

**How Political Control of Bureaucracy Shapes Social Outcomes:
FDA Review Deadlines and Postmarket Drug Safety**

Daniel Carpenter*
Jacqueline Chattopadhyay*
Susan Moffitt#
Clayton Nall*

August 2010

ABSTRACT

Politicians deploy numerous policy instruments to control bureaucracy, yet we know little about whether particular instruments of control affect policy, market or health outcomes. We address these neglected questions by examining agency review and decision deadlines, asking whether deadlines (1) influence decision timing, and (2) are associated with significant market or health outcomes. We test our hypotheses on a unique dataset of FDA drug approvals from 1980 to 2008. Using dynamic time-to-decision models, we find that deadlines are associated with a piling of approvals around the deadlines. Using generalized linear models supplemented by matching analyses, we find consistent evidence that “just-before-deadline” approvals are associated with higher rates of market withdrawals and severe safety warnings, and modest evidence these approvals are associated with more postmarket safety alerts and dosage discontinuations. These postmarket outcomes are in turn associated with thousands of additional reported hospitalizations, disabilities and deaths due to adverse drug reactions.

* Department of Government, Harvard University

Department of Political Science, Brown University

Elected officials deploy a broad scope of institutional and administrative tools to influence government agencies' authority and decisions. Scholarship on agency design and appointments suggests instruments of political control bear on general agency performance (Lewis 2008), on agency processes (Yackee and Yackee 2010) and on the accuracy of agency outputs (Krause and Douglas 2006). However, do instruments of political control affect the content of bureaucratic outputs and the quality of ultimate policy outcomes? Do they shape the functioning of markets, social and health outcomes, or the everyday activities of government in its interaction with citizens? Theory predicts that greater political control over agency decisions constrains an agency's ability to deploy agency expertise (Bawn 1995, 63). Such constraints have the potential to affect not only agency decisions but also the substantive policy and human outcomes associated with those decisions. We put this prediction to an empirical test in the context of a ubiquitous yet common instrument of political control: deadlines on the timing of agency decisions.

Well-established scholarship specifies a range of procedures elected officials use to constrain agency actions (McCubbins, Noll, Weingast 1987, 1989; Moe 1990; Bawn 1995; Carpenter 1996; Huber and Shipan 2002; Shipan 2004). However, bureaucrats also have strong incentives to develop and sustain their capacities and expertise. These capacities can serve to strengthen their reputations for producing decisions and services, policy choices that the public as well as legislators recognize as good ones (Carpenter 2001, 2010; Krause and Corder 2007). Assessing the effects of deadlines on policy outcomes offers new evidence on whether and how agencies handle competing pressures to respond to elected officials and promote agency interests in their reputations.

Moreover, assessing the effects of deadlines on *policy outcomes* (e.g. poverty rates, pollution levels) extends scholarship beyond its attention to *agency decisions* (e.g. benefit eligibility criteria, licenses). In doing so, we are able to discern the impact of instruments of political control on policy outcomes, distinct from the effects that derive from policy uncertainty and bounded rationality.

Bureaucrats face uncertainty about their agencies' futures and make decisions under bounded rationality (Krause 2003), yielding implications for policy outcomes (including error) even when decisions are rooted entirely in expertise.

We draw upon new data and refined statistical methods to address these two dimensions of regulatory choice: decisions and outcomes. We first rely upon a statistical model rarely used in the study of administrative and regulatory durations – the dynamic Cox model with time-varying covariates – to “peer inside” the regulatory decision and assess whether the deadline influenced its ultimate timing. We apply these analyses to the review of new drug applications (NDAs) by the U.S. Food and Drug Administration (FDA), a subject of increasing interest in contemporary medicine and politics. The enactment of the Prescription Drug User Fee Act (PDUFA) in 1992 and its revisions in 1997, 2002, and 2007 imposed review time goals, widely known as “PDUFA clocks,” upon the FDA. These review goals, which can be viewed as deadlines with implicit penalties for violation, have influenced the timing of FDA decision-making.

We then examine whether deadlines bear on policy and health outcomes: whether drugs approved shortly before a deadline experienced post-marketing safety issues at a greater rate than drugs approved at other points in the review process. We supplement these models with matching analyses.

These outcomes are different from those usually examined in the study of administrative or regulatory politics. They are, first and foremost, *market outcomes* and system outcomes – if a drug is withdrawn or discontinued it will be generally unavailable to patients or their physicians who would wish to use it. If a drug receives a black-box warning or a safety alert it will change the market for the drug, and also the character of medical practice (it may become “standard practice” not to prescribe it, or to prescribe it only after other therapies have failed). In addition, the outcomes we examine are *health outcomes* (both in what they cause and in what they reflect); that is, they reflect

quantifiable human impacts in terms of hospitalization, disability and even death. By correlating these postmarket outcomes with adverse event reports on drugs, we adduce observational evidence that human lives, hospitalizations and disabilities may be associated with these outcomes. In short, the policy outcomes we study are social outcomes with social welfare implications – they involve markets, professions, lives and deaths.

Our analysis assesses the impact of Food and Drug Administration deadlines for drug reviews on drugs' subsequent post-marketing experiences for the period ranging from 1980 to 2008.ⁱ This expansive time frame allows our analysis to assess policy outcomes across multiple political regimes. Moreover, a wide variety of administrative decisions are characterized by timing phenomena, including licensing (dam license renewal by the Federal Energy Regulatory Commission (Kosnik 2006), for instance), and rulemaking. While our empirical analysis focuses on drug approval in the United States, applications to other national settings and regulatory policies are possible. Drug regulators in the European Union also operate under review time deadlines, as do many other agency decision processes.ⁱⁱ Though our inquiry focuses on one form of political control (deadlines) and one agency (the FDA), the prevalence of deadlines across government agencies gives our analysis broad applicability.

Deadlines as Potential Instruments of Political Control and Policy Influence

The duration of choice is a crucial dimension of government decision-making. Many agencies and regulators possess discretion over not only their decisions, but also over when those decisions are made. While scholarship in institutional political science, economics and sociology has begun to consider questions of bureaucratic timing (e.g., Ando 1999; Carpenter 2002; Kosnik 2006; Olson 1997, 2002; Spence 1999; Whitford 2005; Yackee and Yackee 2010), few scholars have considered how externally imposed administrative institutions, such as deadlines, influence regulatory decisions and their substantive policy outcomes.

Statutorily codified deadlines appear for a range of policy domains, ranging from environmental regulation (Morgenstern 1993), to traffic safety regulation (Mashaw and Harfst 1990, 69-83), to the collection of educational statistics. The 1988 amendments to the Federal Insecticide, Fungicide and Rodenticide Act, for example, required the Environmental Protection Agency to re-register 700 substances within a period of ten years (Morgenstern 1993, 245). Subsequent amendments imposed additional deadlines on the EPA's re-registration of pesticides even as they provided a stream of user fees to the agency (Schierow 2008, 8). In the field of pharmaceutical regulation, the FDA faces six month deadlines for the review of more therapeutically vital "priority" drugs and (since 1997) ten months for most other drug reviews. Other health regulators worldwide including the European Medicines Evaluation Agency (EMA) have adopted similar regulatory review deadlines.ⁱⁱⁱ

Despite the prevalence of deadlines, the vast scholarship on political control over the bureaucracy has devoted relatively little attention to deadlines as a potential instrument for influencing bureaucratic action and the policy consequences of those actions. According to McCubbins, Noll and Weingast (1987, 1989), and Moe (1989), "making an agency rely on private deadlines, imposing tight deadlines, or making certain types of decisions difficult to promulgate" are classic examples of procedural controls that reduce agency independence.^{iv} Reducing agency independence can influence agency decisions and ultimate policy outcomes by affecting the volume and kind of information agencies may receive before making decisions (Abbott, 1987). Structuring the flow of information may constrain an agency's ability to apply its expertise and make bureaucrats particularly responsive to certain interest groups (Balla and Wright, 2001; McCubbins, Noll and Weingast, 1987: 675).

The limited scholarship on the effects of deadlines, however, is inconclusive. Some surveys suggest bureaucrats consider deadlines a weak form of Congressional oversight (Furlong 1998).

Other studies report concern that deadlines limit bureaucrats' abilities to apply their expertise to policy decisions (HHS 2002). With the exception of Gersen and O'Connell's (2007) hypotheses on the benefits and risks of congressionally imposed deadlines on agency behavior, we are unaware of any literature that yet systematically and empirically tests how deadlines influence agency decision making, and little that specifies how deadlines influence organizational learning and policy outcomes more generally.^v

To understand the potential effects of deadlines, we begin by considering their mechanisms. Deadline institutions impose a penalty (explicit or implicit) for the endurance of a decision or administrative process beyond a specified point in time (the "deadline"). Where the deadline is absolute, this penalty may be conceived as "infinite" or large enough to outweigh all other factors in the regulator's decision, including the agency's reputation for product quality.^{vi} In other cases, the deadline penalty is smaller, such that the deadline becomes one of plural factors to influence the timing of regulatory behavior. Consider the possibility of a Congressional deadline penalty. Congress can choose to enforce deadlines by conditioning agencies' future resource streams on meeting those deadlines. Deadlines, thus, may pit an agency's interest in using its expertise to produce accurate, valid decisions—and in so doing building and protecting its reputation for expertise and the power therein—against its fundamental interest in having sufficient staff and funding in the future. Resource cuts are a certain, measurable, and quickly enforced punishment for agency failure to meet a deadline. In contrast, the reputational and power costs resulting from lower quality agency decisions might or might not ever materialize, are difficult to measure, and, absent a crisis tied to a swiftly made agency decision, will likely develop only over the long run, if at all.

