The basic structure of pharmaceutical regulation differs in several crucial respects from the governance of other goods and other economies. The most salient contextual distinction concerns the pervasive uncertainty that governs pharmaceuticals – neither producers nor consumers nor regulators can be said to know with plausible certainty the quality or hazards of a drug product. The fundamental dynamic is far more than asymmetric information; it is also that no participant in the game of pharmaceutical regulation can be said to know the true quality of any drug. The most salient institutional distinction rests in gatekeeping: the necessity of governmental pre-market review for new products, where any approval is based in part upon an experimental (non-market) history of the product. This gatekeeping feature of pharmaceutical regulation is adopted by nation states and regional federations around the globe, and this fact gestures less to the pure optimality of these arrangement than to the symbolic centrality of the American example (Carpenter 2010, Chapter 11). In the United States, the gatekeeping role is played by the U.S. Food and Drug Administration (FDA), a storied organization that exercises significant administrative discretion and has deeply shaped the laws and regulations by which it has been given regulatory authority (Marks 1997; Carpenter 2001, 2010; Carpenter and Sin 2007).
The emergence of gatekeeping power at the FDA is one of the most compelling narratives in American public law over the past century, and for students of public choice, gatekeeping power fascinates because it fundamentally shapes and reorders the incentives, concepts and procedures of the global pharmaceutical market. These incentives, concepts and procedures govern scientific researchers, drug and medical products companies, physicians and other prescribers of medicine, and patients and consumers. The FDA’s position as sentry – at the gate where ideas and molecules become commodities – induces highly strategic and particular investment behavior on the part of pharmaceutical companies, and this position reinforces the agency’s power over how researchers and companies define and use scientific and medical concepts. Pre-market review authority implies veto power not unlike that studied in institutional political science (Cameron 2000) – the ability to shape the “agenda” of drugs submitted to the agency (and hence the issues and controversies that will come before the regulator), as well as the power to shape human thought and learning.

I devote this essay to introducing readers in public law and public choice to central principles governing pharmaceutical regulation, including its historical development and some theoretical precepts and models that can serve as a partial guide for its understanding. I open with a critical survey of the state of current research in this field, pointing especially to the lack of uncertainty and dynamic considerations in theories and models constructed for the field, and the lack of nuance and complexity in historical studies of the subject. I then briefly allude to several critical developments in the history of pharmaceutical regulation, and then turn to discuss two modeling frameworks in which the operation of pharmaceutical institutions in the United States and other regions may be better understood.
At the core of this endeavor are three concepts and related models – reputation, information and confidence. The regulatory power of the FDA stems from its reputation for scientific expertise and consumer protection. So too, the spread of American arrangements throughout the world (democratic and non-democratic, industrialized and developing) is a product of the agency’s international reputation. Information is arguably the central product of American pharmaceutical regulation; one of the most important effects of regulatory institutions is to generate greater information (and higher-quality data) about therapies than would be generated in the absence of regulation. Confidence, finally, composes the core dynamic of pharmaceutical regulation and its institutions – the agency must have confidence in the products it approves and in their commercial sponsors, the public (including patients and their prescribers, as well as those who pay for drugs) depends upon the confidence induced by regulation for its well-being and for its optimal utilization of therapies, and the agency depends upon public and congressional confidence for its operation and its power.

Modern Research in Pharmaceutical Markets and Regulation – A Lack of Rigor and Realism

Despite their centrality in law, government and economics, and despite the controversy they occasion in modern political debate, arrangements for the regulation of pharmaceuticals have been poorly and simplistically analyzed in academic studies.¹ The poverty of these analyses is most apparent in modern economics and in industrial organization, where

¹ Important exceptions occur in the work of legal scholar Richard Merrill (1997) and historian Harry Marks (1997).
theoretical models as well as quantitative studies analyze a blunt tradeoff between “safety” and “access.” On one side, the “benefits” of pharmaceutical regulation are claimed to come in keeping unsafe drugs off of the market, whereas the “costs” of regulation come from impeding access to new medicines (Peltzman 1973, Temin 1980; Berndt, Gottschalk, Philipson and Strobeck 2005; Philipson and Sun 2008). Using the language of statistical decision theory, the benefits come in the form of reduced “Type I errors” (accepting or marketing a drug that is in fact unsafe beyond its therapeutic value) whereas the costs of regulation come in “Type II errors” (rejecting a drug whose benefits outweigh its safety). The regulator or the FDA in these models is assumed (without supporting evidence or theory other than anecdote or bland references to the thalidomide crisis of 1962) to favor caution over access, or to avoid Type I errors more than Type II errors.

Despite the lack of theory and evidence supporting it, this thinking informs a wide variety of academic writings on the subject of pharmaceutical regulation (Peltzman 1973, Temin 1980; Berndt, Gottschalk, Philipson and Strobeck 2005; Philipson and Sun 2008), and it informs

---

2 See Carpenter and Ting (2007) for theoretical definitions of these errors, as well as an equilibrium model containing simple dynamics and stochastic uncertainty, as well as statistical analysis of correlates of regulatory error.

