explored further in the following section. First, because acceptance with certainty will induce all types to submit, the expected quality of submitted and accepted products will often be exactly R's reservation value of k. R thereby benefits from its gatekeeping power, as $k > \frac{m}{n}$ by assumption. Despite the errors introduced by incomplete information, approval regulation in effect allows R to "skim" the best

products from the set of types, even when a product is *ex ante* unacceptable. Second, 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 information. This information, in addition to type, determines submission strategies. Finally, R makes both Type I (accepting bad products) and Type II (rejecting good products) errors at the regulatory phase. This contrasts with a decision-theoretic world in which a regulator knows product types and can experiment on her own.

3. Main Results

We now turn to the repeated-experimentation model. Two types of equilibria are possible, depending on how players react to success in a period 1 experiment. We begin by deriving each equilibrium, and then apply the submission equilibrium refinement (where necessary) to choose between them.

Throughout the game, each player's actions 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 otherwise. Thus:

$$\rho^*(h_t) = \begin{cases} 0 & \text{if } \bar{x}(h_t) > k \\ \in [0,1] & \text{if } \bar{x}(h_t) = k \\ 1 & \text{if } \bar{x}(h_t) < k. \end{cases} \tag{8}$$

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:

$$(1-\rho^*(h_t)) \left[\frac{\theta + \sum_{i=1}^{t} e_{i-1}}{n+t-1} \right] - c_s > v(\theta, h_t). \tag{9}$$

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

The derivation of equilibrium strategies is simplified greatly by the fact that the following are dominant strategies for R:

$$\rho^*(h_t) = \begin{cases} 1 & \text{if } h_t = \emptyset, 0, \text{ or } (0,0) \\ 0 & \text{if } h_t = (1,1). \end{cases} \tag{10}$$

That is, when the experimental record contains no successes, the expected product quality must be below k. Thus R would reject a submission with such a history, and so $\sigma^*(\theta, h_t) = 0$ for all such

h_t. At $t = 3$, the inability to continue experimentation therefore requires that $\omega^*(\theta, (0,0)) = 1$. Likewise, when h_t has two successes, both types have expected quality exceeding k, and would thus be accepted. Hence, $\sigma^*(\theta, (1,1)) = 1$.

This leaves three possible experimental histories to consider: $h_1 = 1$, $h_2 = (0,1)$, and $h_2 = (1,0)$. The following discussion restricts attention to these histories.

3.1 Early Submission Equilibrium

The first type of equilibrium is the *Early Submission Equilibrium* (ESE). In it, F submits "early" (*i.e.*, in period 2) with positive probability. Since $(m+1)/(n+1) > k > m/(n+1)$, R wishes to accept only "high" types in period 2.

As with the single-period benchmark in Section 2.3, there are two cases, which depend on whether P is willing to accept a product with expected quality implied single successful experiment. Recall that $\tilde{x} = \frac{m}{n+1}\left(1 + \frac{p}{m-1+p}\right)$ is the expected period 2 quality conditional upon a single experimental success, given that both types experiment in period 1. If $k \leq \tilde{x}$, then R is willing in expectation to accept the set of all successful first-period experimenters. Since additional experiments are costly and do not improve product quality in expectation, F submits if $h_2 = 1$.

When $k > \tilde{x}$, the existence of an ESE depends on another condition, which ensures that an initially successful high type prefers submission to continued experimentation. We label this the *Early Submission* (ES) condition:

$$c_s - c_e > \frac{m(m+1)}{(n+1)(n+2)}. \tag{11}$$

This expression is counter-intuitive because it suggests that F has less incentive to delay submission when submission costs are high and experimental costs are low. Under these conditions, it might be expected that F would have a greater incentive to wait for a (possibly) better product to submit. However, this incentive is reversed by the fact that the probability of acceptance is increasing in c_s and decreasing in c_e in equilibrium. This reflects the credibility of submissions when c_s is high (which encourages high quality submissions) and c_e low (which encourages information acquisition).

Equilibrium play also depends on c_s and c_e in another way. When both cost parameters are sufficiently high, then it is possible that upon a single failure, the low type will prefer withdrawal to continued experimentation. This is summarized by the following condition, which we label the *Early Withdrawal* (EW):

$$\frac{m-1}{n+1}\left(\frac{m}{n+2} - c_s\right) - c_e \leq 0. \tag{12}$$

When EW is true, R correctly believes that only a high type would remain in the game after a failure, and thus has the same estimate of x as F. Note that under assumptions (1) and (2), EW implies ES.

The first result derives the ESE strategies over the set of game histories that do not have dominant strategies.

Proposition 1 *(Early Submission Equilibrium) If $k \leq \tilde{x}$ or ES holds, then there exists a PBE in which equilibrium path strategies are as follows.*

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

For F (type $\underline{\theta}$): $\eta^*(\underline{\theta}, \emptyset) = 1$, $\sigma^*(\underline{\theta}, 1) = \begin{cases} 1 & \text{if } k \leq \tilde{x} \\ \frac{n(n+1)k - pm(m+1)}{(1-p)m(m-1)} & \text{if } k > \tilde{x}, \end{cases}$ $\eta^*(\underline{\theta}, 1) = 1 - \sigma^*(\underline{\theta}, 1)$,

$\eta^*(\underline{\theta}, 0) = \begin{cases} 0 & \text{if EW holds} \\ \frac{kn(n+1)(n+2) - p(n-m)m(m+1)}{(1-p)(n-m+1)(m-1)m} & \text{otherwise,} \end{cases}$ $\omega^*(\underline{\theta}, 0) = 1 - \eta^*(\underline{\theta}, 0)$, $\omega^*(\underline{\theta}, (1,0)) = 1$, $\sigma^*(\underline{\theta}, (0,1)) = 1$.

For R: $\rho^*(1) = \begin{cases} 0 & \text{if } k \leq \tilde{x} \\ \frac{n-m+1}{n+2} + \frac{n+1}{m} c_e - \frac{n+1-m}{m} c_s & \text{if } k > \tilde{x}, \end{cases}$ $\rho^*(1,0) = 1$,

$\rho^*(0,1) = \begin{cases} 0 & \text{if EW holds} \\ 1 - \frac{n+2}{m}\left(\frac{n+1}{m-1} c_e + c_s\right) & \text{otherwise.} \end{cases}$ ■

Proof All proofs are provided in the Appendix. ■

In most parameter configurations and histories, the average quality of F's submission is k, and thus R is made indifferent between rejection and acceptance. There are two exceptions. First, when the average quality following a single experimental success is high (*i.e.*, $k < \tilde{x}$), then submissions at $t = 2$ will have higher quality than k. The prior distribution of firms or technologies therefore affects the quality of early submissions. Second, when c_s and c_e are high so that EW holds, low types will withdraw from experimentation if $h_2 = 0$. Hence submissions made when $h_3 = (0,1)$ will be of higher quality than k.

It is useful to trace two sample game histories in the ESE. First, suppose that costs are not high (so that EW is violated), and consider the path of a high type firm whose first experiment fails. F cannot submit, but is sufficiently confident to conduct another experiment. With probability $m/(n+1)$, he succeeds and submits. But because the low type would also experiment with positive probability and submit if successful, R mixes between acceptance and rejection. Second, suppose that $k > \tilde{x}$, F is of the low type, and $e_1 = 1$. Because k is high, R does not accept all period 2 submissions and instead mixes. This makes F indifferent between submission and further experimentation. Because all high types would submit, R infers that F is of the low type if he experiments. Thus a further experimental success ($e_2 = 1$) results in submission and acceptance, while a failure results in withdrawal because $k > m/(n+2)$.

3.2 Late Submission Equilibrium

The second equilibrium type is the Late Submission Equilibrium (LSE). Here there are no submissions at $t = 2$, either because the expected quality of successful experimenters at $t = 2$ does not warrant acceptance ($k > \tilde{x}$). It is guaranteed to exist when the ESE does not, and additionally does not impose any requirements on c_s or c_e (as in ES). As with the ESE, however, the equilibrium path of play depends EW, with some low types withdrawing from play endogenously upon receiving negative experimental information.

Analogously to Proposition 1, the next result derives the LSE strategies over the set of histories without dominant strategies.