Deadlines can affect regulatory choice in at least two ways. They can influence the duration of decisions by preventing regulatory processes from elapsing beyond a certain time (the "timing effect"), and they can influence the quality of those decisions. By quality we mean both procedural

quality (how much care was taken with the decision, the breadth of deliberation involved or the span of the information search involved) as well as outcome quality (the likelihood that the decision will result in associated errors or harms). Consider an analogy from the scholastic realm. Suppose the dean of a university instituted a new rule requiring professors to spend no more than 30 minutes with a term paper before grading it. If professors spent no more than 15 minutes grading papers before the limit was imposed, it would be expected to have little if any effect. If on the other hand the usual professor spent one hour with each paper before grading it, we would be interested in whether the new deadline really did shorten the grading time, and in questions such as whether students felt shortchanged. We might ask, then, not only (1) whether the timing of grading changed, but (2) whether the distribution of grades was changed by the imposition of a deadline.

Deadlines may bear on the quality of subsequent outcomes by constraining the amount of time, information, deliberation or other resources a decision-maker can apply to an uncertain policy decision. Deadlines may impair the quality of policy outcomes by restricting the expertise an agency can mobilize or apply to a policy problem. Deadlines may also convey political information: deadlines may send signals to the agency to attend to some interests over others. In this sense, deadlines may alter the decision criteria the agency uses, leading bureaucrats to privilege timeliness over other considerations, regardless of decision resources available to the agency.

To form hypotheses, we introduce the following two concepts.

Definition: *deadline regime*. A policy, passed by politicians, that imposes decision time goals or rules upon an agency, with associated penalties (implicit or explicit) for failing to meet the deadlines.

Definition: *deadline time* [t^{dead}]. The time-interval after which the decision is regarded as “late” or past the deadline. For a 90-day deadline, the deadline time is the 90th day; for a nine-month deadline, the deadline time is the ninth month.

Combined with these definitions, our examination of deadlines' mechanisms yields several general hypotheses:

H1 Timing: Deadline institutions that posit a deadline time will induce a concentration of decisions that occur at or immediately before the deadline time $[t^{lead}]$. Specifically, under a deadline institution, there will be a greater likelihood of decision at the deadline time relative to

- (a) the pre-deadline time interval (the time intervals before the deadline $t^{lead}-1$, $t^{lead}-2$ and so on)
- (b) the post-deadline time interval (immediately following the deadline; $t^{lead}+1$, $t^{lead}+2$, and so on).
- (c) the pre-regime deadline time (the deadline time intervals $[t^{lead}]$, but before a deadline regime was in place, or before there was a penalty associated with a decision after t^{lead}).

H2 Quality: Decisions that are completed at or just before the deadline in a deadline regime will be associated with higher rates of revisitation (the agency correcting possible mistakes) or higher rates of error or observed harm. (Note, however, that the deadlines may still be net beneficial if they speed time-to-decision and reduce uncertainty.)

H2a Revisitation: at-deadline decisions will be revisited more commonly (in terms of revised decisions, additional information provided, or consumer warnings) than decisions not at-deadline.

H2b Error/Reversal: at-deadline decisions will generate higher error than decisions not at-deadline.

The FDA User-Fee Program and Review Deadlines: A Brief Description

The FDA user-fee program and associated review deadlines offer an exemplary venue for testing these general hypotheses and for laying the foundation for systematic understanding of deadline effects across government agencies.^{vii} Examining deadline effects in the FDA enables our study to build on and yet expand established and expansive scholarship on regulatory agencies (Moe 1987; Scholz et. al 1991; Wood and Waterman 1991; Huber 2007), and more fully develop a recent line of inquiry that has focused on the data-rich environment of the Food and Drug Administration in particular (Heimann 1997; Carpenter 2002, 2004, 2010; Carpenter and Ting 2007; Moffitt 2010; Olson 1997, 2000, 2002, 2004a, 2004b). The FDA, moreover, offers important variation on our primary explanatory variable: deadlines. In the two decades before the Prescription Drug User Fee Act (PDUFA) of 1992 (which was subsequently reauthorized in 1997, 2002 and 2007), the FDA

faced no constraints on the amount of time it took reviewing new drug applications.^{viii} Intended to speed drug approval times, PDUFA represented a bargain struck between the FDA, Congress, and interested groups: user fees associated with PDUFA would augment FDA staffing resources while the pharmaceutical industry and concerned disease advocates would, in principle, receive quicker approvals and quicker access to new therapies.^{ix} PDUFA introduced a “review clock”: from the date of submission of the first new drug application (NDA), a countdown timer begins ticking. The time on the clock (6 months, 10 months or 12 months) differed according to whether the new drug application was designated “priority” or “standard.”^x

Among the features of the user-fee program relevant to FDA’s approval behavior is the absoluteness of the “PDUFA clock” deadline. Once the deadline has elapsed, CDER has far less incentive to hurry a drug, as it no longer can count towards annual review time goals. The user-fee law did not require 2- and 3-month reviews, nor did it forbid the occasional two-year review. Instead, the provision that eventually nine of ten drugs must be reviewed by the deadlines means that the PDUFA clock uniformly governs most all of FDA’s drug review behavior.^{xi} By imposing specific deadlines and linking agency compliance to the agency revenue source, PDUFA represented a significant effort to influence FDA actions.

We hypothesize that the PDUFA clock deadlines changed the timing of FDA decision behavior. Specifically, when we operationalize H1 for the case of FDA drug approval, we expect that the deadline regime (PDUFA) will have introduced a temporal discontinuity into drug review by prompting FDA to rush out approvals in the months just before the deadline passes.

Operationalized H1 (FDA Timing): The hazard of drug approval is higher at t^{dead} than in the months before t^{dead} or the months after t^{dead} .

Consider the 12-month review clock for standard drugs, and suppose we focus attention on the incentives of the agency to approve a drug in the next two months, however long the review has lasted to date. When the eleventh month of the review cycle starts, the incentives for completing

NDA review in the next two months are quite high since near-term completion will allow the agency to meet the review clock for this drug. If the agency fails to meet the review time goal however, there should be much less incentive to approve the drug in months thirteen or fourteen. Hence we should observe a high proportion of approvals concentrated or “piled up” in the months just before the deadline, and relatively few concentrated just after. The same logic should obtain for other deadlines.

Moreover, we expect that the deadline-induced changes in the FDA’s decision-behavior are associated with changes in the quality of subsequent policy outcomes. Broadly, we expect that incentives to meet the deadline will outweigh other considerations, such as the agency’s reputation for product quality, as measured by events signaling safety concerns during the post-market period. Specifically, we hypothesize that drugs approved in the window just before the deadline elapses were ‘rushed’ to market and thus were less likely to receive sufficient time and expertise applied to their reviews.

Operationalized H2 (FDA Decision Quality): Drugs approved at or just before t_{lead} will have more post-marketing regulatory events (revisitation and error).

While most drugs undergo some degree of regulatory modification in the post-market period as drug are prescribed to new patient groups, post-marketing regulatory events (PMREs) are a reliable measure of post-market safety problems that might have been addressed at the drug-evaluation stage. PMREs include black-box (i.e., safety label) warnings, firm- and FDA-issued safety alerts, safety-based withdrawals, and discontinuations of a form of drug from the market. We hypothesize that such events occur with higher frequency among drugs approved in the two months before the deadline than among comparable drugs approved earlier in the cycle or later in the cycle.

Our specific hypothesis departs from a broader, political signal interpretation of deadline effects. Our hypothesis does not suggest that deadlines are associated with lowering drug quality overall, which could arise if deadlines operated primarily as political signals, altering the agency’s

decision to attend to timeliness over expertise across all drugs reviewed during the deadline regime. Instead, we argue that the specific behavior change induced by the deadline – drug piling right before the deadline – is associated with specific changes in the quality of policy outcomes – more post-marketing regulatory events for drugs approved right before the deadline elapses. Alternatively, post-marketing regulatory events may be associated with inherent task uncertainty or from firm privilege in other aspects of regulatory review, rather than deriving from deadline-induced change in agency decision behavior.

Deadlines and the Duration of Regulatory Review Cycle

Methods and Measures: To test our first set of hypotheses, that deadlines alter the *timing* of agency decisions such that the hazard of drug approval increases in the month the deadline elapses, we retrieve statistical estimates of the conditional probability of approval at each point of the regulatory review cycle. At each month of the review cycle, what is the relative hazard rate of approval in the month, given that the drug has not yet been approved? We conduct likelihood-based hazard analyses of review times and retrieve month-specific hazard estimates that allow us to construct this desired statistical portrait of the FDA review cycle. To minimize dependence upon parametric statistical assumptions we employ Cox proportional hazard models. We reserve formal elaboration of our statistical models for the Online Appendix.

Our sample consists of New Molecular Entities approved between 1950 and 2008, for a total of 1034 drugs. Each model controls for staffing levels in the FDA’s Center for Drug Evaluation and Research and includes indicators for the 50 largest firms. We estimate separate models for priority and standard drugs,^{xii} and use the Efron method for handling ties.^{xiii}

Results: We hypothesized that deadlines induce drug approval ‘piling’ in the months just before a deadline elapses. We expect to find higher hazards in the month leading up to drug review

deadlines. For standard drugs, the original PDUFA deadline was 12 months, changed to a 10 month deadline after 1997. The deadline for priority drugs is 6 months.

[Figure 1 here]

Our analyses suggest that the PDUFA clocks have dramatically changed the behavioral structure of the FDA drug review cycle. Figure 1 plots the hazard ratios for priority drugs approved in the fourth, sixth and eighth month of the review cycle in the period before and after the introduction of deadlines.^{xiv} For priority drugs submitted after the introduction of six-month PDUFA deadlines, we observe a sizable increase in approval hazards in the sixth month of the review: an 8-fold increase in the hazard over the fourth month ($\chi^2=11.88$; $p=0.0006$) [test of H1(a)]. The hazard ratio drops nearly twelve-fold the eighth month, after the deadline has passed ($\chi^2=13.18$; $p=0.0003$) [test of H1(b)]. In the period before the introduction of deadlines, the hazard ratio in the sixth month of the review cycle is significantly lower than in either the fourth or eighth month ($\chi^2=13.18$; $p=0.0003$) [test of H1(c)].