3 The lack of rigor in these analyses is most apparent in Philipson and Sun (2008), who argue for a national pre-emption standard (essentially the invalidity of state-level lawsuits against pharmaceutical manufacturers for “failure to warn” of the hazards in their products). Their argument is tied to a broader initiative by industry-affiliated scholars and attorneys to curtail state-level product liability lawsuits (Viscusi et al., 1994); the U.S. Supreme Court heard oral arguments in the broadest of these cases (Wyeth v. Levine) in the fall of 2008. Philipson and Sun conclude that “joint regulation of drug safety can be inefficient when the regulatory authority mandates a binding and well-enforced level of safety investment,” which is implausible when considering the obvious fact that safety (and any level of safety guaranteed by a regulator) are subject to uncertainty. The argument also ignores an entire mathematical literature on the potential advantages of redundant arrangements (Bendor 1985; Heimann 1997; Ting 2002).
a wide variety of popular discussions as well (Kinsley 1989, Miller 2004). And it is almost certain wrong.

Scholarship focused upon the safety-versus-access tradeoff misses two crucial principles that must inform rigorous analysis of regulatory institutions. The first is a principle of modeling and theoretical understanding. Drug quality and safety are characterized by dynamic uncertainty, and no one – not the company producing the drug, not the physicians studying it or prescribing it, and not the regulator approving and controlling it – knows the certain truth about its safety or its therapeutic benefits. This inescapable uncertainty is resolved over time, through sequential experiments of the randomized controlled trial (RCT) variety and others. Hence *any scientifically valid and informative model must have both dynamic and stochastic elements*. Dynamics imply that the agents are rational over time and compute the “continuation value” of different actions – the value of staying with the current choice next period, when the same decision must be confronted again and it is assumed that the agent will act rationally then, too. Stochastic realism implies that agents assign subjective probabilities over time to different outcomes, where the agents’ subjective expectations are revised according to the information they observe over time as it unfolds in a partially random manner. A large number of existing models of markets and regulation in the pharmaceutical and health sectors fail to include these critical elements and are thus fundamentally flawed (Ma and Burgess 1993, Karlsson 2007, Philipson and Sun 2008).  

---

4 For models and empirics that include both stochastic/uncertainty considerations and dynamic rationality, see Crawford and Shum (2005), Coscelli and Shum (2004), and Carpenter (2004), Carpenter and Ting (2007) and Carpenter, et al (2010).
A second critical principle that must govern the study of pharmaceutical regulation is one of historical development. In the realm of pharmaceutical regulation, both in the United States and as well in modern Europe – there is not an exogenous industry or market or production process that was then regulated ex post by the government. Instead, regulatory institutions such as the FDA, professions (academic pharmacology and statistics), pharmacopoeial societies and others helped to create the methods and technologies by which drug companies became modern (Marks 1997). In the United States particularly, but also in Europe, *regulation did not intrude upon a pre-existing marketplace; it constituted a new marketplace*. The new marketplace was, moreover, characterized by greater and higher-quality information, and one of greater confidence among physicians, patients, investors and others.

**Public Law and Organizational Reputation**

The regulatory authority of the Food and Drug Administration (FDA) in the area of pharmaceuticals has largely arisen in response to democratic valuation of the risks and uncertainty associated with the pharmaceutical marketplace and the subjectively perceived capacity of regulatory institutions to address those risks (Young 1990; Marks 1997; Hilts 2003; Carpenter 2010). This process has been complicated and is not easily simplified, but its broad outlines suggest a politics of reputation. Where the risks and uncertainty of the pharmaceutical marketplace have been deemed high or salient, and where the capacity of the regulator to address these dilemmas and reduce uncertainty has been perceived as high – as in 1906, 1937-1938 and 1960-1963 – national institutions in America have strengthened and endowed further capacity and discretion upon the national government’s drug regulatory agency. The
organizational reputation of the FDA, combined with the uncertainty and scientific transformation of the pharmaceutical realm, has been the central dynamic in the evolution of American pharmaceutical regulation. Whereas public choice and legal scholars have drawn upon capture and rent-seeking theories of regulation to account for the origins of many regulatory statutes and institutions (Stigler YEAR; Temin 1980; Bartel and Thomas 1987), these theories fail to explain salient patterns of American and global pharmaceutical regulation. In particular, firms that have stood to benefit from the application of new and more stringent regulation have consistently opposed the introduction of these requirements (Jackson 1970, Carpenter and Sin 2007).

The legal authority of the FDA in the area of pharmaceuticals has resulted from the co-evolution of administrative practice, scientific change and statute. By virtue of the 1938 Federal Food, Drug and Cosmetic Act, the FDA became endowed with the power of pre-market review of new drug applications (NDAs). The 1938 Act defined the category of new drug and defined the reference standard of “generally recognized as safe” (GRAS). The Act followed a long process of proposal generation by the FDA’s parent agency at the time, the U.S. Department of Agriculture, but the power of pre-market notification was not proposed until after the sulfanilamide tragedy of 1937 (Jackson 1970, Marks 1997, Carpenter and Sin 2007). In that episode, the excipient of a new sulfanilamide drug contained diethylene glycol (the operative ingredient in antifreeze, and usually fatal if ingested). Over 100 people died from taking the medication, and FDA officials were roundly praised for quickly finding and collecting the remaining supplies of the drug before further harm could be caused. In the weeks following the episode, Department of Agriculture officials proposed a system of pre-market review, and
the essential veto power of American pharmaceutical regulation was embodied in legislation passed and signed in June 1938 (Jackson 1970, Carpenter and Sin 2007).