Proposition 2 *(Late Submission Equilibrium) If $k > \tilde{x}$, then there exists a PBE in which equilibrium path strategies are as follows.*

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

For F (type $\underline{\theta}$): $\eta^*(\underline{\theta}, \emptyset) = \eta^*(\underline{\theta}, 1) = 1$, $\eta^*(\underline{\theta}, 0) = \begin{cases} 0 & \text{if EW holds} \\ \frac{kn(n+1)(n+2) - p(n-m)m(m+1)}{(1-p)(n-m+1)(m-1)m} & \text{otherwise,} \end{cases}$

$\omega^*(\underline{\theta}, 0) = 1 - \eta^*(\underline{\theta}, 0)$, $\sigma^*(\underline{\theta}, (1,0)) = \frac{kn(n+1)(n+2) - pm(n-m)(m+1)}{(1-p)(m-1)(n-m+1)m}$, $\sigma^*(\underline{\theta}, (0,1)) = 1$.

For R: $\rho^*(1) = 1$, $\rho^*(1,0) = 1 - \frac{n+2}{m} c_s$, $\rho^*(0,1) = \begin{cases} 0 & \text{if EW holds} \\ 1 - \frac{n+2}{m}\left(\frac{n+1}{m-1} c_e + c_s\right) & \text{otherwise.} \end{cases}$ ■

As with the ESE, the average product quality of submitted products is usually k, and thus R is made indifferent between rejection and acceptance. The only exception occurs when submissions are made when $h_3 = (0,1)$ and EW holds, since low types will have withdrawn from experimentation. This results in submission quality higher than k and acceptance by R.

It will again be instructive to trace two example game histories in the LSE. First, consider the path of a high type firm whose first experiment succeeds. R does not accept submissions at $t = 2$, so F experiments again. He submits regardless of the outcome. If $e_2 = 1$, then R accepts. If $e_2 = 0$, then the F's submission is still acceptable to R, but the low type's submission has an expected quality below k. Thus, R mixes between acceptance and rejection. Second, suppose that F is of the low type, EW holds, and $e_1 = 0$. Even though R does would accept a submission if $e_2 = 1$, the costs of experimentation and submission, as well as F's reduced assessment of the product's quality, induce him to withdraw.

3.3 Equilibrium Selection

Propositions 1 and 2 establish the existence of equilibria under two sets of parametric conditions. The ESE requires that either $k \leq \tilde{x}$ or ES hold, while the LSE requires only that $k > \tilde{x}$. Thus when $k \leq \tilde{x}$, the model uniquely predicts the ESE. And when $k > \tilde{x}$ and ES does not hold, the model uniquely predicts the LSE. If $k > \tilde{x}$ and ES holds, then both equilibria exist. Here the submission

equilibrium refinement selects the ESE.[9]

Because they will be important in the subsequent analysis of regulatory error, the following table lists the submission equilibrium rejection probabilities as a function of the experimental history. We include only non-trivial histories, in which rejection and acceptance are not dominant strategies.

Table 1
Rejection Probabilities

h_t	$\rho^*(h_t)$
1	$\begin{cases} 0 & \text{if } k \leq \tilde{x} \\ \frac{n-m+1}{n+2} + \frac{n+1}{m}c_e - \frac{n+1-m}{m}c_s & \text{if } k > \tilde{x} \text{ and ES holds} \\ 1 & \text{otherwise (no submission).} \end{cases}$
1,0	$\begin{cases} 1 & \text{if } k \leq \tilde{x} \text{ or ES holds (no submission)} \\ 1 - \frac{n+2}{m}c_s & \text{otherwise.} \end{cases}$
0,1	$\begin{cases} 0 & \text{if EW holds} \\ 1 - \frac{n+2}{m}\left(\frac{n+1}{m-1}c_e + c_s\right) & \text{otherwise.} \end{cases}$

4. Regulatory Error

We now examine in detail the model's predictions about the likelihood of regulatory errors. Recall that the regulator commits a Type I error by approving a product with expected quality below k, or a Type II error by rejecting a product with expected quality above k. Both kinds of error may occur at any history of the game in which submissions are pooled, because R cannot distinguish between types and mixes between acceptance and rejection in these cases. We restrict attention to errors from the standpoint of the regulator's preferences (k) only, and thus cannot consider whether her "values" are optimal.[10]

4.1 Type I Errors

First consider the probability of equilibrium Type I errors for each history in which F submits with positive probability. Let $\psi^I(h_t)$ denote the equilibrium probability of a Type I error given h_t and a submission. The probability of a Type I error is then $\psi^I(h_t) = (1 - \rho^*(h_t))(1 - \mu(h_t, S))$.

[9]Because EW implies ES, the refinement therefore implies that the LSE strategies that depend on EW (see Proposition 2) are never used.

[10]Peltzman (1973, 1976), among others, considers the question of optimal regulatory standards.

Calculating $\mu(h_t, S)$ is facilitated by the fact that $\bar{x}(h_t) = k$ whenever the rejection probability is interior. In these cases, we may use (4) to obtain:

$$\mu(h_t, S) = k(n + t - 1) - m + 1 - \sum_{i=1}^{t} e_{i-1}. \tag{13}$$

When the rejection probability is not interior, $\mu(h_t, S)$ must be calculated directly from F's equilibrium strategies.

There are four histories for which submissions occur with positive probability. If either $h_3 = (1, 1)$, or $h_3 = (0, 1)$ and EW holds, there are no errors because only the high type submits. For the other histories, we may use Table 1 to calculate the error probabilities conditional upon submission for each. Table 2 summarizes the results.

Table 2
Type I Error Probabilities

h_t	$\psi^I(h_t)$
1	$\begin{cases} \frac{(1-p)(m-1)}{m+p-1} & \text{if } k \leq \tilde{x} \\ \left(\frac{m+1}{n+2} - \frac{n+1}{m}c_e + \frac{n+1-m}{m}c_s\right)(1 + m - k(n+1)) & \text{if } k > \tilde{x} \text{ and ES holds.} \end{cases}$
1, 0	$\frac{n+2}{m}c_s(1 + m - k(n+2))$
0, 1	$\begin{cases} 0 & \text{if EW holds} \\ \frac{n+2}{m}\left(\frac{n+1}{m-1}c_e + c_s\right)(1 + m - k(n+2)) & \text{otherwise.} \end{cases}$
1, 1	0

For many histories, Type I errors are increasing in c_s, because c_s can be interpreted as a one-time signal of quality. Likewise, c_e has a similar effect late in the development phase. In these cases, high costs convey product quality and thus increase acceptance probabilities, which in turn raise the possibility of errors from "bad" acceptances. This effect does not hold over the entire range of c_e and c_s, however, as values that violate EW cause initially unsuccessful low types to withdraw, which reduces the probability of Type I errors to zero.

For early submissions, the effect of c_e on $\psi^I(h_t)$ can be the opposite of that for late submissions. Since c_e reduces the incentive of low types to continue experimentation, high values can 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, it will be useful for testing purposes to report comparative statics based on the length of the pre-submission experimental

history. Let $\overline{\psi}^I_\tau(\cdot)$ denote the probability of a Type I error conditional upon submission after τ periods of experimentation. Aggregating the results from Table 2, it is straightforward to obtain the following comparative statistics:

- *Short experimentation* ($\tau = 1$): $\begin{cases} \frac{\partial \overline{\psi}^I_1}{\partial c_e} < 0, \frac{\partial \overline{\psi}^I_1}{\partial c_s} > 0 & \text{if ES holds and } p < \frac{(m-1)[(n+1)k-m]}{2m-(n+1)k} \\ \frac{\partial \overline{\psi}^I_1}{\partial c_e} = 0, \frac{\partial \overline{\psi}^I_1}{\partial c_s} = 0 & \text{otherwise.} \end{cases}$
- *Long experimentation* ($\tau = 2$):[11] $\begin{cases} \frac{\partial \overline{\psi}^I_2}{\partial c_e} > 0, \frac{\partial \overline{\psi}^I_2}{\partial c_s} > 0 & \text{if EW does not hold} \\ \frac{\partial \overline{\psi}^I_2}{\partial c_e} = 0, \frac{\partial \overline{\psi}^I_2}{\partial c_s} = 0 & \text{otherwise.} \end{cases}$

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

4.2 Type II Errors: Rejection

An analogous analysis can be performed on Type II errors from rejection decisions. Let $\psi^{II}(h_t)$ denote the equilibrium probability of a Type II error given h_t and a submission. The probability of a Type II error is then $\psi^{II}(h_t) = \rho^*(h_t)\mu(h_t, S)$.

Again, there are four histories for which submissions occur with positive probability. Note that $\psi^{II}(h_t) = 0$ for all cases in which all submissions are accepted. For the other histories, the error probabilities conditional upon submission for each history are easily calculated using (13) and Table 1. Table 3 summarizes the results.