These patterns generally apply to standard drugs as well. Figure 2 plots the monthly approval hazard ratios retrieved from dynamic Cox estimation for the first 24 months of the review cycle, where these can be estimated, for standard drugs. For drugs submitted before 1993, no discontinuity is seen at the tenth or twelfth months of review that is consistent with drug piling. For New Molecular Entities submitted before PDUFA, the approval hazard ratio for drugs in the tenth month was not statistically differentiable from that in the twelfth month ($\chi^2=0.05$; $p=0.83$). Nor is the hazard for drugs approved in the twelfth month statistically differentiable from that in the fourteenth month ($\chi^2=2.37$; $p=0.12$). That hazard for drugs approved in the tenth month is statistically differentiable from that in the eighth month ($\chi^2=6.68$; $p=0.01$), but in the opposite direction: the hazard ratio is lower in the tenth month than in the eighth.

[Figure 2 here]

For standard drugs submitted from 1993 to 1997 and falling under the provisions of the first user-fee law, we observe a sizable increase in approval hazards for the twelfth month of review compared to the same month in the second user-fee regime (starting in 1997) when the deadline changed to 10 months ($\chi^2=5.73$; $p = 0.02$) [test of H1(a)]. Moreover, as hypothesized, approval hazards fall off appreciably for the month after the review clock deadline, comparing the twelfth month with the fourteenth ($\chi^2=6.69$; $p = 0.01$) [test of H1(b)].

For the period since 1997, when the relevant deadline for non-priority NDAs was ten months, we observe a large increase in approval hazards in the tenth month of the review cycle, compared with the tenth month under the previous twelve month deadline regime ($\chi^2=8.22$; $p = 0.004$) [test of H1(c)]. Moreover, the approval hazard in the month before the ten-month FDAMA clock deadline elapses is two fold greater than the approval hazard in the month after the review deadline elapsed, though this increase does not approach standard levels of significance relative to the eleventh or twelfth month.

Across models, an increase in FDA staff is associated with an increased approval hazard, controlling for the year of drug submission. This is consistent with the implicit trade that characterized PDUFA: user fees (and associated staff) in exchange for faster review times. The user-fee laws, however, specified that additional staff acquired through user fees would be devoted to drug reviews, *not* post-marketing drug surveillance or safety.

Deadlines and Policy Outcomes

Results from the Cox proportional hazard models provide evidence that deadlines associated with PDUFA clocks are associated with specific changes in the timing of FDA review behavior: drug pile up around the review deadlines. Yet, do these changes in FDA decision behavior bear on

policy outcomes? In particular, do drugs approved in the window just before the deadline elapses have qualitatively different post-marketing experiences? We now turn to investigate the second set of hypotheses, namely that the review clock institutions have influenced not just the timing, but also the quality of the FDA's decision, particularly drug safety and post-marketing regulatory issues. While some analysts (including the FDA itself) have examined whether the *overall* rate of drug safety problems has risen or fallen since 1993,^{xv} we conduct a different, more focused comparison. We compare the post-marketing experiences of drugs approved in the months before the deadline elapses to the post-marketing experiences for drugs approved during other times in the different user-fee regimes.

Measures of Post-marketing Regulatory Events: We examine four measures of post-marketing regulatory events. One measure comes from a dichotomous indicator of whether the approved drug received a new post-market *black-box warning* on its label for a significant new adverse drug reaction (ADR) (Lasser et al. 2002).^{xvi} A second measure is a dichotomous indicator of whether an NME faced a *safety-based withdrawal*.^{xvii} A third measures FDA drug-specific *safety alerts*. These alerts include all safety-related letters, press releases, and health advisories circulated by the FDA and by pharmaceutical firms from 1996 through 2008, for all NMEs approved since 1996.^{xviii}

We also examine one variable that is not explicitly safety-related, but which may reveal issues of efficacy and clinical uptake: the rate at which *dosage-forms of the drug are discontinued* from the market place.^{xix} Though discontinuation “indicates drugs that have been discontinued from marketing or that have had their approvals withdrawn for other than safety or efficacy reasons,” that dosage withdrawal may be indirectly related to safety. When healthcare providers notice safety problems, clinical demand for a drug may decline. In many cases, dosage discontinuation arises from weak clinical demand.^{xx} We report summary statistics for these measures in Table 3.

[Table 3 about here.]

Specification and Estimation of Generalized Linear Models. We use generalized linear models to consider potential implications of review deadlines on subsequent drug post-marketing experiences. Drugs vary in numerous ways that are unobservable to the analyst and even to the researchers who study them. One advantage of a GLM testing framework is that it allows the analyst to control statistically for numerous sources of variation as long as the sample size permits it.

For our sample of new molecular entities, we introduce one main vector of parameters for estimation – a set of terms for each primary indication, or disease category (modeled as a random effect). The immediate result is that dozens (sometimes hundreds) of separate error terms or parameters are added to the models we estimate. Our indexation of primary indications, though, has the advantage of being much more refined than that used by other analysts of FDA drug approval (e.g., Carpenter 2002, Lasser 2002, Olson 1997, Olson 2004b). Other analysts control for generic therapeutic category terms (for example, a binary indicator for all anti-neoplastic drugs or all central nervous system (CNS) drugs), but not for particular primary indications. Because drugs are assigned to divisions based primarily upon their primary indication, the primary indication index \mathbf{y} is a sufficient index for CDER reviewing divisions, so any static factors associated with the division-level review organization are captured by this set of hundreds of terms.^{xxi}

In the regression results we report, we employ the generalized linear model (GLM) framework (McCullagh and Nelder 1989) for panel data and mixed effects models. Through the details in the appendix, recall that for any drug i , its primary indication is indexed by \mathbf{y} and its sponsor by k . Recall, too, that S_k and D_y serve as binary indicators for the drug's sponsor and primary indication, respectively. We observe several different indicators of a post-marketing regulatory event, which we denote by $y_{\psi ki}^{PMRE}$, and we estimate models of the form

$$y_{\psi ki}^{PMRE} = f\left(\alpha^S S_k + \gamma' Z_{\psi ki} + u_{\psi} + e_{\psi ki}\right) \quad (1)$$

where f is a function (whose arguments are always linear) to be specified, where u is a random effect term which is assumed uncorrelated with Z , where \mathbf{a}^s represents firm-specific coefficients, and where ϵ is a model disturbance. Notice that primary-indication-specific effects are modeled as random effects. Again, the appendix contains more details of modeling empirics.

For each of the four post-market event variables (black box warning indicator, safety-based withdrawal indicator, safety alert count, and dosage discontinuation indicator), we regress the regulatory event variable on the submission year (to capture the time trend), an indicator of whether the drug received approval within the month leading to or of a deadline (the primary explanatory variable of interest), and selected other measures in extended models. We thus create the following variables:

1. Agency Decision Criteria: Pre-Deadline Approval. Our previous analyses demonstrated changes in the timing of FDA reviews following the introduction of deadlines. We hypothesize that the specific change in FDA approval behavior – piling of approvals near the deadline – is associated with changes in drugs’ post-marketing outcomes. Thus, for any deadline month and its preceding month ($\tau^{\text{deadline}}, \tau^{\text{deadline}} - 1$), we construct a “pre-deadline” approval indicator Z^{PRE} scored 1 if the drug in question was approved in τ^{deadline} or $\tau^{\text{deadline}} - 1$, and 0 otherwise. This represents our primary explanatory variable. Where the deadline is 12 months, for instance, then approvals in the 11th and 12th month after submission are coded as 1. This variable appears in tables 4 through 7 as “approved within 2 months of deadline.”

2. Agency Uncertainty: Drug Priority and Novelty (Log Order of Drug Entry). To assess claims that lower quality policy outcomes may arise from inherent drug uncertainty, we include two measures to capture aspects of drug uncertainty. For one, we include an indicator representing whether the FDA designated the drug as a priority or standard drug review.^{xxii} The FDA identifies drugs that represent significant innovation over existing therapies for priority review. Such

innovation holds potential for greater drug safety and efficacy, and yet also greater uncertainty as new formulations are introduced into the market for the first time. We also measure drug novelty by taking the log of the number of drugs previously approved to treat the drug's primary indication. The more drugs to treat a particular disease, the more experience the FDA may have with therapies to treat that disease and possibly less uncertainty over how that drug will behave in the market.^{xxiii} Fewer drugs on the market to treat a disease may capture potential uncertainty about how the therapy will perform in patient populations outside of experimental conditions.

3. Count of Previous Firm NME Approvals. To assess firm effects on drugs' post-marketing experiences and test claims that post-marketing problems arise from established firms' relationship with regulators, we include a measure of firm experience, derived from a count of the number of NMEs that a sponsoring firm has had previously approved.

The most basic, "baseline" models we estimate include only a constant term, the year of submission, and the pre-deadline approval indicator. The expanded models we report include these variables plus the count of firms' previous drug approvals, the measure of NME novelty, and the indicator of priority review. The baseline models are regressions on all NMEs approved from 1950 through 2008. We report results for the expanded models for NMEs approved between 1980 and 2008, and then for NMEs approved between 1993 and 2008.

[Tables 4, 5, 6 and 7 about here.]

Results: GLM Estimation. We report the generalized linear model results in Tables 4 through 7. We observe that our measure of the deadline-induced change in FDA approval behavior – approval within the month of or the month before a deadline – is positively related to safety alerts, safety-based withdrawals, new black-box warnings, and sometimes to dosage form discontinuations at standard levels of statistical significance.^{xxiv}

The results in the first columns of tables 4, 5, and 7 report logit coefficients rather than readily interpreted quantities of interest. For this reason, the second column of results in these tables reports the odds ratio for each covariate. We see for instance that the probability of a drug approved in the month of or the month before a deadline receiving a new black box warning is 3.82 times greater for an at-deadline approval, according to the full model for NMEs approved between 1980 and 2008, and 3.28 times greater according to the full model specified in the NMEs approved after 1992 [test of H2].