Although Congress did not authorize the agency to consider the effectiveness or efficacy of new drugs in the 1938 Act, the FDA did so nonetheless. Careful and thorough research by historian Harry Marks (1997) shows that FDA medical officers were explicitly taking account of “efficacy” and “therapeutic value” considerations in their review of the very first new drug applications submitted after the 1938 statute took effect. In the decades that followed, efficacy regulation became something of a procedural reality. The agency honed its new drug application form, established salient precedents in decisions on several important medications (not least the drugs Enovid and Altafur) and relied upon a sequenced battery of pharmacological tests developed by its own personnel to assess the merits and risks of new drugs (Swann 1997; Carpenter 2010, Chapter 3).

The FDA hence began to rely upon efficacy considerations, sequenced clinical trials (some of them randomized) and other elements of a procedural vision of drug regulation a decade or more before the thalidomide tragedy of 1960-1962. In that tragedy, a sedative (trade name Contergan in Germany) was marketed widely; when taken during a particular interval of human gestation, use of thalidomide was associated with stark and vivid birth defects. The epidemic of these birth defects (phocomelia, or “seal’s limbs”) became widespread in Europe and Australia in 1960 and 1961, and at first medical and regulatory authorities were uncertain as to its cause. Then in the fall of 1961 medical authorities in Australia and Germany announced that thalidomide ingestion by pregnant mothers was the cause.
The United States essentially avoided the tragedy because FDA medical officer Frances O. Kelsey refused to authorize the marketing of the drug. When Kelsey’s resistance was publicized by Washington Post writer Morton Mintz in July 1962, Kelsey became a medical and regulatory luminary and was roundly celebrated for her rigor in reviewing the drug. Just as important, the legislative momentum for a dramatic and broad strengthening of the federal government’s control over new prescription drugs had launched again. The Drug Amendments of 1962 were passed later that autumn, and they converted a regime of pre-market notification (where the burden of proof for drug approval lay with the FDA) into a regime of pre-market approval (where the burden of evidence was shifted to the drug’s sponsor, usually a drug company). The 1962 Amendments were accompanied by broad federal rules issued in 1962 and 1963 that literally redefined the process of clinical research and drug development in the United States and in most of the industrialized world. Drug development would now proceed according to planned clinical trials that were sequenced in three phases; a “Phase 1” for safety investigations in largely healthy volunteer subjects, a “Phase 2” for further investigation of safety and early investigation of efficacy, and a pivotal “Phase 3” where the largest clinical trials were required. This phased system of experiment now governs the modern pharmaceutical world in and across numerous countries.

The thalidomide and sulfanilamide episodes reveal the centrality of organizational reputation in the evolution of economic regulation. Regulatory agencies with stronger reputations – not just stronger beliefs about the agency but also stronger and more independent networks in which those beliefs are supported – are more likely to gain policymaking autonomy, discretion and crucial resources (Carpenter 2001). The sulfanilamide
and thalidomide tragedies provided a crucial basis of reputation for the FDA, and the agency’s organizational image (as protector of American citizens and as scientifically and administratively competent) functioned as a crucial public and legislative consideration in the various decisions to grant FDA vast authority over the pharmaceutical marketplace. In the wake of the laws, the FDA acquired other powers through rulemaking and through deference of the courts and medical and scientific authorities, and these moments of power too were shaped by reputation.

Since the early 1960s, the FDA’s powers have expanded gradually and forcefully, but contingently. An important mechanism of power elaboration has come in rulemaking and the issuance of guidance documents. The agency’s rulemaking ability has generally been used to broaden the set of experimental requirements upon new drug sponsors, and rulemaking and guidance have together been used to articulate the paradigm of modern medical research, the randomized, double-blinded, placebo-controlled clinical trial. Other rules have elaborated the equivalence criteria for generic drugs, have outlined governance standards for animal and clinical laboratories (including “Good Laboratory Practices” and “Good Clinical Practices”), and have reordered drug labeling and advertisements.5

At the same time that political and judicial authorities sanctioned expansive reach for the FDA in the realm of pharmaceuticals, they also recognized important limits upon the agency’s power. While agency scientists and others have long wished for FDA authority to regulate nutritional supplements (including herbal remedies, vitamins and other compounds whose manufacturers make implicit or explicit therapeutic claims about their products) in the same way as new drugs are governed, Congress has repeatedly refused to grant specific authority to the FDA over these products. The agency floated new rules for governance of dietary supplements and vitamins in 1973, but Wisconsin Democratic Senator William Proxmire sponsored an amendment (eventually named after him as the “Proxmire Amendment” of 1976), which established a new section 411 of the Federal Food, Drug, and Cosmetic Act. Section 411 enjoins the FDA from establishing standards to control the potency of vitamins in food supplements, and also prohibits the agency from governing supplements as drugs based entirely on their potency. In 1994, after FDA Commissioner David Kessler proposed to expand regulation of dietary supplements and herbal remedies, Congress passed the Dietary Supplement Health and Education Act (DSHEA) of 1994, which negated the FDA’s ability to engage in pre-market regulation of dietary supplements and also restricted the agency’s authority to test these products once on the market. Kessler’s attempt to regulate cigarettes and other tobacco products as devices for the delivery of nicotine was struck down not by Congress but by the federal courts, which examined the Federal Food, Drug and Cosmetic Act (September 27, 1977) 49612-30. Current FDA regulations are summarized in 21 CFR Part 50 (Informed Consent), part 56 (IRB Standards), part 312 (rules on Investigational New Drugs) and parts 812 and 813 (investigational devices). The origins and evolution of, and political controversy occasioned by, these rules is examined at length in Carpenter (2010, Chapters 4 and 8).
and its legislative history found no congressional authorization or intent to allow the agency to regulate tobacco products.⁶