[11] While "long" experimentation can imply any of three histories, calculation of these comparative statics is simplified by the observation that the history $(1, 0)$ cannot occur if ES holds. Since EW implies ES, $\psi^I(1, 0)$ does not enter the calculation of $\overline{\psi}^I_2$ when EW holds. And when EW does not hold, the effect of $\psi^I(1, 0)$ is in the same direction as that of $\psi^I(0, 1)$.

Table 3 **Type II Error Probabilities**	
h_t	$\psi^I(h_t)$
1	$\begin{cases} 0 & \text{if } k \leq \tilde{x} \\ \left(\frac{n-m+1}{n+2} + \frac{n+1}{m}c_e - \frac{n+1-m}{m}c_s\right)(k(n+1)-m) & \text{if } k > \tilde{x} \text{ and ES holds.} \end{cases}$
$1,0$	$(1-\frac{n+2}{m}c_s)(k(n+2)-m)$
$0,1$	$\begin{cases} 0 & \text{if EW holds} \\ \left[1-\frac{n+2}{m}\left(\frac{n+1}{m-1}c_e + c_s\right)\right](k(n+2)-m) & \text{otherwise.} \end{cases}$
$1,1$	0

Perhaps unsurprisingly, the results in Table 3 are symmetric to those in Table 2. The same factors that increase acceptance rates (and Type I errors) will reduce Type II errors. Thus, no value of c_e or c_s is dominated in the sense of increasing both types of error.

We may carry out the analogous exercise to that in Section 4.1 to derive comparative statics based on the length of experimentation. Let $\overline{\psi}_\tau^{II}(\cdot)$ denote the probability of a Type II error conditional upon submission after τ experiments.

- *Short experimentation* ($\tau = 1$): $\begin{cases} \frac{\partial \overline{\psi}_1^{II}}{\partial c_e} > 0, \frac{\partial \overline{\psi}_1^{II}}{\partial c_s} < 0 & \text{if ES holds and } p < \frac{(m-1)[(n+1)k-m]}{2m-(n+1)k} \\ \frac{\partial \overline{\psi}_1^{II}}{\partial c_e} = 0, \frac{\partial \overline{\psi}_1^{II}}{\partial c_s} = 0 & \text{otherwise.} \end{cases}$
- *Long experimentation* ($\tau = 2$): $\begin{cases} \frac{\partial \overline{\psi}_2^{II}}{\partial c_e} < 0, \frac{\partial \overline{\psi}_2^{II}}{\partial c_s} < 0 & \text{if EW does not hold} \\ \frac{\partial \overline{\psi}_2^{II}}{\partial c_e} = 0, \frac{\partial \overline{\psi}_2^{II}}{\partial c_s} = 0 & \text{otherwise.} \end{cases}$

Note finally that for all histories, $\frac{\partial \psi^{II}}{\partial k} > 0$ and $\frac{\partial \psi^{II}}{\partial c_s} \leq 0$, and so at the margin raising the quality standard or lowering submission hurdles increases Type II errors. Accordingly, our analysis suggest that foresightful disease advocates, by placing more weight on Type II errors, would support reducing quality standards and raising submission costs.

4.3 Type II Errors: Delay

In addition to rejecting "good" products, Type II errors can also occur if the submission of a product that R would accept is delayed. Delay-induced errors have played a large role in the recent history of pharmaceutical regulation. 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 the time required for approval. In 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.[12]

Our model provides a simple way of uncovering submission delays. After one experiment, only the high type's product is acceptable to R. In the ESE, these products are always submitted, while in the LSE they never are. Thus in the LSE, an unwanted submission delay occurs with probability pm/n.

The LSE requires that $k > \tilde{x}$, which implies a low value of p. Thus, as intuition would suggest, low *ex ante* confidence in the product inhibits early acceptances. The LSE also requires that ES not hold (*i.e.*, $c_e \geq c_s - \frac{m(m+1)}{(n+1)(n+2)}$). High experimental costs make early submissions impossible because they raise the low type's incentive to submit immediately, and thus reduce the quality of possible period 2 submissions. Low submission costs also make early submissions impossible because they reduce the credibility of submissions.[13]

It is worth noting that a final class of Type II errors occur if F stops development of a potentially acceptable product. In the equilibria of our model, this does not occur because the only withdrawals occur when EW holds and the low type withdraws after one failure. Under the parametric assumptions of our model, such products will never be acceptable from R's perspective. 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 because of an initial experimental failure.

5. Data

Our data include 32,216 molecular entities that have developed to some stage of product maturity over the past thirty years. We have collected these data from a variety of sources, including Freedom of Information Act requests to the FDA, FDA Annual Reports from its drug reviewing divisions, F-D-C (Food, Drug and Cosmetic) Reports, and the trade database Pharmaprojects^{TM}. Where our data concern yet-to-be submitted products, we rely more heavily upon trade reporters and databases. Where our data concern drugs submitted to the FDA and then either rejected or approved and launched, our data rely much more heavily upon FDA sources, checked against the data of Dranove and Meltzer (1994) and Pharmaprojects^{TM}.

[12]See Section 112 of the FDA Modernization Act of 1997 (21 USC 351), and the Accelerated Approval Rule (21 CFR 314, 601).

[13]Note that because of the equilibrium selection argument of Section 3.3, we are being conservative in predicting such errors.

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 percent, or 2,789, were eventually submitted to the FDA. The others have been abandoned or are currently in a limbo R&D status which (following Pharmaprojects^{TM}) we signify as "no development reported."

Our analyses here sample only approved drugs, and it is worth remarking that a richer estimator would consider the quasi-selection effect by which some drugs get submitted and others get abandoned. We are currently working on this problem.

Measuring Type I Errors. We confine our focus here to Type I errors, or "bad" approvals. 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 has probably 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 General Accounting Office (GAO) in 1990, when it produced a report on "postapproval risks" of FDA-approved drugs. For a sample of 198 molecular entities, the GAO asked 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 of Type I errors is placed elsewhere, we adopt straightforwardly the coding of the GAO in their report, and code two dependent variables.

- *GAOLIST* is an indicator scored 1 if the drug was listed by the GAO as having been subject to "significant labeling revisions," and scored 0 otherwise.

- *GAOLINES* is the number of lines of text describing significant labeling revisions in the original GAO report. *GAOLINES* is scored 0 for all drugs for which $GAOLIST = 0$.

Our use of *GAOLINES* 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 a set of new molecular entities approved from 1976 to 1985, and then only on those that the GAO selected for study (198 drugs). We are missing firm data on two of the drugs that the GAO studied, 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 (*WITHDRAW1*). 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 (*NOENTRY5*), 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. 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 the log of this variable *LNFIRMINDS*. 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.

Another plausible source of variation in c_e is by disease. Diseases that have a higher hospitalization burden are, on net, more costly per patient in clinical trials, because the trial will often have to pay the cost of hospitalization for treatment and placebo arms. The treatments will also be associated with delivery of care in a higher-cost setting. 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.

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 costly 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*.

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). There are two interpretations of PDUFA's relationship to submission cost. Most patently, PDUFA raised explicit submission costs by requiring all sponsors to pay for each new NDA.

We argue instead that a better way to think of PDUFA is in the reverse. 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 (CDER). 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. Since the fees are quite small (ranging from $240,000 to about $1m) relative to the capitalized benefits of earlier approval (Carpenter *et al* 2003, Carpenter 2004), the implicit acceleration of benefits probably far outweigh the explicit user fee costs, and so the onset PDUFA should be conceived as having reduced c_s.

Controls. For all analyses, we control for firm fixed effects. Where relevant, we also control for a brute time trend by including the year in which the NDA was submitted (*SUBYEAR*).

The other 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 (HCUP) data, the average number of days of hospitalization per hospitalization. We call this variable *HHOSLENG*. This variable is zero whenever there are no recorded hospitalizations for a disease in the HCUP data. This variable covaries positively with the mortality of a disease and with the expected cost of treatment for those individuals who present with the disease. This variable was also shown in Carpenter (2002) to be negatively and significantly associated with FDA approval times for new molecular entities. For now, we attach no theoretical meaning to this variable, except to say that it may covary positively with experimentation cost, insofar as clinical trials for more expensive and high-morbidity diseases are probably more costly per-trial and per-patient.

6. 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 to capture essential features of the data.