Drugs approved shortly before the FDA deadline elapses also appear more likely to experience safety-based withdrawals. For NMEs approved between 1980 and 2008, the probability of a post-market withdrawal is 8.66 times higher for drugs approved in the two months before a deadline than for those approved at other times. And for NMEs approved between 1993 and 2008, the full model indicates that the probability of a safety-based withdrawal is 12.24 times greater for a drug approved in the two months leading up to its approval deadline than for drugs approved at other times [test of H2]. The difference in the probability of safety-based withdrawals for drugs approved within two months of their approval deadlines, and drugs approved at other times, is largest for when the sample of NMEs approved is restricted to the post-PDUFA era.

The odds ratios for dosage form discontinuations also support this conclusion. For NMEs approved after 1979, the probability of a dosage form discontinuation is 1.83 times greater for drugs approved in the two months leading up to their approval deadlines than for drugs approved at other times. Drugs approved between 1993 and 2008 were 1.96 times more likely incur a dosage form discontinuation [test of H2]. We also see a consistently positive, statistically significant relationship between safety alerts and approval in the two months approaching a deadline [test of H2]. On the whole, these results suggest that we cannot reject our hypothesis that changes in the timing of FDA

decisions – resulting in drug piling around the specified deadline – are associated with greater drug post-marketing regulatory events.

Our analyses yield more mixed results for alternative explanations: that greater post-marketing problems stem either from inherent drug uncertainty (reflected in our measures of drug novelty and drug priority) or from firm privilege. Drug priority status is not associated with greater post-marketing regulatory events for any of the safety outcomes we consider. Moreover, contrary to predictions about inherent therapeutic uncertainty, having more drugs on the market to treat a particular disease is associated with a greater likelihood of post-marketing black box warnings. Drugs submitted by more established firms appear more likely to experience safety alerts at standard levels of statistical significance.

Matching Analyses. We also seek to minimize the possibility that our results are dependent on modeling assumptions as a consequence of large differences between our implicit treatment group (approval in the two months before a looming deadline) and our implicit control group (drugs approved at other times). For a variety of reasons related to non-random assignment, these groups may not be equivalent in all relevant aspects except for the just-before-deadline approval.^{xxv} There is, to be sure, no way of eliminating non-random assignment as an explanation for our results, which are generated from an observational research design, especially when accounting for the linear effect of a small number of covariates. To reduce the likelihood that our results are due to baseline differences between the treatment and control groups, we match observations from the treatment and control groups on observables (e.g., Hansen 2004, Iacus, King, and Porro 2009) to construct a sample that allows estimation of the causal effect of the treatment on the population represented by the matched group.

We use two matching methods: optimal matching (Hansen 2004) and coarsened exact matching (Iacus, King, and Porro, 2009).^{xxvi} Using these methods, we match on a range of covariates

widely believed to predict pre-deadline approval.^{xxvii} We examine issues of covariate balance and present graphical results in the Online Appendix.

[Table 8 about here.]

Results from optimal matching and CEM appear in Table 8. Though they apply to different target populations, the results are consistent with our original GLM analysis, though they indicate larger uncertainty around some estimates than results obtained without matching. Given the differences in target populations for the two methods, we present the results under each method separately.

Optimal matching followed by modeling largely supports findings on post-market regulatory events produced from the non-matched sample. Among drugs approved during a looming deadline, the deadline can be expected to increase the chance of a safety-related withdrawal by a factor of 2.0 (95% CI: [1.3, 21]). The probability of a black-box warning increases by a factor of 1.8 (95% CI: [1.5, 14.4]), drug-form discontinuation by a factor of 1.4 (95% CI: [0.9, 3.3]). The count of safety alerts can be expected to increase by a factor of 1.2 (negative binomial model 95% CI: [0.9, 2.2]).

Coarsened exact matching, which permits inference to a population represented by the matched sample, also yields evidence supportive of deadlines' effect on post-market regulatory events, but with less certainty in some PMRE categories. Among drugs approved during a looming deadline, the deadline can be expected to increase the chance of a safety-related withdrawal by a factor of 2.7 (95% CI: [0.3, 14]). The probability of a black-box warning increases by a factor of 2.9 (95% CI: [0.4, 26]) and the probability of drug-form discontinuation by a factor of 1.9 (95% CI: [0.9, 11]). The count of safety alerts can be expected to increase by a factor of 1.6 (negative binomial model, 95% CI: [0.3, 2.2]).

The Human Impact of Postmarket Safety Issues – Examination through Correlates of PMREs with Adverse Event Reports. We have already adduced evidence for the hypothesis

that at-deadline approvals are associated with significant market and medical outcomes. At-deadline approvals are more likely to be withdrawn or discontinued from the marketplace and are associated with a higher rate of post-approval labeling revision, including for black-box warnings. Whether lives and deaths and other human health outcomes are affected is a different question. Some information from the plausible human health effects of these dynamics can be estimated from examining adverse drug reactions that are correlated without postmarket safety outcomes. If the average drug withdrawal is associated with 50,000 deaths, say, then a small increase in the probability of a drug withdrawal may be expressive of an appreciable number of lives lost or damaged. Using a database of all reported adverse drug reactions from 1996 to 2006, we estimate the average change in reported adverse drug events associated with two of our post-market regulatory outcomes – drug withdrawal and safety alerts.

The database, collected from the FDA's MEDWATCH system, classifies drug adverse events into a number of categories, but three are of interest here: (1) adverse events associated with patient's subsequent death, (2) adverse events associated with a patient's subsequent hospitalization, and (3) adverse events associated with a patient's subsequent disability. Each safety-based drug withdrawal is associated with 30,621 additional adverse events associated with death, 164,183 additional adverse drug events associated with hospitalization, and 14,407 additional adverse events associated with a reported disability. If we take a one-unit change in the natural logarithm of safety alerts as a predictor for adverse events, such a change is associated with 13,191 more death-related adverse events, 44,181 hospitalization-related adverse events, and 6,988 disability-related adverse drug reactions.

For a number of reasons, these numbers must be taken as suggestive only of possibly correlated effects of safety-based withdrawals and safety alerts. One problem is that there is massive underreporting of adverse events, such that many pharmacologists and drug safety experts see them

as only indicative of drug safety problems. Another problem is that one can never be certain that a drug caused a death or hospitalization in a given case; the MEDWATCH system is based upon reporting of physicians' and manufacturers' suspected associations of an adverse event with use of a particular drug. For this reason, we do not attempt here to causally tie at-deadline approvals to higher rates of adverse events. These correlates, however, do suggest that post-marketing regulatory events such as drug withdrawals and safety alerts are not merely social and economic outcomes. They reflect real human outcomes, in that they are correlated with induced hospitalizations, induced disabilities and deaths.

Discussion

The puzzle of whether and when instruments of political control affect the quality of agency outcomes has motivated scholarship for decades. Yet, few have put this puzzle to a rigorous empirical test. Our results suggest that deadline institutions influence the timing of agency decisions – resulting in drug piling around the deadline – and that these deadline-induced behavior changes are associated with different policy outcomes. Rather than consider aggregate effects (i.e. does the introduction of deadlines affect overall policy quality), we thus offer a refined approach that links specific deadline-induced changes in agency behavior (drug piling) with ultimate policy consequences. Taking this more refined approach provides a window into the mechanism through which deadlines may affect policy outcomes. By linking specific behavior changes with changes in policy quality, we are able to discern that deadlines do not appear to affect the quality of all drugs. Instead, the drugs approved just before the deadline elapses are the ones more likely to experience post-marketing problems. By assessing policy outcomes, our analyses offer important extensions to scholarship that focuses solely on agency decisions. Moreover, our analysis suggests how institutional choices can affect agency decision criteria in ways that subordinate other agency considerations, such as reputation for product quality.

We note that our results are not uniform, and that some vary by model specification (hence our reporting of multiple specifications of the GLMs) and under different matching specifications. However, we observed no statistically significant *negative* relationships between deadline approvals and post-marketing regulatory events (PMREs) across all of the robustness checks done on the results reported here. Moreover, we *do* observe a large number of positive relationships statistically significant at the 5% level between deadline approvals and PMREs that hold across statistical specifications. While it is undoubted that detection methods for adverse post-market events and other safety issues have increased in the past two decades, the starkness of “deadline effects” suggests that broader trends in pharmacoepidemiology are unlikely to have contributed to the positive association between approval under a looming deadline and adverse post-market events.

For all four of the regulatory events modeled, our analyses suggest a positive relationship between approval in the two months leading up to a deadline and adverse post-market events. The results suggest then, that the rate at which drugs experience post-marketing regulatory events is appreciably higher for drugs approved in the months before the PDUFA clock deadlines, compared to other drugs. The finding of an increase in safety-based withdrawals, in particular, contrasts with other findings in the literature which reported drug safety withdrawals within the United States unrelated to the speed of FDA drug approval (e.g., Tufts Center for the Study of Drug Development, 2005). While psychology research (e.g. Huber and Kunz, 2007) suggests that deadlines would lead to stronger scrutiny of drugs, the findings here suggest reason to doubt this hypothesis.

Our findings offer useful implications for drug regulation, and regulation more generally, both within the world drug regulation system and in other regulatory agencies. We have focused the present analysis upon new molecular drug approval in the United States. The drug reviews of the European Medicines Evaluation Agency (EMA) are governed by a system of review-time

deadlines, and thousands of other FDA reviews not analyzed here (of medical devices, biologic drugs and generic drugs, among others) are governed by deadlines nearly identical to those examined here. There are, furthermore, many administrative and regulatory decisions characterized by timing phenomena (O’Connell and Gerson 2009; Yackee and Yackee 2009). Wherever deadlines apply to administrative timing processes in ways that are implicit (deadlines implemented as goals or soft constraints) or explicit (explicitly mandated and absolute statutory deadlines), our models will be of potential analytic value.