Despite the restrictions that federal courts have placed upon FDA regulation in recent years – including not only the Brown and Williamson verdict but also the Washington Legal Foundation case in which the agency’s ability to conduct pre-market review of drug promotional materials was held in violation of the First Amendment’s recently derived “commercial speech” protections – the broad trends of American jurisprudence have generally provided additional power and deference to the agency. In the Dotterweich case of 1943, the U.S. Supreme Court held that company executives could be held criminally liable for violations of the Federal Food, Drug and Cosmetic Act even if they were not personally involved in the wrongdoing (320 U.S. 815, 64 S.Ct. 367). And in a set of crucial decisions in the 1970s, the Supreme Court allowed the agency to remove thousands of drug products from the market without holding an administrative hearing for each and every one; several years later the Court held that even cancer drugs for terminally ill patients did not qualify for an exception from the FDA’s pre-market approval authority.⁷

Theoretical Review

⁶ The central federal judicial decision on tobacco was handed down by the U.S. Supreme Court in FDA v. Brown and Williamson Tobacco Corp (98-1152) 529 U.S. 120 (2000); 153 F.3d 155, affirmed.

⁷ Washington Legal Foundation v. Henney, 202 F.3d 331 (D.C. Cir. 2000). The agency’s power of summary judgment and of rulemaking in Weinberger v. Hynson, Westcott and Dunning, 412 U.S. 609 (1973); several other cases were decided at the same time. The cancer case concerned the unproven treatment Laetrile; U.S. v. Rutherford, 442 U. S. 544 (1979). These cases are discussed at length in Carpenter (2010, Chapters 5 and 6).
Drug Review as a Stochastic Optimal Control Problem

The agency’s fundamental authority rests in pre-market approval of new drugs. It is this authority that tethers pharmaceutical companies so tightly to the FDA, it is the shadow of a rejection that leads companies to restructure or abandon drug development projects years (even decades) before they are possibly submitted to the regulator, and it is this authority that allows the FDA and its scientists to shape the very content of technical concepts that are used in medical research. It is useful, then, to begin theoretical analysis with a characterization of this decision. Elsewhere (Carpenter 2002, 2004b) I have described this choice process as an optimal stopping problem. In such a decision, a government agent must decide not simply whether to approve the drug but when to do so and with what base of evidence. These models embed two key concepts and requirements: (1) a random unfolding of evidence in the form of a stochastic process that is partially unanticipable, moving in both continuous and discrete movements, and (2) dynamic rationality in the sense that optimal stopping policies take into account the value of continuation (waiting for more information or better opportunities).

In several papers (Carpenter 2004b, Carpenter et al 2010), I have modeled regulatory decision making and other optimal stopping decisions using some form of a continuous-time stochastic process, most often a “diffusion” or Wiener process, which has normally distributed movements. More recently, Carpenter and Grimmer (2009) have used a more generalized stochastic process known as a Levy process, which includes both continuous diffusions and “jumps.” The idea is that much of the evidence for the quality or efficacy in a drug comes in more continuous movements of a disease state, but there are also unanticipable negative events such as safety hazards that can arise suddenly and impose disproportionate negative
utility upon patients. We can represent a “spectrally negative” Levy process by the following simple equation for $X(t)$, which is the evidence on drug $i$ observed by the regulator at time $t$.

$$X_{ij}(t) = m_{ij}t + s_{ij}w(t) + l_{ij}xt - \sum_{k=1}^{J(t)} Z_k$$

(1)

In all such equations in this paper, drugs are subscripted by $i$ and the diseases they treat are subscripted by $j$. On the right-hand side of equation (1) there are two deterministic terms (the first and the third) and two stochastic terms (the second and the fourth). The first term of the equation represents the “drift” of the continuous diffusion process ($m_{ij}$ can be positive or negative), while the third term of the right-hand side represents a “compensator” for the jumps (the product of terms $l_{ij}x$ is non-negatively valued). These terms can be combined according to their common factor $t$, of course, but are left separate here for illustrative purposes. The variable $w(t)$ is the Brownian motion or Wiener process, a normally distributed variable with mean 0 and variance $t$, with $s_{ij} > 0$ its “diffusion coefficient.” The final term sums up the number of downward jump shocks $J(t)$, where each jump is of size $Z_k$ and is governed by a distribution $G(Z)$ that takes values in the non-negative real numbers ($\mathbb{R}^+$). This fourth and final term of equation (1) represents a discontinuous “safety” variable that represents “all of a sudden” events like myocardial infarction, severe liver damage (hepatotoxicity), organ failure, allergic reaction, or something else. The intuition of this equation is that continuous data reveal good and bad outcomes incrementally (better or worse healing of the condition treated), whereas discrete data represent safety-related events, hence bad data can reveal itself discretely whereas good data does not.
The problem faced by the regulator is then to approve the drug at time $t_{app}$ if and only if the benefits of the drug outweigh its hazards and the value of waiting for more information.