Summary descriptive statistics for a few of our variables appear in Table 4. The first half of the table details the GAO data. Notice that by this coding, exactly one-half of in-sample new molecular entities approved from 1976 to 1985 were subject to "serious" postapproval labeling revisions. When we count the number of lines of serious labeling revision described by the GAO, the average is 5.6 lines per drug but runs to 48 lines for the maximally revised labeling.

Table 4
Summary Statistics

Variable	Mean	SD	N
GAOLIST	0.50	0.50	196
GAOLINES	5.60	8.80	196
LNFIRMINDS	1.91	0.96	196
HHOSLENG	2.84	3.84	196
WITHDRAW1	0.061	0.241	1071
NOENTRY5	0.029	0.168	1071
LNFIRMINDS	2.86	1.53	1071
HHOSLENG	2.14	3.77	1071

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 percent of new molecular entities approved by the FDA from 1984 to the present were approved but then withdrawn in one or more highly developed markets (*WITHDRAW1*). The fraction of entities approved by the FDA but that failed to enter five or more highly developed foreign markets is 2.9 percent for this period (*NOENTRY5*). 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.

6.1 GAO Data

We begin with analysis of the GAO data. The results from maximum likelihood regressions of the binary dependent variable *GAOLIST* upon relevant regressors, including a heteroscedasticity correction, are presented in Table 5. In Table 6, we regress *GAOLINES* on the same variables using negative binomial regression. By the results of Section 4.1, we have two predictions. First, for drugs that are submitted relatively early and by low-c_e firms, we predict that Type I error is decreasing in c_e across the such subsamples. From these results, there is support for this hypothesis. Looking at columns (1) and (2) of both tables, the coefficient on *LNFIRMINDS* is generally positive and statistically significant. However, in some cases this variable attains statistical significance only at the $p < 0.10$ level for a two-tailed test.[14]

Our second prediction is that, for products characterized by longer experimentation times and submitted by larger firms, Type I errors should be increasing in c_e. By the results of Section 4.1, long experimentation and low c_e should be associated with a *reversal* of the comparative static on c_e. In Columns (3)-(5) of Table 5 and Columns (3)-(5) of Table 6, we present estimates

[14]This is due in part to the inclusion of fixed effects, which raise the standard error associated with *LNFIRMINDS*.

derived from truncated samples in which analysis is restricted to larger firms and drugs with longer development times. We do this in order to construct an artificial sample similarity with empirical regimes where lower experimental cost and longer experimentation are observed. The probit results show a sign change on the coefficient for *LNFIRMINDS*, which now has a negative effect. In the probit regressions, the coefficient on *LNFIRMINDS* is positive and statistically significant in the high-experimentation-cost, short-development samples (columns (1) and (2)), and negative in the low-experimentation-cost, long-development samples. However, the negative coefficient estimates of columns (4) and (5) of Table 5 are statistically differentiable from zero only at the $\alpha = 0.10$ level. Similar results are obtained in Table 6, where the coefficient estimate on *FIRMINDS* is positive for the high-experimentation-cost, short-development sample (though only at the $\alpha = 0.10$ level), but negative and statistically significant for the low-experimentation-cost, long-development subsample.[15]

Although we do not report them here, these results are generally replicated if we use generalized linear regressions with heteroscedasticity corrections for these variables. In particular, regressing *LNGAOLINES* (where $LNGAOLINES = \ln(1 + GAOLINES)$) upon *LNFIRMINDS* yields results that are substantively identical to those of the maximum likelihood models.

Table 5
Probit Analyses of GAO-Coded Error

Dependent Variable: *GAOLIST*

	(1)	(2)	(3)	(4)	(5)
LNFIRMINDS	2.267	0.797	−0.409	−0.854	−0.792
	(0.900)	(0.394)	(0.351)	(0.530)	(0.551)
HHOSLENG	0.522	0.048	−0.017	−0.042	−0.036
	0.189	(0.055)	(0.046)	(0.089)	(0.095)
SUBYEAR	−0.286	−0.126	0.041	0.049	0.045
	0.123	(0.068)	(0.023)	(0.031)	(0.021)
Constant	561.8	247.5	−79.99	−93.61	−85.88
	(244.0)	(134.9)	(45.46)	(58.60)	(39.94)
N	31	52	57	28	25
LNFIRMINDS	< 2	< 3	> 2	> 2	> 2
INDTIME (mos.)	< 120	< 120	NR	> 96	> 120

[15]One difference between these tables is that, for the negative binomial regressions with *GAOLINES* as the dependent variable, the variable *FIRMINDS* seems to produce a better fit than the variable *LNFIRMINDS*, hence we use the former for these regressions.

Table 6
Negative Binomial Regression Analyses of GAO-Coded Error

Dependent Variable: *GAOLINES*

	(1)	(2)	(3)	(4)	(5)
FIRMINDS	0.352	0.114	−0.257	−0.059	−0.047
	(0.208)	(0.081)	(0.138)	(0.023)	(0.017)
HHOSLENG	−0.058	−0.025	−0.016	−0.023	0.001
	(0.068)	(0.048)	(0.046)	(0.079)	(0.086)
SUBYEAR	−0.004	0.149	−0.030	0.075	0.073
	(0.007)	(0.066)	(0.010)	(0.024)	(0.025)
Constant	9.222	0.227	−56.44	−145.23	−142.23
	(12.90)	(20.89)	(18.82)	(47.95)	(48.57)
N	46	73	66	50	44
LNFIRMINDS	< 2	< 3	> 2	> 2	> 2
INDTIME (mos.)	< 120	< 120	> 72	> 96	> 120

Note: Maximum likelihood estimations for samples with *LNFIRMINDS* > 3 ($N = 14$) fail to converge.

6.2 International Data

We now turn to the sample of drugs approved over the past twenty-five years ($N = 1071$), and employ our measures of Type I error as coded via international comparison. 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, *PDUFADUM*. Following the analysis of Section 4.1, 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. Hence anytime that a firm has never had a foreign withdrawal in our dataset, or anytime that a firm has never been denied 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.[16] The estimates should, therefore, be interpreted with caution.

Table 7 shows the results of a probit analysis examining the "full sample" for effects of secular changes in the cost of submission. Notice that, once the full sample is examined without theoretically relevant conditions and restrictions, simple linear inclusion of the variable *LNFIRMINDS* does not add to the model's predictive power and does not yield coefficient estimates statistically differentiable from zero. In specifications (1) and (3), submission cost is measured as the (differenced)

[16]The dependent variables range from 0.03 to 0.06, depending on the model estimated.

series in median approval time, calculated for the submission year of the molecule in question as $\Delta E(APPTIME)$. In toto, this represents firms' generally unconditioned expectations about the time- or delay-cost of their submission. In specification (1) this variable is positively related to Type I error rate as measured by *WITHDRAW1*, but the association is statistically significant only at the $p < 0.10$ level for a two-tailed test. In specifications (2) and (4), we instrument for $\Delta E(APPTIME)$ with the PDUFA dummy variable (scored 1 for products submitted after 1993). The idea here is that PDUFA represents a shift in expected approval times that was anticipable by submitting firms. Here we find more robust associations between the instrumented change in median approval time and Type I error rate. Notice that all four models include a "time trend," which is the year in which the molecule in question commenced development.

Rather strikingly, the estimates suggest that reductions in approval time are associated with less Type I error, not more. 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). It is important, though, to note that our data offer no measure of severity. Hence we cannot make statements here about the desirability or aggregate welfare effects of the 1992 PDUFA legislation.

Table 7
Probit Analyses of Internationally-Coded Error Tests for the Effects of PDUFA

	Dependent Variable: *WITHDRAW1*		Dependent Variable: *NOENTRY5*	
	(1)	(2)	(3)	(4)
LNFIRMINDS	0.195	0.177	0.360	0.251
	(0.200)	(0.192)	(0.270)	(0.255)
HHOSLENG	−0.070	−0.068	−0.201	−0.233
	(0.030)	(0.030)	(0.069)	(0.085)
$\Delta E(APPTIME)$	0.035	–	0.012	–
	(0.018)		(0.025)	
$\Delta E(APPTIME)^*$	–	0.513	–	0.722
(* instrumented by PDUFA)		(0.161)		(0.265)
Constant	92.06	67.53	−2.12	23.41
	(51.74)	(53.41)	(0.22)	(69.05)
N	533	535	410	412
Time Trend?	Yes	Yes	Yes	Yes

Finally, Table 8 presents probit and GLS results for the dependent variable *WITHDRAW1*. Here the results are quite interesting for the *LNFIRMINDS* variable. When the sample is variously

restricted to drugs produced by smaller firms and excludes drugs with very long IND times, the coefficient estimates are positive and statistically significant for specifications (2) and (3). However, when the sample is restricted to large firms (specifications (4)-(7)), we observe negative and significant coefficients. In many of these subsamples collinearity induces the dropping of the *PDUFADUM* variable, but when it is included it is never statistically differentiable from zero.