Moreover, the mechanisms through which deadlines impair policy outcomes – by limiting the amount of time and information an agency may bring to bear on decisions, by conveying political information to the agency, by restructuring agency decision criteria – manifest in other forms of political control. Appointments, for instance, can change the information an agency has at its disposal as well as alter the agency’s criteria and culture. Budgets, for their part, embody concrete resources and political symbols. The fundamental similarity in the mechanisms that enable instruments of political control give us confidence that our deadline results apply broadly to other institutions designed to shape and constrain bureaucratic behavior.

Our models offer compelling evidence that political control through regulatory deadlines affects the timing of agency decisions and the quality of their outcomes. We, thus, offer an important empirical extension to scholarship that predicts a tradeoff between political control and agency expertise (Bawn 1995). According to Bawn, Congress’ motive in promulgating controls is often specifically to “enfranchise some groups and exclude others,” or to delimit “who has access to the agency” (1995: 62). Deadlines may privilege firm information over other information sources, for instance. Yet, our nuanced story of piled approval decisions and more troubled post-marketing experiences suggests future scholarship must address several questions on firm privilege. While firms may benefit from faster approval times, do they benefit more (or less) from approvals that do

not occur right around the deadline window? Moreover, to what extent do the costs of more troubled post-marketing experiences erode the benefits of faster approval times? Though future scholarship will address these questions explicitly, we can currently observe that winners and losers in the empirical story we present are not as straightforward as a stylized model might imply.

We find consistent results across deadline regimes, over time. Future research, however, should consider two additional aspects of agency learning. First, when agencies develop a reputation for meeting the ‘spirit’ of the deadline (i.e. getting therapies on the market more quickly), are they better able to violate the specific terms of the deadline without political consequence? As agencies and political overseers learn, in other words, do deadlines become less ‘absolute?’ Second, as agencies adapt to deadline regimes, do they learn how to adjust their work processes to afford *both* product quality and attentiveness to meeting the deadline? These future lines of inquiry will help illuminate the durability of deadline institutions and their consequences.

Considerable scholarship over the past twenty years has explored the range of tools elected officials have at their disposal to shape bureaucratic behavior, thus eroding the image of a runaway or impervious bureaucracy. Scholarship has attended much less systematically to the consequences of such political control. Our study offers an important contribution to this ongoing discussion by demonstrating how one largely overlooked instrument – deadlines – shapes both agency actions and the quality of subsequent policy outcomes.

References

- Abbott, Alden. 1987. “Case Studies on the Costs of Federal Statutory and Judicial Deadlines.” *Administrative Law Review* 39:467.
- Ando, Amy Whitenour. 1999. “Waiting to be Protected under the Endangered Species Act: The Political Economy of Regulatory Delay.” *Journal of Law and Economics*, 42: 29-60.

- Balla, Steven J., and John R. Wright. 2001. "Interest Groups, Advisory Committees, and Congressional Control of the Bureaucracy." *American Journal of Political Science*, 45(4): 799-812.
- Bawn, Kathleen. 1995. "Political Control versus Expertise: Congressional Choices about Administrative Procedures." *American Political Science Review*, 89(1): 62-73.
- Berlin, Robert J. 2009. "Examination of the Relationship Between Oncology Drug Labeling Revision Frequency and FDA Product Categorization," *American Journal of Public Health* 99 (Supplement 2) (Publication Date: September 2, 2009) 1-6.
- Bohte, John. 2001. "School Bureaucracy and Student Performance at the Local Level." *Public Administration Review*, 61(1): 92-99.
- Carpenter, Daniel P. 1996. "Adaptive Signal Processing, Hierarchy, and Budgetary Control in Federal Regulation." *American Political Science Review* 90(2): 283-302.
- Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton: Princeton University Press.
- Carpenter, Daniel P. 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval." *American Journal of Political Science*, 46(3): 490-505.
- Carpenter, Daniel P., A. Mark Fendrick, Michael Chernew, and Dean Smith. 2003. "Approval Times For New Drugs: Does The Source Of Funding For FDA Staff Matter?" *Health Affairs* [Web Exclusive], W3-618-624.
- Carpenter, Daniel P. and Michael Ting. 2007. "Regulatory Errors with Endogenous Agendas," *American Journal of Political Science* 51(4): 835-852.
- Carpenter, Daniel P., Evan James Zucker, Jerry Avorn. 2008. "Drug-Review Deadlines and Safety Problems." *New England Journal of Medicine*, 358(13): 1354-1361.
- Carpenter, Daniel P. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation in the FDA*. Princeton: Princeton University Press.

- Fershtman, Chaim and Daniel Seidmann. 1993. "Deadline Effects and Inefficient Delay in Bargaining with Endogenous Commitment." *Journal of Economic Theory* 60:306-321.
- Food and Drug Administration. 2007. *Manual of Policies and Procedures: Center for Drug Evaluation and Research* MAPP 6020.3.
- Furlong, Scott R. 1998. "Political Influence on the Bureaucracy: The Bureaucracy Speaks." *Journal of Public Administration Research and Theory*, 8(1): 39-65.
- Gerson, Jacob and Anne O'Connell. 2008. "Deadlines in Administrative Law," *University of Pennsylvania Law Review* 156: 923.
- Gneezy, Uri et al. 2003. "Bargaining Under a Deadline: Evidence from the Reverse Ultimatum Game," *Games and Economic Behavior* 45:347.
- Hansen, Ben B. 2004. "Full Matching in an Observational Study of Coaching for the SAT," *Journal of the American Statistical Association*, 99 (467) (September 2004): 609-618.
- Heimann, C.F. Larry. 1997. *Acceptable Risks: Politics, Policy and Risky Technologies*. Ann Arbor, MI: University of Michigan Press.
- Ho, Daniel and Kosuke Imai and Gary King and Elisabeth Stuart. 2007. "Matching as Nonparametric Preprocessing to Reduce Model Dependence in Parametric Causal Inference." *Political Analysis*, 15: 199-236.
- Huber, Gregory A. 2007. *The Craft of Bureaucratic Neutrality: Interests and Influence in Governmental Regulation of Occupational Safety*. New York: Cambridge University Press.
- Huber, John and Charles R. Shipan. 2002. *Deliberate Discretion: The Institutional Foundations of Bureaucratic Autonomy*. New York: Cambridge University Press.
- Huber, Oswald and Urs Kunz. 2007. "Time Pressure in Risky Decision-making: Effect on Risk Defusing." *Psychology Science*, 49(4): 415-426.

- Iacus, Stefano, Gary King, and Giuseppe Porro. 2010. Multivariate Matching Methods That Are Monotonic Imbalance Bounding.” Unpublished.
- Kosnik, Lea-Rachel D. 2006. “Sources of Bureaucratic Delay: A Case Study of FERC Dam Relicensing.” *Journal of Law, Economics and Organization*, 22(1): 258-288.
- Krause, George A. 2003. “Coping with Uncertainty: Analyzing Risk Propensities of SEC Budgetary Decisions, 1949-97.” *American Political Science Review*, 97(1): 171-188.
- Krause, George A., and J. Kevin Corder. 2007. “Explaining Bureaucratic Optimism: Theory and Evidence from U.S. Executive Agency Macroeconomic Forecasts.” *American Political Science Review*, 101(1): 129-142.
- Krause, George A., and James W. Douglas. 2006. “Does Agency Competition Improve the Quality of Policy Analysis? Evidence from OMB and CBO Fiscal Projections.” *Journal of Policy Analysis and Management*, 25(1): 53-74.
- Lasser, Karen E., Paul D. Allen, Steffie J. Woolhandler, David U. Himmelstein, Sidney M. Wolfe, David H. Bor. 2002. “Timing of New Black Box Warnings and Withdrawals for Prescription Medications.” *Journal of the American Medical Association*, 287(17): 2215-2220.
- Lewis, David E. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton: Princeton University Press.
- Lexchin, Joel. 2005. “Drug Withdrawals from the Canadian Market for Safety Reasons, 1963–2004,” *Canadian Medical Association Journal*, 172(6): 765-767.
- Mashaw, Jerry and David Harfst. 1990. *The Struggle for Auto Safety*. Harvard University Press.
- McCullagh, Peter and J.A. Nelder. 1989. *Generalized Linear Models*. London: Chapman and Hall.
- McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast. 1987. “Administrative Procedures as Instruments of Political Control.” *Journal of Law, Economics, & Organization*, 3(2): 243-277.

- Moe, Terry M. 1987. "An Assessment of the Positive Theory of 'Congressional Dominance.'" *Legislative Studies Quarterly*, 12(4): 475-520.
- Moffitt, Susan L. 2010. "Promoting Agency Reputation through Public Advice: Advisory Committee Use in the FDA," *Journal of Politics* (72) 3: 880-893.
- Morgenstern, Richard. 1993. "Science, Engineering, and Regulation," in *Keeping Pace with Science and Engineering*. (Washington: National Academic Press), pp. 243-250.
- Nardinelli, Clark, Michael Lanthier and Robert Temple; Daniel Carpenter. 2008. "Drug Review Deadlines and Safety Problems [Correspondence]," *New England Journal of Medicine* 359 (July 3, 2008) 95-98.
- Olson, Mary K. 1997. "Firm Characteristics and the Speed of FDA Approval." *Journal of Economics and Management Strategy*, 6: 377-401.
- Olson, Mary K. 2000. "Regulatory Reform and Bureaucratic Responsiveness to Firms: The Impact of User Fees in the FDA." *Journal of Economics and Management Strategy*, 9: 363-95.
- Olson, Mary K. 2002. "Pharmaceutical Policy Change and the Safety of New Drugs." *Journal of Law and Economics*, 45(S2): 615-642.
- Olson, Mary K. 2004a. "Are Novel Drugs More Risky for Patients than Less Novel Drugs?" *Journal of Health Economics*, 23(6): 1135-1158.
- Olson, Mary K. 2004b. "Managing Delegation in the FDA: Reducing Delay in New-Drug Review." *Journal of Health Politics, Policy, and Law*, 29(3): 397-430.
- Olson, Mary K. 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 2008): 175-200
- Scholz, John T., Jim Twombly, and Barbara Headrick. 1991. Street-Level Political Controls Over the Bureaucracy. *American Political Science Review* 85 (3): 829-850.