We consider a fixed and known “approval payoff” $A$ and allow the regulator to consider calculable posterior moments of the observed continuous diffusion process, as follows. The posterior mean can be calculated as $E(m \mid \infty) = \frac{m}{s} + \frac{(\infty \cdot s^2)}{(1/s) + (t/s^2)}$, where $m$ and $s$ denote the known mean and variance of a background “generating distribution” of drugs that the regulator expects to see over time. Similarly the posterior variance of the mean is

$$V(t) = \frac{1}{(1/s) + (t/s^2)}.$$

Then with $y$ and $w$ serving as variables of integration, and $d > 0$ a constant discount factor, the regulator’s problem can be written as

$$\max E e^{-d(t_{app})} \left[ A + E \sum m_{i,j} Z_{k} \hat{\phi}_{i} e^{-d(y-w)} \hat{\phi}(s,w)- \sum_{k=1}^{l} \hat{\phi}_{i} Z_{k} + l_{i,j} x_{i, j} \right]$$

$$= E e^{-d(t_{app})} \left[ A + \sum m_{i,j}^{*} \hat{\phi}_{i}^{*}, w_{i,j} \hat{\phi}_{i}^{*} Z_{k} + l_{i,j} x_{i, j} \right]$$

(2)

A key contribution of the reputation-based perspective on regulation is that regulators may be reluctant to revisit past decisions. This generates an endogenous “irreversibility of approval” for many decisions (Carpenter 2002, 2004b). The point that the FDA’s public responsibility is founded in public expectations about the agency is put succinctly by former FDA general counsel Richard Merrill.

Citizens may complain when local police fail to curtail unlawful or violent activity, but few believe that even the best-functioning police force can solve, much less prevent, all
crimes. FDA is believed to have a different role, a responsibility to prevent harm before it occurs. The law makes it unlawful, without proof of intent or demonstration of actual injury or deception, to market drugs that the agency has not approved. In some sense, the agency becomes a warrantor of manufacturer compliance with the rules that govern drug development and marketing. This responsibility is implicitly acknowledged in the agency’s own publications, is frequently referred to in press accounts of its performance, and historically has permeated the dialogue between the agency and congressional oversight committees. FDA is repeatedly reminded, and often reminds us, that it shares responsibility for any drug that causes harm.8

This irreversibility of regulatory approval bestows a particular form upon the optimal policy for the regulator in this model. Carpenter and Grimmer (2009) derive a generalized approval policy for the regulator in the form of a first-passage time policy (see Figure 1).9 The space of possible data points is divided into a continuation region (where the evidence suggests waiting to approve) and an approval region (where the evidence suggests approval). The data (the Levy “evidence process” whose law of movement is described in equation (1)) starts in the continuation region (else all drugs would be approved immediately), and the drug is approved when and only if the evidence process crosses the barrier separating the continuation region from the approval region for the first time. The equation for the barrier expresses the dependence of the regulator upon information from both the continuous and discrete data of the process, as well as dynamic rationality and stochastic uncertainty. The policy equation is

\[ h^*(t) = d(l x - A) + \frac{1}{2s^2} V(t) F_{\text{crit}}(\hat{g}(t), t) + I_{\hat{O}_{R_i}}(\hat{g}(t), t - Z, t) - F(\hat{g}(t), t) W(Z) \]  (3)


9 Important technical details necessary for the understanding of these models occurs in the mathematical literature on stochastics and dynamics. See Billingsley (1979) and Karatzas and Shreve (1997).
where $F$ is a valuation function that maps evidence into values (subscripts denote partial derivatives with respect to the variable subscripted) and $\hat{g}(t)_j = \dot{m}_t + l_{xt} - \sum_{k=1}^{j(t)} \hat{a}_k Z_k$. The barrier policy is depicted in Figure 1, where the Levy process is simulated over 100 different evaluations and the bold stochastic process represents the median of these simulated evidence variables. The downward sloping, convex deterministic line is the approval policy described in equation (3). Note that because irreversibility heightens the value of waiting for more information or better outcomes, even a highly-valued evidence process is unlikely to generate drug approval early in the review [technically this comes from the $V(t)$ term in equation (3)]. This generates a result called the scarcity of quick approval (Carpenter 2004b) for which there is abundant evidence (Carpenter 2002). The scarcity of quick approval in this model is not a result of “slowness” but is dynamically rational and a result of uncertainty aversion (not risk aversion strictly defined).
This model, while technical, is very general and offers intuitive and instructive guidance to understanding FDA therapeutic review decisions for drugs, medical devices, biologics and other new products. It also expresses some features of the internal culture of drug review at the FDA since the 1950s (Carpenter 2010, Chapter 3, Chapter 7). In addition to the scarcity of quick approval, different versions of this model offer a number of predictions that have received consistent support in empirical research. These include:

- *Early Entrant Protection* – a simple application of the model to the scenario of repeated drug reviews for a single therapy products the simple prediction that drug review times are an increasing function of the number of drugs already approved for the target disease. Evidence for this hypothesis is provided in Carpenter (2002) and Carpenter et al (2010).
• _The Advantage of Regulatory Familiarity_ – Olson (1997) was the first scholar to show that FDA review times for new drug applications were a decreasing function of the number of drug previously submitted by the firm. The modeling framework here predicts this relationship under a wide variety of parameter values and assumptions; the more familiar is the regulator with the firm’s products (based upon past regulatory interactions), the lower the implicit “uncertainty penalty” that the drug starts out with (or the lower the barrier in Figure 1).