Table 8
Probit and GLS Analyses of Internationally-Coded Error

Dependent Variable: *WITHDRAW1*

	(1) Probit	(2) Probit	(3) Probit	(4) GLS	(5) Probit	(6) GLS	(7) Probit
LNFIRMINDS	0.811	0.647	0.770	−0.030	−0.380	−0.043	−0.360
	(0.578)	(0.332)	(0.386)	(0.017)	(0.168)	(0.022)	(0.173)
HHOSLENG	0.002	–	–	−0.005	−0.082	−0.006	−0.082
	(0.041)			(0.004)	(0.042)	(0.003)	(0.047)
PDUFADUM	0.043	–	0.421	−0.016	–	−0.027	–
	(0.387)		(0.487)	(0.036)		(0.039)	
Constant	−2.603	−2.548	−3.372	0.092	0.242	0.226	−0.122
	(0.551)	(0.456)	(0.813)	(0.051)	(0.730)	(1.086)	(0.814)
N	213	24	90	383	234	262	215
LNFIRMINDS	< 1.5	< 2	< 3	> 2	> 3	> 3	> 3
INDTIME (mos.)	NR	< 360	< 360	> 96	> 72	> 96	> 96

6.3 Summary and Caveats

We find appreciable support for our primary hypotheses that (1) where experimentation cost is high and development times are shorter, reductions in experimentation cost (c_e) (increases in firm experience) are accompanied by increases in Type I regulatory error; and (2) where experimentation cost is low and development is longer, increases in firm experience are accompanied by declines in Type I regulatory error. This inversion of statistical associations occurs across two datasets, and in each of two measures of Type I error in each dataset.[17] The results (especially Tables 6 and 8) lend solid support to the predicted reversal of the comparative static on c_e for the subsample of low-c_e, long-IND times submissions.

We can test for effects of submission cost only brutely, and only in our second set of analyses on more recent data. There, the coefficient estimates for $\Delta E(APPTIME)$ in Table 7 broadly support the hypothesis that by reducing the cost of submission, PDUFA lowered the probability of Type I

[17]Somewhat stronger results are typically obtained if firm fixed effects are removed in all models.

errors. The coefficient for expected approval time is positive in all estimations and significant in all but one, and across a variety of specifications. If it can be safely interpreted that a known effect of PDUFA was to reduce implicit submission costs for sponsors, then we have evidence that Type I error rates are decreasing in c_s. 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 technologies for detecting drug hazards *ex ante* is one plausible candidate.

We do not attempt to attach any theoretical meaning to the estimated effect of *HHOSLENG*. However, it should be noted that *HHOSLENG* is only weakly associated with GAO-coded errors, but negatively and robustly associated with error as coded via international comparisons.

Finally, we note a problem with our second measure of Type I error as coded through international comparison. It is possible that *NOENTRY5* is 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 *NOENTRY5* is affected by firm-level variables that are not responsive to anticipation of regulatory strategy (portfolio factors in the company's pipeline), then it is possible that our inferences are subject to error from confounding variables.

7. Case Studies of Regulatory Error: Vioxx and Beta-Blockers

Any quantitative analysis rests upon suspect assumptions about sample weights. Is an error always of the same magnitude? Strictly speaking, our regressions reported in Tables 5-8 would suppose so. While we are comfortable with the assumption made in analyses of *GAOLINES* that more elaborate labeling revision is a proxy for greater regulatory error in approval, even here the relative weight of one regulatory error vis-a-vis another is probably difficult to measure reliably with data based on labeling revisions.

Type II errors are in general harder to measure than are Type I error, because all but a few drugs approved in overseas markets (and not withdrawn) eventually do get approved in the United States, so must equate delay with *de facto* rejection (see Carpenter 2004). It is for this reason, and to further develop and refine intuition, that we briefly examine some case narratives of Type I and Type II errors, in order to round out our empirical account.[18]

Does the theoretical framework elaborated here help to illuminate more classic cases of regulatory error by approval regulators? We consider a classic case of Type II error (delayed availability),

[18] A more elaborate and rigorous empirical examination of Type II errors lies outside the parameters of this paper. We are currently working on data collection for this problem, and believe that the interpretive issues presented by Type II error simply demand a separate and sustained analysis.

and a more recent and highly publicized case of what may have been a Type I error, the quick approval (and failure of the FDA to induce withdrawal of) Vioxx (rofecoxib).

7.1 A Type II Error: The 1970s Delay of Beta Blockers

Libertarian 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 (Wardell and Lasagna 1975). By blocking beta-adrenergic receptors to which adrenaline binds in heart cells, beta blockers reduce effective cardiovascular "workload." The delayed introduction of propanolol and practolol as treatments for essential hypertension and angina was estimated by Wardell to have resulted in 10,000 to 20,000 deaths per year in the United States. Other commentators put these costs lower and argue that the FDA is not primarily to blame (Hilts 2003: 191-92). We step aside from the political debate on this question and assume, for purposes of this paper, that the delayed entry of propanolol and other beta-blockers into the U.S. market represents a Type II error. Just how catastrophic the delay was, just who was to blame, and whether the severity of this error casts doubt upon the desirability of FDA regulation generally – these are issues we leave to other researchers.[19]

We seek to understand the potential contribution of an approval regulation framework to the following three questions. First, why generally were proplanolol and other beta-blockers available in the United Kingdom years before they were available for such conditions in the United States? Other beta blockers with greater cardioselectivity – practolol and oxprolenol – were also available in the UK five to ten years before their introduction into U.S. markets. Second, much more specifically, why was propanolol available relatively early for arrhythmias and much later for hypertension? The patent for propanolol hydrochloride was granted in 1962 and it entered the U.K. market in 1965. It was approved by the FDA in November of 1967, but according to Wardell and Lasagna (1975: 61) it was not until 1973 that the FDA approved its use for hypertension and angina.[20] Third, why did the FDA essentially reject oxprolenol for the better part of five years (its eventual approval consumed 54 months)?[21]

The nature of these delays merits reflection. In most cases, the aggregate delay was in part due to delays in FDA approval (a *de facto* rejection) as well as delayed submission to the FDA (the rational failure to submit a good product). Although a complex constellation of scientific factors

[19]While certainly plausible, Wardell and Lasagna's claims were made without a clear statistical methodology, and we are aware of no replication to date which has confirmed or disconfirmed them. Hilts' account is also quite threadbare in its particulars. The statistical, scientific and policy history of beta-blockers represents a promising agenda for careful research.

[20]Wardell and Lasagna do not address the important question of whether propanolol was available as a treatment for hypertension from the inception of its entry into the British market, or whether it followed the U.S. path of being approval for hypertension only later.

[21]Data on patent dates and U.S. approval taken from Carpenter FDA approval database, maintained at the Department of Government, Harvard University.

was at play, it nonetheless appears that the following three factors played a role in these delays.

Low Credibility. For both Type II errors committed by the regulator and for "slow submission" Type II errors, society's *ex ante* belief in product quality plays an important role. Low expectations for product value reduce the likelihood of an early submission equilibrium, and in the 1960s and early 1970s, many FDA officials viewed pharmaceutical science with considerable suspicion and distrust. FDA Commissioner James Goddard delivered a powerful broadside to the PMA in 1966 saying that he was "shocked at the quality of many submissions to our IND staff," that he was "shocked at the materials that come into us," and "shocked at clear attempts to slip something by us." Goddard's successor, Commissioner Charles Edwards, frankly lamented to the National Academy of Sciences a in 1970 that the FDA was "receiving massive submissions of valueless data." For firms large and small, old and new, FDA officials were generally seen as reluctant to place faith in companies' therapeutic claims. The reasons for this distrust include both a history of fraudulent and suspect product applications in the 1950s and 1960s, and the growing reliance of FDA officials upon rigorous principles of research pharmacology.[22]

Experimentation Costs: High for Early Beta-Blockers, Low for Later Beta-Blockers. Consider first the delayed introduction or propranolol for hypertension. Once submitted for arrhythmias in 1966, propranolol received a relatively quick review (17 months, or one year less than the 29-month average for all NMEs submitted in 1966). The costs of experimentation were relatively high for propranolol and other early beta-blockers because so little was known about this class of drugs and one of the main firms developing these therapies – Imperial Chemical Industries (ICI) – was relatively small by global pharmaceutical standards of the time. An important question, then, is why it took so long for ICI and other sponsors to submit their drug applications to the FDA. Hilts (2003: 192), for instance, reports that ICI "simply hadn't done the scientific studies necessary to get the drug approved" for hypertension and angina.