- Smith, Kevin B., and Christopher W. Larimer. 2004. "A Mixed Relationship: Bureaucracy and School Performance." *Public Administration Review*, 64(6): 728-736.
- Spence, David B. 1999a. "Agency Discretion and the Dynamics of Procedural Reform." *Public Administration Review*, 59(5): 425-442.
- Spence, David B. 1999b. "Managing Delegation Ex Ante: Using Law to Steer Administrative Agencies." *Journal of Legal Studies*, 28: 413-459.
- Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag.
- Tufts Center for the Study of Drug Development. 2005, September/October. "Drug Safety Withdrawals in the U.S. Not Linked to Speed of FDA Approval." *Tufts CSDD Impact Report*.
- United States Department of Health and Human Services Office of the Inspector General 2002. *FDA Review Process for New Drug Applications*. OEI-01-01-00590. Washington, DC: Department of Health and Human Services
- Whitford, Andrew B. 2005. "The Pursuit of Political Control by Multiple Principles." *The Journal of Politics*, 67(1): 29-49.
- Yackee, Jason Webb, and Susan Webb Yackee. 2009. "Administrative Procedures and Bureaucratic Performance: Is Federal Rule-Making 'Ossified?'" *Journal of Public Administration Research and Theory*: 1-22.

Figure 1: Approval Hazard Ratios for Priority NMEs before and after Deadlines

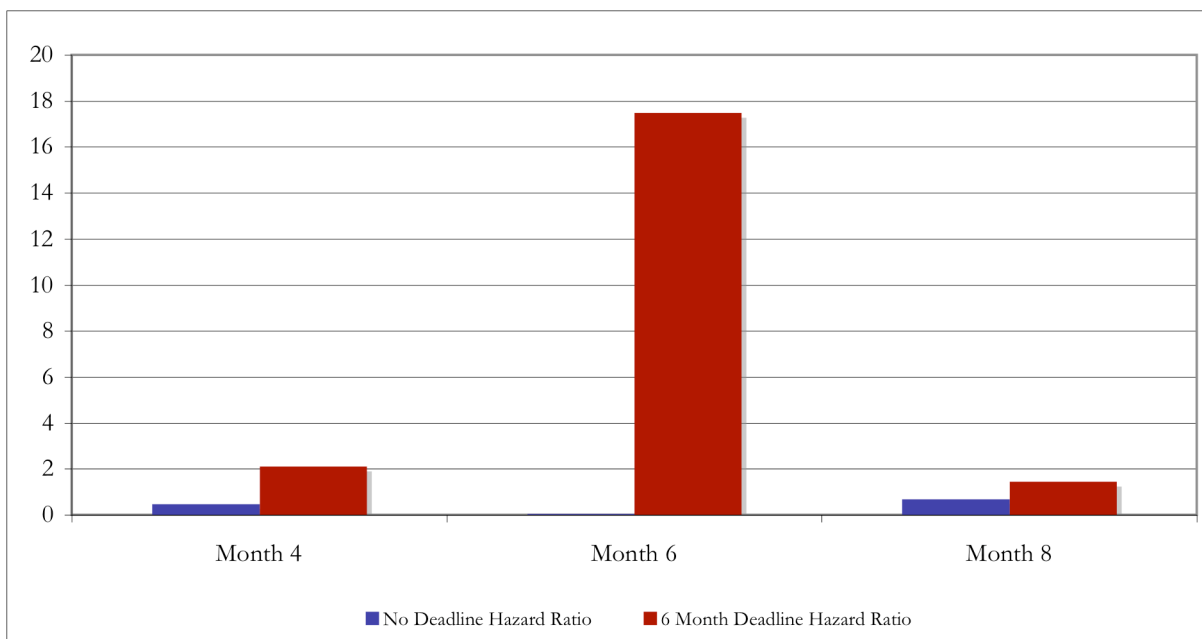


Table 1: Priority Drug Hazard Ratios, Before and After Deadlines

Variable	No-Deadlines Hazard Ratio (SE)	6 Month Deadline Hazard Ratio (SE)
Year of Drug Submission	0.965 (0.009)	0.965 (0.009)
Staffing Levels	1.001 (0.0002)	1.001 (0.0002)
Month 1	1.056 (1.233)	0.947 (1.106)
Month 2	2.316 (1.804)	0.432 (0.336)
Month 3	1.352 (0.910)	0.739 (0.498)
Month 4	0.471 (0.237)	2.121 (1.067)
Month 5	0.854 (0.532)	1.171 (0.730)
Month 6	0.057 (0.024)	17.479 (7.265)
Month 7	0.090 (0.041)	11.108 (5.011)
Month 8	0.684 (0.398)	1.462 (0.850)
Month 9	0.103 (0.049)	9.718 (4.590)
Month 10	0.467 (0.275)	2.139 (1.258)
Month 11	1.167 (1.248)	0.857 (0.916)
Month 12	0.155 (0.083)	6.467 (3.453)

Number of drugs: 484; Number of drug months: 9171. Bold-type connotes statistical significance at the $p < .05$ level (all tests are two-tailed); + indicates significance at the $p < .1$ level. Table does not include firm controls and months 13-24. Results available from authors.
Non-Linear Likelihood test comparing months 4 & 6; $\chi^2=11.88$, $p=0.0006$; comparing months 6 & 8: $\chi^2=13.18$; $p=0.0003$

**Figure 2: Hazard Ratios for Standard Drugs,
Before Deadlines and During Different Deadline Regimes**

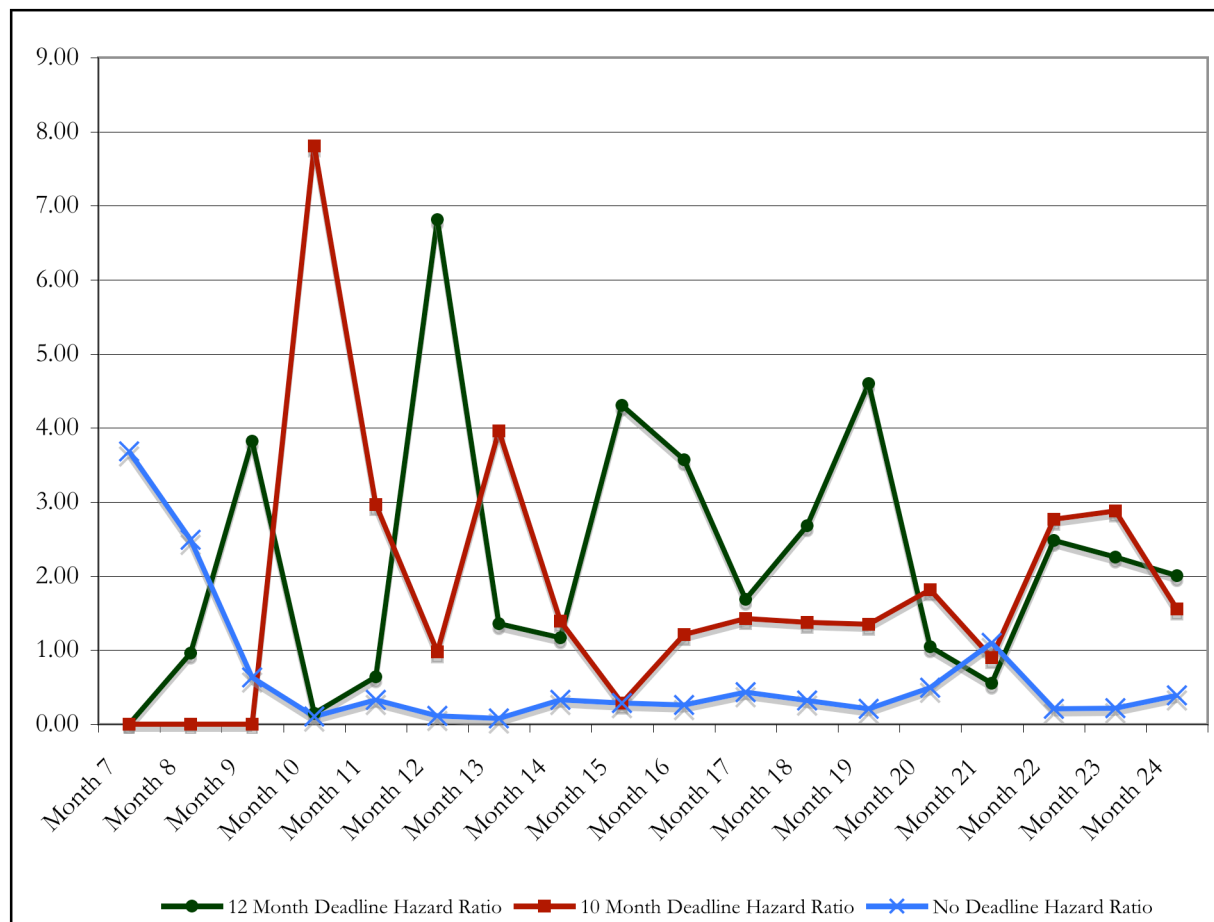


Table 2. Hazard Ratios for Standard Drugs

Variable	No Deadline	12 Month Deadline (PDUFA)	10 Month Deadline (FDAMA)
Year of Drug Submission	0.943 (0.009)	0.943 (0.009)	0.943 (0.009)
CDER Staff Levels	1.001 (0.0002)	1.001 (0.0002)	1.001 (0.0002)
Month 8	2.492 (2.805)	0.799 (0.899)	0
Month 9	0.633 (0.524)	3.166 (2.614)	0
Month 10	0.100 (0.056)	0.401 (0.428)	7.774 (3.451)
Month 11	0.327+ (0.191)	1.062 (0.876)	3.508 (2.166)
Month 12	0.117 (0.051)	11.704 (5.226)	3.804 (2.166)
Month 13	0.082 (0.053)	2.949+ (1.931)	6.363 (3.526)
Month 14	0.326 (0.172)	1.497 (1.011)	1.793 (1.215)
Month 15	0.286 (0.132)	4.531 (2.134)	0.538 (0.568)
Month 16	0.257 (0.132)	4.860 (2.749)	2.279 (1.596)
Month 17	0.435 (0.239)	2.197 (1.504)	1.991 (1.369)
Month 18	0.318+ (0.204)	3.570+ (2.631)	2.209 (1.867)
Month 19	0.204 (0.133)	6.476 (4.624)	2.753 (2.408)