• _Effects of Deadlines and Other Institutions_. Adding deadline institutions to the model, Carpenter and Grimmer (2009) find that deadlines can exponentially heighten regulatory error rates and can actually increase the uncertainty associated with regulatory decision times. In a recent empirical study, Carpenter, Zucker and Avorn (2008) find evidence that drugs approved right before congressionally-imposed deadlines are several times more likely to be withdrawn from the market for safety problems at a later date or to experience black-box warnings added to their label.

**Approval Regulation Game**

The stochastic model of drug approval presented in the previous section suffers from an important limitation, namely the exogeneity of the drugs under review by the regulator. In reality, firms develop therapeutic products strategically and using somewhat different goals than are pursued by regulatory officials. In these development strategies, firms confront much of the same uncertainty that buffets the regulator during drug review. So too in research and development (R&D) conducted by the firm, the same conditions of dynamic optimality that
governed regulatory drug review are applied to over-time sequences of decisions to invest in or abandon drugs, or to stop the R&D process by submitting the drug for approval.

The interaction of regulator and firm in an uncertain and dynamic environment points to a game-theoretic representation of equilibrium drug regulation, unlike the decision theoretic representation in the previous section. Carpenter and Ting (2007) present a very simple model of this interaction with a single firm and a single regulator. The uncertainty and distribution of drug “quality” is represented by a Beta distribution, which maps parameters into a “curing probability” between zero and one. The key assumption of the model is that neither the firm nor the regulator know the true efficacy of the drug, but that the firm is better informed than the regulator in the sense that its prior beliefs are more precise. The game begins with a decision of the firm to either experiment (E) with the drug (this is a “continuation strategy,” and is modeled as paying a certain cost to observe the outcome of a 0-1 experiment, or a “Bernoulli trial”), to abandon or withdraw the drug (W), or to submit the drug to the regulator for possible approval (S). After each experiment the firm chooses whether to abandon the drug, to submit it to the regulator, or conduct the experiment again and face the same choice structure in the following period. The product can be experimented upon a maximum of two times, such that in the final “fish-or-cut-bait” period of firm’s possible moves, the firm must submit the drug or drop the project entirely. If the firm submits the product to the regulator, then the regulator approves (A) or rejects (R) the product. The structure of the interaction is observed in Figure 2.
Carpenter and Ting derive perfect Bayesian equilibria of this game, and show that the most compelling equilibria for analysis lay in mixed strategy equilibrium in which the regulator chooses a rejection strategy that induces the firm to abandon some moderately good drugs out of fear that the regulator will reject them. A crucial parameter of the model is the firm’s cost of experimentation. This is presumed to be smaller for large and established firms that have experience and capitalized development projects, and is assumed to be higher for smaller and newer firms. An interesting prediction of the model is that, conditioned on a longer period of
experimentation (two periods), then firms with higher cost of experimentation will be more likely to have their products accepted, as their higher development cost functions as a credible signal of the underlying quality of the product. This prediction is the reverse of that implied by standard games of incomplete information (such as signaling games), and it points to the importance of sequential interaction and two-sided uncertainty in the game form. It also has empirical support, in that less experienced drug sponsors are actually more likely to have their products approved by the FDA, conditioned on a longer period of experimentation (see Carpenter 2010, Chapter 7).

**Consistency with Empirical and Historical Observations**

The approval regulation game framework offers a set of predictions and perspectives that is unique among existing models (verbal and mathematical) of pharmaceutical regulation as well as other regulatory processes in which a government actor must license or approve a product or firm prior to its entry into an established marketplace. I review two of these predictions here and adduce some evidence for them; each is subject to much greater scholarly inquiry and entails a new research agenda of its own.

- **Equilibrium (Induced) Drug Project Abandonment** – A crucial perspective from the approval regulation model of Carpenter and Ting is that many drugs are abandoned not because the evidence for them is poor in and of itself, but because the evidence is insufficient to overcome the firm’s fear or rejection or delay by the regulator. Statistical detection of such rejections is tricky, because by the argument and model of Carpenter and Ting (2007), such abandonments are endogenous to regulatory action, so any
simple correlation between regulatory action and observed drug abandonment
decisions is insufficient to establish causation. Carpenter (2010, Chapter 8) considers a
different research tactic, namely looking at the stock price movement in the sponsor
company following an adverse decision by the FDA. If these stock price movements
associated with regulatory decisions are unanticipable (and the argument for this claim
can be based upon rational expectations and efficient asset pricing markets theories),
then they are plausibly exogenous to subsequent decisions by other firms to abandon
drugs. Under the measurement intuition that larger stock price shocks following an
adverse FDA decision can be considered a larger surprise relative to the decision,
Carpenter (2010) examines the correlation of (a) one-day movements in stock prices for
the firm rejected or delayed by the FDA with (b) next-period (subsequent month)
number of drugs abandoned by other firms. Table 1 presents a coefficient estimate
from a simple time-series regression of abandonments on “surprise” shifts in the stock
price of a company rejected or delayed by the FDA in the previous month. A standard-
deviation shift in the sponsor’s stock price (14.79 percentage points) after an FDA
rejection or delay corresponds to the abandonment of 4.58 drugs in the month
following the FDA decision. Note that if there were no surprise value in the FDA’s
decisions, the stock price shift would be small (expectation zero), and hence the FDA
rejection would not be measured.