Although historical data are scarce to answer this question, it is also quite possible that the *per-trial* experimentation cost for any hypertension drug was larger than that for an arrhythmia drug during this time.[23] If true, this differential would help to explain why, for a uncertain regulator

[22]Goddard remarks from Hilts (2003: 168). Consider also the tirade of FDA general counsel William Goodrich against the advertising practices of the "big eight" pharmaceutical companies; Mintz (1967: 92L). Edwards remark from "Remarks to be presented at a meeting with the Drug Research Board of the NAS/NRC, February 20, 1970," Charles Edwards Papers, Box 2 ("Speeches, FDA, 1970-1971"); Mandeville Special Collections Library, University of California, San Diego.

[23]The historical context of arrhythmia treatment and hypertension treatment in the 1960s and 1970s renders this a very difficult question to address. Using figures from hospitalization costs, the treatment of arrhythmia can be either more or less costly than the treatment for hypertension, depending entirely upon which complications accompany the disease. Yet more recent figures would include the costs of implantable cardiac defibrillators (ICDs), which currently run about 30,000 U.S. dollars per quality-adjusted life-year (see ARHQ). Other surgery-based treatments – including pacemaker implantation or catheter radioablation of accessory nerve pathways – are also quite costly. However, purely

such as the FDA, propanolol for arrhythmia was easy to accept but propranolol for hypertension was not.

FDA regulators would not accept an "early submission" for propranolol for hypertension, then, because of the decreased credibility of such a submission. In the approval regulation model, higher experimental costs can prevent the existence of an early submission equilibrium because they raise the "low type's" incentive to submit immediately, and thus reduce the expected quality of period 2 ("early") submissions.

The experimentation costs for later beta-blockers were lower for two reasons: more was known about them, and they were developed by much larger firms. With the approval of propranolol in 1967, subsequent beta-blockers were seen as "me-too" drugs for which much less scientific progress had been made relative to the original. Hence the testing of second-generation beta-blockers was cheaper per-patient than was the testing of propranolol.

Consider, then, the case of oxprolenol. It was patented in 1965, but not until 14 years later (in 1979) was it submitted to the FDA for treatment of essential hypertension. This fact has two implications for our purposes. First, much more was known about beta-blockers in the late 1970s, and so the experimentation cost (per-patient and per-trial) had been considerably reduced from the late 1960s. Second, the amount of experimentation conducted on oxprolenol before its submission was 14 years, which in our model corresponds to the long experimentation category. In this category we predict that Type II errors are a non-increasing function of experimentation cost, or a non-decreasing function of firm size. Ciba-Geigy was, at the time of oxprolenol's approval, a European juggernaut with 14.7B (Swiss Francs) in annual sales. In the 1990s, before its eventual merger into Novartis, it employed over 80,000 people worldwide. And in 1980, as the FDA was considering Oxprolenol, four of Ciba's new molecular entities had just been approved since 1976.[24] Finally, it merits note that the FDA's Bureau of Drugs had assigned a priority rating of "1C," or "standard" to oxprolenol (two more generous ratings, 1A and 1B, were available at the time). These ratings may have been a reflection of the low strategic credibility of the drug, or perhaps a low *ex ante* belief in oxprolenol's value may have inhibited an early acceptance.

Both early and late, then, beta-blockers may have been adversely affected by doubts about the

pharmacological treatment of arrhythmias in the 1970s was undoubtedly cheaper than contemporary alternatives that entail hospitalization. Our guess is based on the fact that hypertension trials would likely have been larger because of essential hypertension's greater prevalence; from National Health Interview Survey and Vital Statistics data, the prevalence of essential hypertension would appear to be about four times that of (aggregated) cardiac arrhythmias over the past twenty years. All of these discussions presume the validity of treatment cost measures as proxies for clinical trial costs. For oncologic settings, this appears to be a reasonable assumption (Goldman, et al (2003)), but we do not have comparable figures for arrhythmia and hypertension.

[24]These were Cibalith (Lithium citrate), Transderm-Scop (Scopolamine), Lioresal (Baclofen), and Ludiomil (Maprotilene HCL).

strategic credibility of the applications, in at least two ways that are illuminated by our model. Beta-blockers made their entrance into a regulatory climate shot through with doubts about the credibility of claims made and studies undertaken by pharmaceutical companies. In our analysis, this lack of general credibility prevented early submission equilibria that might have resulted in quicker availability in the United States. Later on, beta-blockers such as oxprolenol were associated with much lower experimentation costs that lowered the strategic credibility of a product application, particularly for large firms such as Ciba-Geigy that could easily surmount these hurdles.

7.2 Vioxx

On September 30, 2004, Merck announced the voluntary withdrawal of Vioxx (rofecoxib), a COX-2 inhibitor for the treatment of osteoarthritis and other indications, from the worldwide market. 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. Soon after the withdrawal, other evidence from clinical trials and epidemiological studies surfaced suggesting that Vioxx was associated with a trebling or more of the risk of adverse cardiovascular events, and perhaps with tens of thousands of deaths due to its wide use. The FDA soon published a 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 alone. This memorandum and the reception it received at the agency have been the subject of considerable public controversy. Whatever the controversy, it now appears that the FDA quite likely erred in giving Vioxx too quick an approval or in being too late to induce its withdrawal from the marketplace, or both.

For purposes of this discussion, we assume that the initial approval of Vioxx was itself a regulatory Type I error. In saying this, it is important to keep in mind that the FDA has been heavily criticized less for approving Vioxx in the first place, but much more acting too slowly to induce its removal. Still, given that adverse cardiovascular outcomes were associated with Vioxx after 18 months of use, it is possible that the FDA relied on insufficient clinical trial evidence in approving the drug – that is, clinical trials whose endpoints were observed only six months, 12 months and 18 months after administration of the treatment. The interpretation of regulatory error in this case, while we leave the medical and epidemiological details to others, seems defensible in this case. What does our theoretical framework say about it? In fact, FDA review documents and recent congressional investigations of the Vioxx episode reveal that FDA reviewers saw evidence of increased hypertension and thromboembolic events with Vioxx, suggesting that more studies should perhaps

have been required.[25]

Rofecoxib was patented on January 10, 1994 and its clinical development and testing was quite rapid. It was submitted as Vioxx to the FDA on November 23, 1998, was granted priority review status, and was approved on May 20, 1999. From the standpoint of approval regulation, the most important thing to know about Vioxx is credibility and size of its sponsor: Merck. Merck is one the largest and most experienced pharmaceutical firms in the world, and it has long enjoyed a good reputation at the FDA. Although our data are impressionary and not scientifically collected, interviews with officials at the FDA and within the pharmaceutical industry support this point of view. When asked an open-ended question about which firm enjoys the best reputational relationship, pharmaceutical industry officials volunteer the name "Merck" more than any other company by far. Because of Merck's reputation and the trust that scientists and regulators placed in its research and development, Vioxx came to the FDA with a high degree of credibility. In terms of our model, this generated a higher *ex ante* belief in the quality of Vioxx (a high p), and hence for an enhanced likelihood of early-submission equilibrium. The priority review status given to Vioxx for a non-lethal disease such as osteoarthritis would support this interpretation.

The other noteworthy feature of Vioxx's development was the relatively brief period of experimentation it enjoyed. The drug was submitted for NDA approval just four years after its patent had been granted. Clearly, then, this was an "early submission." Under short experimentation in the ESE of our model, Type I error is decreasing in c_e, or increasing in firm size. For short experimentation phases, high costs are no longer a credible signal of quality to the regulator. Not only is Merck one of the most reputationally established firms in the pharmaceutical industry, it is also one of the most experienced and largest. Our model predicts that with short experimentation, Type I errors will be concentrated among larger firms.

8. Discussion

A model based upon endogenous submissions to a regulator produces several counterintuitive predictions that would not be generated by existing approaches to agency failure, either structural or cognitive. The model suggests a number of empirical tests, and on this front our preliminary results appear promising.