Number of drugs: 550; Number of drug months: 14, 562. Bold-type connotes statistical significance at the $p < .05$ level (all tests are two-tailed); + indicates significance at the $p < .1$ level. Table does not include firm controls and months 1-7, 20-24. Results available from authors.
 Non-Linear Likelihood test comparing months 10 for PDUFA & FDAMA: $\chi^2=8.22$, $p=.004$; comparing months 12 for PDUFA & FDAMA: $\chi^2=5.73$; $p=.02$; comparing months 12 & 14 for PDUFA: $\chi^2=6.68$, $p=0.01$; comparing months 10 & 12 for FDAMA: $\chi^2=1.04$, $p=0.309$

Table 3: Summary Statistics for Post-Marketing Regulatory Event (PMRE) Variables

STANDARD NMEs	Valid NMEs	Mean	Std. Dev.	Minimum	Maximum
Safety-Based Withdrawal	340	0.021	0.142	0	1
Black-Box Warning Addition	340	0.029	0.169	0	1
Safety Alert (1996-2008)	133	0.714	1.132	0	7
Dosage Form Discontinuations	340	0.332	0.472	0	1
PRIORITY NMEs	Valid NMEs	Mean	Std. Dev.	Minimum	Maximum
Safety-Based Withdrawal	286	0.017	0.131	0	1
Black-Box Warning Addition	286	0.038	0.192	0	1
Safety Alert (1996-2008)	114	0.842	1.537	0	8
Dosage Form Discontinuations	286	0.290	0.455	0	1

Table 4: Black Box Warnings, Random Effects Logistic Regressions

Variable	NMEs Approved Post 1979		NMEs Approved Post 1992	
	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)
Pre-Deadline Approved within 2 Months of Deadline	1.34 (0.53)	3.82 (2.04)	1.19 (0.49)	3.28 (1.62)
Firms Firms' Previous Drug Approvals	0.03 (0.02)	1.03 (0.02)	0.03+ (0.02)	1.03+ (0.02)
NME Novelty Log Order of Drug Market Entry	0.43 (0.21)	1.53 (0.32)	0.45 (0.20)	1.57 (0.32)
Drug Priority Priority Review = 1	0.25 (0.51)	1.29 (0.66)	0.53 (0.50)	1.69 (0.85)
Time Year of Submission	0.10 (0.05)	1.11 (0.05)	-0.05 (0.07)	0.95 (0.07)
Constant	-212.07 (90.49)		115.52 (138.16)	
Number of Primary Indications		187		137
NMEs (number of observations)		626		337
Log Likelihood		-76.35		-70.37

Note: Bold-type connotes statistical significance at the $p < .05$ level (all tests are two-tailed); + indicates significance at the $p < .1$ level

Table 5: Safety-Based Withdrawals, Random Effects Logistic Regressions

Variable	NMEs Approved Post 1979		NMEs Approved Post 1992	
	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)
Pre-Deadline Approved within 2 Months of Deadline	2.16 (0.76)	8.66 (6.58)	2.5 (0.87)	12.24 (10.67)
Firms Firms' Previous Drug Approvals	0.04 (0.02)	1.04 (0.03)	0.03 (0.03)	1.03 (0.03)
NME Novelty Log Order of Drug Market Entry	-0.01 (0.32)	0.99 (0.31)	0.13 (0.43)	1.14 (0.48)
Drug Priority Priority Review = 1	-0.49 (0.71)	0.61 (0.44)	0.01 (0.72)	1.29 (1.14)
Time Year of Submission	0.04 (0.06)	0.96 (0.04)	-0.23+ (0.14)	0.79+ (0.11)
Constant	-77.29 (117.19)		456.56+ (270.19)	
Number of Primary Indications		187		137
NMEs (number of observations)		626		337
Log Likelihood		-50.26		-40.14

Note: Bold-type connotes statistical significance at the p<.05 level (all tests are two-tailed); + indicates significance at the p<.1 level.

Table 6: Safety Alerts, Random Effects Negative Binomial Regressions

Variable	NMEs Approved Post 1995	
	Coefficient	(SE)
Pre-Deadline Approved within 2 Months of Deadline	0.50	(0.20)
Firms Firms' Previous Drug Approvals	0.01+	(0.01)
NME Novelty Log Order of Drug Market Entry	0.15	(0.16)
Drug Priority Priority Review = 1	0.09	(0.24)
Time Year of Submission	-0.13	(0.04)
Constant	267.80	(74.11)
Number of Primary Indications		112
NMEs (number of observations)		247
Log Likelihood		-272.52

Note: Bold-type connotes statistical significance at the p<.05 level (all tests are two-tailed); + indicates significance at the p<.1 level

Table 7: Dosage Form Discontinuations, Random Effects Logistic Regressions

Variable	NMEs Approved Post 1979		NMEs Approved Post 1992	
	Coefficient	Odds Ratio	Coefficient	Odds Ratio
	(SE)	(SE)	(SE)	(SE)
Pre-Deadline Approved within 2 Months of Deadline	0.60+ (0.32)	1.83+ (0.58)	0.67 (0.33)	1.96 (0.65)
Firms Firms' Previous Drug Approvals	0.01 (0.01)	1.01 (0.01)	0.01 (0.01)	1.01 (0.01)
NME Novelty Log Order of Drug Market Entry	-0.01 (0.08)	0.99 (0.08)	0.11 (0.12)	1.12 (0.14)
Drug Priority Priority Review = 1	-0.31 (0.20)	0.73 (0.15)	-0.04 (0.32)	0.96 (0.31)
Time Year of Submission	-0.13 (0.02)	0.88 (0.01)	-0.15 (0.05)	0.87 (0.04)
Constant	260.25 (30.85)		294.26 (95.4)	
Number of Primary Indications		187		137
NMEs (number of observations)		626		337
Log Likelihood		-341.56		-148.40

Note: Bold-type connotes statistical significance at the p<.05 level (all tests are two-tailed); + indicates significance at the p<.1 level

Variable	Safety Withdrawal (Logistic)		Black-Box Warning (Logistic)		Safety Alerts (Neg. Binomial)		Discontinuation (Logistic)	
	Optimal	CEM	Optimal	CEM	Optimal	CEM	Optimal	CEM
Constant	383+ (220)	-178 (350)	170 (160)	2.79 (394)	164 (63.7)	187 (177)	292 (103)	559 (261)
Approved In Two Months Before Deadline	1.65 (0.712)	0.703 (0.985)	1.53 (0.58)	1.17 (1.07)	0.368+ (0.221)	-0.138 (0.470)	0.524 (0.338)	1.16+ (0.642)
Year of NME Submission	-0.194+ (0.11)	0.088 (0.175)	-0.088 (0.080)	-0.004 (0.197)	-0.082 (0.032)	-0.094 (0.089)	-0.147 (0.052)	-0.281 (0.131)
Log(Order of Entry in Disease Niche)	0.093 (0.292)	-0.456 (0.554)	0.459+ (0.238)	0.087 (0.509)	0.155+ (0.09)	0.375 (0.238)	0.106 (0.14)	-0.048 (0.344)
Priority Review	0.230 (0.696)	-0.661 (1.124)	1.160+ (0.608)	0.025 (1.17)	0.140 (0.223)	0.047 (0.581)	0.234 (0.344)	-0.648 (0.849)
Unique NME Submissions By Firm	0.017 (0.024)	0.130 (0.07)	0.027 (0.0185)	0.164 (0.076)	0.025 (0.008)	0.077 (0.038)	0.000 (0.013)	-0.009 (0.062)
NMEs (number of observations)	258	125	258	125	258	125	258	125

Note: Bold-type indicates $|Z| > 2.0$. + indicates $|Z| > 1.7$.

ⁱ The analyses here are different and much more comprehensive than those that appear in Carpenter, Zucker and Avorn (2008), which has provoked debate among political scientists and FDA officials (see Nardinelli et. al. 2008). First, our sample covers a longer period. Our analysis of the timing of drug approvals considers drugs from 1943 through 2008. Our analysis of postmarket outcomes generally considers the period from January 1980 to December 2008. Carpenter et. al. focus on a narrow set of drugs submitted from January 1993 to December 2004. Second, our models include controls for the therapeutic priority status of the drugs reviewed (“priority” versus “standard” reviews); observers, including FDA officials themselves, have previously argued that priority drugs account for both quicker reviews and greater post-market safety issues (see, for example, Berlin 2009). Third, compared to previous studies, we conduct much more complex and systematic statistical analyses of the association between just-before-deadline approvals and postmarket safety issues, using generalized linear models with random effects, as well as other methods of observational causal inference (optimal and coarsened exact matching) to assess the robustness of causal inferences concerning deadline institutions and regulatory behavior.

ⁱⁱ Other FDA decision processes include the regulation of medical devices, biologic drugs and generic drugs. Replication of these analyses in other contexts would require (1) data on the duration of administrative processes, (2) an instance in which deadlines were imposed upon the policymaker or agency, and (3) data on the “outcomes” or post-decision quality of the case.

ⁱⁱⁱ Deadlines are also proposed in other cases of administrative decision making, such as the granting of visas or the examination and review of entrants to a country by a customs service. Business interests in the U.S. and Canada have been pressing for review time goals for border crossing agents, and non-profit groups have constructed an archive of border-time wait statistics to buttress these arguments and to call for more research;

see <http://www.wcog.org/Border/IMTC-Projects/Border-Wait-Time-Archive/68.aspx>

and <http://www.cascadegatewaydata.com> (both accessed April 3, 2010). In other cases, federal agencies have sponsored these studies; see http://www.fhwa.dot.gov/HEP/STEP/success_bordr.htm (accessed April 3, 2010).