table

<table>
<thead>
<tr>
<th>Table 1:</th>
</tr>
</thead>
</table>
## Time-Series Analysis of Regulatory Surprises and Next-Period Drug Project Abandonments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (Std. Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month (Trend Variable)</td>
<td>-0.01 (0.06)</td>
</tr>
<tr>
<td>FDA_REJECTION•(%ΔSTOCKPRICE)$_{t-1}$</td>
<td>-0.32 (0.13)</td>
</tr>
</tbody>
</table>

Model is a Prais-Winsten model with an autoregressive term and with autoregressive error; \( N = 170 \) months. Count regression of number of abandonments gives similar results.

- Development, the Regulatory Agenda, and Regulatory Error. Another implication of the approval regulation model of Carpenter and Ting (2007) is that the Type I and Type II errors of the regulator will be correlated in particular ways with the experimentation cost of the firm submitting the product. In particular, when experimentation with the product is shorter, the. But when experimentation has taken longer, products
submitted by small firms (higher experimentation costs) will be more likely to be approved and, hence, more likely to be subject to Type I error (faulty approval).

Carpenter and Ting find evidence for this proposition in a statistical analysis of safety-based withdrawals from the worldwide pharmaceutical market (2007: Table 4). The following graph (Figure 3) displays the general relationship, namely that as firm size (measured by previous investigational new drugs (INDs) submitted) increases, and as experimentation cost decreases, the likelihood of the sponsor’s approved drug being withdrawn from the market declines.¹⁰

¹⁰ The models control for firm-level fixed effects, severity of the treated disease (length of hospitalizations associated with the drug’s primary indication) and a time trend. Further evidence from the 1970s and 1980s, using a General Accounting Office (GAO) report on postmarket experience of approved drugs, supports the hypothesis. See Carpenter and Ting (2007) for details of the estimation, including an extreme-value regression estimator which takes account of the “rare event” nature of safety-based withdrawals.
Figure 3: For Longer-Developed Drugs, Drugs Sponsored Smaller Firms Are Associated with Higher Postmarket Withdrawal Rates, [Graphed Estimates from Carpenter and Ting (2007)]

Normative Questions of Public Law

Pharmaceutical regulation involves some of the most difficult political and normative issues confronting modern American science and government. What constitutes an appropriate level of “freedom” with regards to medications? Should terminally ill cancer patients have the right (legal, normative or constitutional) to choose and take any drug they wish, whether or not it has been approved by a regulatory agency? More generally, should a government regulator aim to approve new therapies and make them accessible as quickly as possible, erring on the side of provision? In many cases, these questions concern fundamental dimensions of whether governments should regulate markets and how much market-based institutions should be used to match therapies to patients and to regulate safety implicitly.

My own view is that while granting terminally ill patients some access to medication may be laudable ethically, there is no clear or plausible constitutional or legal right to access to medication, as the Court has ruled in U.S. v. Rutherford (1979) and more recently in Abigail Alliance v. FDA (2008). No claim to a right to therapy is legitimately founded in either the property protections of the U.S. Constitution (which protect only articles under the ownership of human agents, and not a broader right to purchase) or in the commonly construed privacy rights embedded in the Constitution. And while access to drugs is praiseworthy, and while the FDA’s acceleration in approval processes in the last three decades carries with it genuine benefits for medical treatment and patients, it is critical to remember that the primary
contribution of FDA drug regulation is not safety, but information. The primary benefit to society of the intensive research and development regimens and rigorous review processes entailed in U.S. pharmaceutical regulation lies not in the greater likelihood the agency has of getting the approval “right,” but in their revelation of much greater and higher-quality information about the therapy in question.

Another crucial normative issue concerns federalism and the legal implications of FDA product approval. In the Supreme Court case Wyeth v. Levine, pharmaceutical companies and others have drawn upon a recent line of thinking – including legal opinion (Viscusi et al 1994) and rules issued by the presidential administration of George W. Bush and the FDA’s general counsel at the time, Daniel Troy – to argue that FDA approval of a drug creates a presumption against tort suits against drug companies for “failure to warn” of drug risks. While the idea of limiting state jury awards and state-level lawsuits against drug manufacturers has merit, I am not persuaded that federal –re-emption is a legitimate or beneficial policy for doing so. The first reason is that, according to a strict construction of the Federal Food, Drug and Cosmetic Act, there is no statutory basis for pre-emption of state court decisions. This fact was not in dispute in the recent oral arguments heard before the Supreme Court in Wyeth v. Levine. The second and more important claim concerns federalism. State legislatures are more than capable enough of restraining their juries and judges, and indeed a number have enacted tort limits upon suits against physicians in recent years. More recently, Michigan has enacted a bill shielding pharmaceutical manufacturers from lawsuit based upon failure to warn, and Georgia is considering a similar bill at the present writing. It is best, I think, to let the states experiment
with these policies, especially where federal statute does not speak clearly to the priority and precedent of FDA decisions.