There remain some unsatisfying features of this framework. First, the fact that much of the significant "action" in our model occurs at the development phase follows from the highly simplified

[25]See M.I. Villalba, "Medical Officer Review: VIOXX (Rofecoxib)" NDA 21-042/21-052, November 19, 1999; Division of Anti-Inflammatory, Analgesic and Ophthalmic Drug Products – HFD-550, U.S. Food and Drug Administration; pp. 86, 90, 100-106. See also A. W. Mathews, "Did the FDA Staff Minimize Vioxx's Red Flags?" *Wall Street Journal*, November 10, 2004, p. B1.

regulatory review phase. Accordingly, the development of a dynamic regulatory phase would enrich the theoretical predictions regarding regulatory error. This may take several forms, including an audit, requests for additional testing, or *ex post* reviews of previously approved products. Second, the development phase itself may better reflect empirical reality, by incorporating a mandated minimum duration of experimentation before submission (*e.g.*, Phase I and Phase II clinical trials, which may usefully serve as a screening mechanism), or "fast track" procedures. We conjecture that such a benchmark would serve to limit Type I errors in the aggregate, while promoting "system" Type II errors from delayed submission.

As it stands, however, the model provides a tractable foundation for studying regulatory decisionmaking and failure in an environment with endogenous firm product development (R&D). We conclude by noting some of the substantively important extensions that the model can be expanded to address. These include:

Structural Model. Ideally we would like to augment our theoretical model with an appropriate "structural" model, but the parsimony of the former (for instance, its restriction to two periods maximum of experimentation) makes this impossible. We are currently exploring ways to simulate a more general model (with endogenous periods of experimentation) and then use the simulations to augment empirical estimation.

Redundancy. Our model posits a single regulator, whereas in contemporary settings many actors (or offices) review the product submissions and decide upon them according to rather complex and obscure voting rules. It would therefore be fruitful to consider the effects of organizational redundancy (*e.g.*, Ting 2003) on not only administrative incentives, but also on the induced pattern of firm submissions.

Bounded Rationality or Adaptation. A boundedly rational regulator and/or firm might be subject to confirmatory bias, memory constraints (the full experimental history might not be cleanly recalled by an agency with high turnover rates), or attention constraints (other products competing for reviewers' energies). Alternatively, a regulator could act to satisfy certain political or scientific aspirations instead of maximizing utilities.

Political Costs for Errors. While purely scientific considerations might yield one form of tradeoff among Type I and Type II errors, political and economic considerations might alter the regulator's preferences and alter the tradeoff. Along these lines, an important variable might be the *ex post* observability of each type of error, which could be modeled as a function of the political strength or visibility of certain lobbies.

APPENDIX

Proof of Proposition 1. We derive the result in three steps. First, we consider histories where $e_1 = 1$, then histories where $e_1 = 0$. Finally, we establish parametric conditions on existence of the equilibrium.

Histories Beginning with $e_1 = 1$. There are two cases. First, if $k \leq \tilde{x}$, 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 $E[E[x \mid h_3] \mid h_2 = 1] = E[x \mid h_2 = 1]$, R can do no better by conducting another costly experiment.[26] Thus, $\sigma^*(\theta, 1) = 1 \ \forall\theta$ and $\rho^*(1) = 0$.

In the second case, $k > \tilde{x}$. 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 $\overline{\theta}$ 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 $\underline{\theta}$ must submit with some interior probability when $h_2 = 1$. Since this implies indifference between submission and continued experimentation, we have:

$$(1-\rho^*(1))\frac{m}{n+1} - c_s = v(\underline{\theta}, 1), \tag{14}$$

which after manipulation yields the equilibrium rejection probability:

$$\rho^*(1) = 1 - \frac{n+1}{m}\left(v(\underline{\theta}, 1) + c_s\right). \tag{15}$$

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 $\sigma^*(\overline{\theta}, 1) = 1$ and $\eta^*(\theta, \emptyset) = 1 \ \forall\theta$. Then indifference by R at $t = 2$ implies:

$$p\frac{m}{n}\frac{m+1}{n+1} + (1-p)\frac{m-1}{n}\frac{m}{n+1}\sigma^*(\underline{\theta}, 1) = k. \tag{16}$$

and thus:

$$\sigma^*(\underline{\theta}, 1) = \frac{n(n+1)k - pm(m+1)}{(1-p)m(m-1)}. \tag{17}$$

Now note that because $\sigma^*(\overline{\theta}, 1) = 1$, 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 $\omega^*(\underline{\theta}, (1, 0)) = 1$. It is then straightforward to derive $v(\underline{\theta}, 1)$:

$$v(\underline{\theta}, 1) = \frac{m}{n+1}\left(\frac{m+1}{n+2} - c_s\right) - c_e. \tag{18}$$

[26] R's out of equilibrium beliefs about F's type in the event that it continues experimenting are therefore inconsequential.

Substituting back into (15), we obtain:

$$\rho^*(1) = \frac{n-m+1}{n+2} + \frac{n+1}{m}c_e - \frac{n-m+1}{m}c_s. \tag{19}$$

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

Note that in this equilibrium, type $\underline{\theta}$ cannot be indifferent between submission and withdrawal. Because $c_e > 0$, experimentation at $t = 2$ would then imply that $v(\underline{\theta},0) < 0$. Thus type $\underline{\theta}$ can only be present at $t = 3$ if it is indifferent between experimentation and withdrawal at $t = 2$; hence:

$$\frac{m-1}{n+1}\left(\frac{m}{n+2}(1-\rho^*(0,1)) - c_s\right) - c_e = 0, \tag{20}$$

which implies $v(\underline{\theta},0) = 0$ and $\rho^*(0,1) = 1 - \frac{n+2}{m}\left(\frac{n+1}{m-1}c_e + c_s\right)$. This probability is interior if $\frac{m-1}{n+1}\left(\frac{m}{n+2} - c_s\right) - c_e > 0$, and is also sufficient to ensure that type $\overline{\theta}$ experiments at $t = 2$ when $h_2 = 0$ (*i.e.*, $\eta^*(\overline{\theta},0) = 1$ and $v(\overline{\theta},0) \geq 0$). At this value of $\rho^*(0,1)$, period 3 submissions yield strictly positive payoffs for both types, and so $\sigma^*(\theta,(0,1)) = 1$ for all θ. This information is then sufficient to characterize $\eta^*(\underline{\theta},0)$:

$$p\left(\frac{n-m}{n}\right)\left(\frac{m}{n+1}\right)\frac{m+1}{n+2} + (1-p)\left(\frac{n-m+1}{n}\right)\left(\frac{m-1}{n+1}\right)\frac{m}{n+2}\eta^*(\underline{\theta},0) = k. \tag{21}$$

Simplifying, we obtain:

$$\eta^*(\underline{\theta},0) = \frac{kn(n+1)(n+2) - p(n-m)m(m+1)}{(1-p)(n-m+1)(m-1)m}. \tag{22}$$

Second, if $\frac{m-1}{n+1}\left(\frac{m}{n+2} - c_s\right) - c_e \leq 0$ (*i.e.*, EW, or (12)), then (20) cannot hold for any $\rho(0,1) \in (0,1)$. In this case type $\underline{\theta}$ must withdraw after a failure (*i.e.*, $\omega^*(\underline{\theta},0) = 1$). This implies that $v(\underline{\theta},0) = 0$ and $v(\overline{\theta},0) \geq 0$, which ensures that if $h_3 = (0,1)$, then $\mu((0,1),S) = \mu((0,1),\emptyset) = 1$. Because $k < \frac{m+1}{n+2}$, $\rho^*(0,1) = 0$ and by (1) and (2), $\sigma^*(\overline{\theta},(0,1)) = 1$.

To complete the derivation, we check whether players follow the prescribed strategies. For all conditions it is sufficient to check that the strategies are followed when $k > \tilde{x}$.

First, at $t = 2$ type $\overline{\theta}$ prefers submitting when $h_2 = 1$ to experimenting if:

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

which is condition ES (*i.e.*, (11)). Thus existence requires that c_s be sufficiently higher than c_e.

Second, at $t = 2$ type $\underline{\theta}$ prefers submitting or experimenting when $h_2 = 1$ to withdrawing if:

$$\begin{aligned}(1-\rho^*(1))\frac{m}{n+1} - c_s &> 0\\ \frac{m(m+1)}{(n+1)(n+2)} - c_e - \frac{m}{n+1}c_s &> 0\\ \frac{m}{n+1}\left(\frac{m+1}{n+2} - c_s\right) - c_e &> 0. \end{aligned} \tag{23}$$

Thus existence requires that c_s and c_e both be sufficiently low.

Finally, we verify that $\eta^*(\theta, \emptyset) = 1\ \forall\theta$. For this purpose it is sufficient to ensure that $\eta^*(\underline{\theta}, \emptyset) = 1$:

$$\begin{aligned}\frac{m-1}{n}v(\underline{\theta}, 1) + \frac{n-m+1}{n}v(\underline{\theta}, 0) - c_e &> 0\\ \frac{m-1}{n}\left[\frac{m}{n+1}\left(\frac{m+1}{n+2} - c_s\right) - c_e\right] - c_e &> 0\\ \frac{m(m-1)}{n+1}\left(\frac{m+1}{n+2} - c_s\right) - (n+m-1)c_e &> 0. \end{aligned} \tag{24}$$

It is easily checked that (24) is satisfied by (1) and (2). It implies (23), and so along with ES is sufficient for existence of the equilibrium. ■

Proof of Proposition 2. We derive the result in three steps. First, we consider histories where $e_1 = 1$, then histories where $e_1 = 0$. Finally, we establish parametric conditions on existence of the equilibrium.

Histories Beginning with $e_1 = 1$. Suppose $h_2 = 1$. Since $\rho^*(1, 1) = 0$, the experimental continuation value for type θ is:

$$v(\theta, 1) = \frac{\theta+1}{n+1}\left(\frac{\theta+2}{n+2} - c_s\right) + \frac{n-\theta-1}{n+1}\max\left\{0, (1-\rho^*(1,0))\frac{\theta+1}{n+2} - c_s\right\} - c_e. \tag{25}$$

With a history of $h_2 = (1, 0)$, R will wish to accept only type $\overline{\theta}$. There are three subcases, of which only the first is possible under the assumptions of the model. In it, R is indifferent between acceptance and rejection when faced with h_2, and type $\underline{\theta}$ is indifferent between submission and withdrawal at $t = 3$ (as with the ESE):

$$(1-\rho^*(1,0))\frac{m}{n+2} - c_s = 0, \tag{26}$$

Thus, $\rho^*(1, 0) = 1 - \frac{(n+2)c_s}{m}$. To calculate the submission probability for type $\underline{\theta}$, we determine the strategy that ensures an average quality of k and leaves R indifferent between rejection and acceptance:

$$p\left(\frac{m}{n}\right)\left(\frac{n-m}{n+1}\right)\frac{m+1}{n+2} + (1-p)\left(\frac{m-1}{n}\right)\left(\frac{n-m+1}{n+1}\right)\frac{m}{n+2}\sigma^*(\underline{\theta}, (1,0)) = k. \tag{27}$$

Solving, we obtain $\sigma^*(\underline{\theta},(1,0)) = \frac{kn(n+1)(n+2)-pm(n-m)(m+1)}{(1-p)(m-1)(n-m+1)m}$, and hence $\omega^*(\underline{\theta},(1,0)) = 1-\sigma^*(\underline{\theta},(1,0))$. Clearly, if type $\underline{\theta}$ is indifferent between submission and withdrawal, type $\overline{\theta}$ must strictly prefer submission, and thus $\sigma^*(\overline{\theta},(1,0)) = 1$.

This case requires that $\eta(\theta,1) = 1\ \forall\theta$. To verify this, type $\underline{\theta}$ experiments at $t=2$ if:

$$\begin{aligned} v(\underline{\theta},1) = \frac{m}{n+1}\left(\frac{m+1}{n+2} - c_s\right) + \frac{n+1-m}{n+1}(0) - c_e &\geq 0 \\ \frac{m}{n+1}c_s + c_e &\leq \frac{m(m+1)}{(n+1)(n+2)}. \end{aligned} \tag{28}$$

This condition is also sufficient for type $\overline{\theta}$ to experiment at $t=2$ (*i.e.*, $v(\overline{\theta},1) = \frac{(m+1)(m+2)}{(n+1)(n+2)} + \frac{n-m(m+2)}{m(n+1)}c_s - c_e \geq 0$).

In the second subcase, R is also indifferent between acceptance and rejection when $h_2 = (1,0)$. But this indifference occurs because (28) cannot hold and type $\underline{\theta}$ is indifferent between experimentation and withdrawal at $t=2$. Thus,

$$\begin{aligned} v(\underline{\theta},1) &= \frac{m}{n+1}\left(\frac{m+1}{n+2} - c_s\right) + \frac{n+1-m}{n+1}\left(\frac{m}{n+2} - c_s\right)(1-\rho^*(1,0)) - c_e \\ &= 0. \end{aligned} \tag{29}$$

Solving, this implies: $\rho^*(1,0) = 1 - \frac{(n+1)(n+2)c_e - m(m+1-(n+2)c_s)}{(n+1-m)(m-(n+2)c_s)}$. But by (1) and (2), $\rho^*(1,0) > 1$, and thus the subcase is ruled out by assumption.

In the last subcase, $v(\underline{\theta},1) < 0$, so low types withdraw even after an initial success ($\omega^*(\underline{\theta},1) = 1$) and $\rho^*(1,0) = 0$. This implies that neither (28) nor (29) can hold; *i.e.*, $\frac{m}{n+1} < c_e + c_s$. This subcase is also ruled out by (1) and (2).

Histories Beginning with $e_1 = 0$. In the case where $h_2 = 0$, the equilibrium is identical to that of the ESE (Proposition 1).

Finally, we check for conditions that ensure existence of this equilibrium. First, to verify that $\sigma^*(\theta,h_2) = 0\ \forall\theta,h_2$, note that R's out of equilibrium beliefs require that $\mu(h_2,S) = p$, and because $k > \tilde{x}$, $\rho^*(h_2) = 1\ \forall h_2$.

Second, to verify that $\eta^*(\theta,\emptyset) = 1\ \forall\theta$, it is sufficient to show that $\eta^*(\underline{\theta},\emptyset) = 1$. This will be true if:

$$\begin{aligned} v(\underline{\theta},\emptyset) = \frac{m-1}{n}v(\underline{\theta},1) + \frac{n-m+1}{n}v(\underline{\theta},0) - c_e &\geq 0 \\ \frac{m-1}{n}\left[\frac{m}{n+1}\left(\frac{m+1}{n+2} - c_s\right) - c_e\right] - c_e &\geq 0 \\ \frac{(n+1)(n+m-1)}{m(m-1)}c_e + c_s &\leq \frac{m+1}{n+2}. \end{aligned} \tag{30}$$

This condition is identical to (24) and is assumed by (1) and (2). It furthermore implies (28). Thus, it is sufficient for existence of the LSE. ■

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: 445-466.

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, CA: 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: 5-39.

Calvert, R. 1987. "Reputation and Legislative Leadership." *Public Choice* 55: 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): 613-631.

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: Princeton University Press.

Dranove, D., and D. Meltzer. 1994. "Do Important Drugs Reach the Market Sooner?" *RAND Journal of Economics* 25: 402-423.

Goldman, D. et al. 2003. "Incremental Treatment Costs in National Cancer Institute-sponsored Clinical Trials," *Journal of the American Medical Association* 289 (22): 2970-7.

Heimann, C. F. L. 1997. *Acceptable Risks: Politics, Policy, and Risky Technologies.* Ann Arbor: University of Michigan Press.

Hilts, Philip. 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: 43-60.

Landau, M. 1969. "Redundancy, Rationality, and the Problem of Duplication and Overlap." *Public Administration Review* 29(4): 346-358.

Mintz, Morton. 1967. *By Prescription Only.* New York: Brown, Little.

Moscarini, G., and L. Smith. 2001. "The Optimal Level of Experimentation." *Econometrica* 69(6): 1629-1644.

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): 363-395.

Olson, M. 2002. "Pharmaceutical Policy Change and the Safety of New Drugs," *Journal of Law and Economics*; 45(2) Part 2: 615-642.

Peltzman, S. 1973. "An Evaluation of Consumer Protection Regulation: The 1962 Drug Amendments." *Journal of Political Economy* 81(5): 1049-1091.

Peltzman, S. 1976. *Regulation of Pharmaceutical Innovation: The 1962 Amendments.* Washington, D.C.: American Enterprise Institute.

Quirk, P. J. 1980. "Food and Drug Administration," in J. Q. Wilson (ed.), *The Politics of Regulation.* New York: Basic Books.

Reinganum, J. F. 1982. "A Dynamic Game of R and D: Patent Protection and Competitive Behavior." *Econometrica* 50(3): 671-688.

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): 274-292.

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, D.C.: The American Enterprise Institute.