Beyond this, the tort system serves several interests that permit for welfare-enhancing redundancy of efforts (Bendor 1985, Heimann 1992). For one, it is well known that the FDA cannot fully police the safety of the modern pharmaceutical and therapeutic marketplaces (Avorn 2004); state-level tort liability therefore provides an extra check upon drug safety. Perhaps more important, important information about drug safety and about its disclosure has surfaced in state-level lawsuits against pharmaceutical manufacturers, and this disclosure threat alone may be sufficient to induce more honest behavior by pharmaceutical companies in a world that is characterized by asymmetric information. Again, it is not merely the safety implications of alternative regulatory arrangements to which we must pay attention, but also the informational implications of those arrangements. Any pre-emption policy – or any other policy concerning the relationship between federal approval regulation and subsequent tort liability – must consider its effects for information (or the lack of it) provided to different actors in the health care system.

**Agendas for Future Research**

The empirical agenda for research in the political economy of pharmaceutical regulation should also be oriented toward information and the examination of factors that induce greater or less “confidence” among different players in the modern pharmaceutical world: drug companies, investors, physicians, patients, health care providers and others. Like much of the current policy discussion, current research is too deeply sunk in a trade-off of limited
imagination, namely that between safety and access (including Carpenter 2002, Philipson, Berndt, Gottschalk and Strobeck 2008). In several ways, information should be the central focus of analysis.

One critical question is how and why drugs make their way onto insurance and health-plan formularies, which is a rough functional equivalent of asking why health providers pay for them. The expansion of prescription drug coverage for health plan formularies is a crucial, and too-little studied, development of the late-twentieth-century health system in the United States. One question is how and whether regulatory requirements and clinical experimentation standards led to changes in these formulary coverage decisions, and whether subsequent changes in formularies and coverage decisions have been affected by regulation.

Another central question is how patients and physicians learn about the drugs they prescribe and take, and how regulation enhances or limits this process. It is well known that many patients take only a fraction of the drugs prescribed to them (see Osterberg and Blaschke 2005 for a summary review). How does the information provided by regulation affect this process? Does the provision of extensive safety information serve to deter patients from taking the drugs that would benefit them? Or does it serve to ward patients away from drugs that they are taking only because the products are being marketed too heavily, and hence the consumption pattern is non-optimal in the first place? What are the effects of drug labeling upon patient adherence? Is there evidence that some patients read the labels, or read certain parts of them? Is there evidence that highly publicized drug safety events, or highly publicized clinical trial results, influence patient behavior?
Another set of issues concern how the current regulatory system produces information. It is again well known that most components of a prescription drug’s label are ignored by physicians, but that boxed warnings do affect physician prescribing and dosing behavior in many cases (Avorn 2004). Labeling thus serves as a highly imperfect, and often discontinuous, instrument of revealing information about drugs. Yet there are other ways in which drug information is revealed and presented to consumers. An important source of information for many consumers is advertising, not only that directed towards physicians but that directed towards patients themselves (so-called “direct-to-consumer advertising,” or DTCA). It is well known that drug advertisements are little regulated (Avorn 2004), and even the constitutionality of regulating these advertisements is an open question. An important policy question thus presents itself. If there are real institutional constraints upon the ability of the FDA and other pharmaceutical regulators to govern advertising directed to patients, then how can the pharmaceutical regulator intervene appropriately and effectively in the health system to ensure that patients are receiving high-quality, unbiased information?

And finally, what about public trust in the institutions of science, medicine and regulation? When physicians, patients, health providers and payers trust the approval decisions of the regulator, what are the implications for utilization, therapeutic innovation and, ultimately, health? And at a time when public, congressional and medical confidence in the FDA’s pharmaceutical governance is declining (Avorn 2004, Carpenter 2010, Chapter 12), are there possible adverse implications of this lack of confidence for the billions of highly decentralized health decisions (diagnosing, prescribing, paying, medicine taking) that occur on a weekly basis in the United States?
I conclude with a gesture to issues that are ripe for analysis in public choice and public law, issues that concern the centrality of agency reputation. The first of these has to do with legislative delegation and legislative and judicial deference to agencies, the second of these has to do with the propriety and efficiency of an agency that is reputation-constrained and reputation-conscious in its decision making. Can the forces of reputation be harnessed to render regulatory agencies better performers of their tasks? Can the politics of reputation serve as another, perhaps more powerful and internal constraint upon administrative discretion and the exercise of power? What are the policy and legal implications of agency reputation and its implications for autonomy and administrative and regulatory power?

How might administrative law and public choice deal with these issues? Empirically, we need much more detailed and historically accurate portraits of agency reputation in the realm of regulation, not just in the pharmaceutical field but in telecommunications, financial security, environmental protection and others. Theoretically, we need an understanding of reputation that conceives of its embedment in a world of multiple audiences. The reality of multiple audiences in modern regulation is created not merely by a separation of power system (creating executive, legislative and judicial audiences for regulatory agencies in the United States), but also by the highly complex nature of modern society. The possibility for productive and complementary interplay between theory and historical analysis is great.

References


----. (2008), ‘The Risk We Bear: The Effects of Review Speed and Industry User Fees on New Drug Safety,’ *Journal of Health Economics* 27 (2) (March): 175-200;