^{iv} Cited in Bawn (1995: 63). Procedural controls, in general, are particularly strong and enduring when they alter agency preferences about decision outcomes (Spence, 1999b: 413).

^v Scholarship on bureaucratic outputs has also largely neglected these issues, but see Olson (2008) for an exception. Her study, however, examines drug safety in relation to *speed* of decision time but not the deadline instrument itself. Some work examines whether redundant, competing agencies collectively exhibit better performance than single agencies with monopolies on expertise in their respective areas (e.g. Krause and Douglas, 2006). Other work considers the relationship between school bureaucracy and student performance (Bohte, 2001; Smith and Larimer, 2004). Most analyses of deadlines in the formal modeling literature in their empirical tests examine bargaining with deadlines – of the sort that occurs when a labor union and company management must reach an agreement by a certain date in order to avoid a walkout or mass layoffs (Fershtman and Seidman 1993; Gneezy et al 2003). These models have limited applicability for situations of policy and politics characterized by uncertainty and discretionary choice.

^{vi} Considerable scholarship has demonstrated the importance of reputation as a motivating factor in bureaucratic decisions (Wilson 1989; Carpenter 2001, 2002, 2004, 2010; Krause and Corder 2007; Moffitt 2010).

^{vii} Drug review and marketing approval by the FDA, moreover, represent one of the most consequential (and controversial) regulatory policies of our time. The agency has been lambasted repeatedly by those who fear that its attention to safety is too lax, and has been excoriated by those who feel that insufficient weight is placed upon patient access to new medicines and the benefits of market access for pharmaceutical companies. We do not intrude into this debate except to note two points. First, much of the debate over has been a debate about the *timing* of FDA decision processes. Second, this political and social debate is largely responsible for giving us the user-fee law that now governs the FDA and pharmaceutical sponsors.

^{viii} The 1997 reauthorization occurred as part of the Food and Drug Administration Modernization Act (FDAMA) of 1997.

^{ix} See testimony of Janet Woodcock, M.D., Acting Commissioner for Operations, FDA, *Drug Safety and the Drug Approval Process*, hearings before the Senate Committee on Health, Education, Labor and Pensions, March 3, 2005; <http://www.hhs.gov/asl/testify/t050303b.html> (accessed October 16, 2005). “Under the PDUFA approach, industry provides additional funding in return for FDA’s efforts to meet drug-review performance goals that emphasize timeliness but do not alter or compromise our commitment to ensuring that drugs are safe and effective before they are approved for marketing.” Mostly concurrent with PDUFA, which provides a steady revenue stream to the FDA, employment in the Center for Drug Evaluation and Research increased from 1,041 in 1981 to 2,395 in 2005. Most recent data on CDER staffing totals are available at:

<http://www.fda.gov/oc/oms/ofm/budget/2006/HTML/Summary/CDER.htm> (accessed July 23, 2005).

^x PDUFA (1992) specified that by 1997, the FDA should review and act upon 90% of standard NDAs in 12 months, 90 percent of priority NMEs in 6 months. FDAMA (1997 also referred to as “PDUFA II”) specified that by FY 1999, 30 percent of standard NDAs in 10 months, by FY 2002 90 percent of standard NDAs in 10 months. It retained the same six-month deadline for priority NMEs. PDUFA III (2002) and PDUFA IV (2007) continued the FDAMA standards. The legislation also set goals such that a large share (usually 90 percent or more) of new molecular entities (NMEs) would be reviewed by a certain date. The embedded incentive in PDUFA was that, if the FDA failed to meet the review time goals, Congress would not renew the user-fee program, depriving FDA of a major revenue source. C. Lewis, “FDA Begins Product Approval Initiative,” *FDA Consumer*, May-June 2003. For review time goals to be reached by FY 2002, see U.S. FDA, Office of the Commissioner, Office of Policy and Planning, “Report on PDUFA Goals: Original New Product Applications,” <http://www.fda.gov/oc/pdufa/report2002/2002-onpa.html> (accessed October 16, 2005). For FY 99 goals and a summary of earlier deadlines and goals, see U.S. FDA, Office of the Commissioner, “Performance on FY 99 FDAMA Goals,” <http://www.fda.gov/oc/fdama/fdamaplresponse/rptgoalsFY99.html> (accessed October 16, 2005).

^{xi} Since March and April 2008, the FDA has begun to relax its adherence to these review-time goals.

^{xii} A drug is deemed “priority” if “no satisfactory alternative therapy exists” or if it represents “a significant improvement compared to marketed products” (FDA 2007).

^{xiii} Supplemental models add disease specific frailties.

^{xiv} These figures compare hazards across time; exact comparisons between the different regimes (lines) are possible but a full comparison requires a fully nested model which we have not presented here. Tables 1 and 2 report the hazard ratios, standard errors, and a series of non-linear likelihood-ratio tests.

^{xv} M. Meadows, Why Drugs Get Pulled Off the Market. *FDA Consumer* 36 (1) (January-February 2002). L. D. Sasich, Comments before the Food and Drug Administration’s Public Meeting on the Prescription Drug User Fee Act (PDUFA). September 15, 2000, (HRG Publication #1536); URL: http://www.citizen.org/publications/print_release.cfm?ID=6737 [accessed December 22, 2005]; T. Moore, Psaty, B. M., Furberg, C. D. Time to Act on Drug Safety. *JAMA*, 279(19) (1998):1571-1573.

^{xvi} We constructed this measure based on Lasser et. al. (2002), but expanded it (by searching the *Physicians’ Desk Reference* and FDA safety alerts) to include drugs that received a new black-box warning from 2000 through December 2008.

^{xvii} We gathered safety-related market withdrawals from the global market using data from three sources. The first is Lasser et al. (2002). Second, SCRIPS reports and Pharmaprojects identify all NMEs approved in the United States and then withdrawn for safety reasons in at least one industrialized nation since 1980. In Pharmaprojects, this includes most European nations as well as Japan, Australia, New Zealand, India and the United States. (Few drugs are withdrawn in just one country.) Third, we examined safety-related withdrawals from Canada since 1963. We code a drug as experiencing a safety-based withdrawal if it was withdrawn in the United States or by two or more foreign regulators in Europe or a single European country, Japan, Canada, Australia, or New Zealand. The European Union’s drug regulatory agency, the EMEA, would count as one regulator in this calculation, but member countries’ regulators might make separable decisions regarding withdrawal. Adjusting this definition slightly in different ways does not affect the measurement. All of our drugs were withdrawn from the U.S. market or from the United Kingdom, Australia and the European Union. In two of three cases where the drugs were not withdrawn officially from the U.S. market (trovafloxacin mesylate and alatrofloxacin mesylate), their producers stopped manufacturing their dosage forms and the drug was withdrawn unofficially. The other case is Tolcapone, which was withdrawn in the EU, the United Kingdom (separately from the EU), Canada, and Australia. We examine global withdrawals because we would like a measure of withdrawals that is less dependent upon FDA decision making.

^{xviii} Data for this measure came from the FDA’s MedWatch database. We model a count of safety alerts for each NME, restricting the count to unique alerts that pertain to an NME’s potential adverse events from 1996-2008.

^{xix} Discontinuation is a code (“3”) in the Product Market Status of approved products, as tracked by CDER, and available at the Drugs@FDA website. It tracks discontinuations of particular dosage and administration versions of an NDA. Accessed July 28, 2005. Our indicator is *not* coded as 1 for discontinuations accompanied by a statement that there exists a Federal Register determination that discontinuation occurred for reasons other than safety or efficacy; 22 NMEs were coded as 0 in our indicator as a result of having all of their discontinuations tagged with this qualifying statement.

^{xx} See for instance the discontinuation of Agenerase® (amprenavir) 150 mg capsules by GlaxoSmithKline in December 2004. According to the company’s letter to the FDA, the product was discontinued “because the clinical demand for AGENERASE 150 mg capsules has diminished significantly. Additionally, in the recent treatment recommendations by the Department for Health and Human Services (DHHS), AGENERASE is no longer recommended as a component

of a preferred or alternative initial regimen.” “Dear Healthcare Professional” letter, September 2004. http://www.fda.gov/cder/drug/shortages/AgeneraseLetter_E2.pdf (accessed November 2, 2005).

^{xxi} Including fixed effects for firms that sponsored a large number of drug reviews did not substantively change the results that we report here but did introduce an even greater number of terms into the analysis. In the results reported, we therefore simply include a measure of the number of previous drug approvals the drug’s sponsoring firm received from the FDA.

^{xxii} This variable is an indicator, coded as 1 if the drug received priority review.

^{xxiii} However, fewer drugs on the market to treat a disease may also reflect patient demand for new therapies.

^{xxiv} Although not reported, these results held up well with modeling disease category using fixed rather than random effects, and when running extreme value regressions.

^{xxv} Only drugs approved after 1992 were potentially exposed to the “looming deadline” treatment.

^{xxvi} The former has the advantage of permitting use of a larger fraction of observations to be used in the analysis and has other desirable properties relative to nearest-neighbor propensity-score matching. The second method, coarsened exact matching (CEM), places covariates in multivariate bins of varying dimensions. Observations in either the treatment *or* control group are discarded if observations of the opposite treatment status do not appear in the same bin. While this method produces inferences about an undefined and unknowable target population, it potentially provides better balance when treated observations cannot be matched easily.

^{xxvii} Before applying matching algorithms, we discard observations approved before 1992, since these observations were not potentially exposed to treatment (their propensity score was, by definition, zero). Most good matching procedures would discard pre-1992 observations, but their inclusion complicates the matching procedure. We then match on four covariates that predict drugs’ approval time: year of NME submission, priority review status, log of order of disease-niche entry, and number of previous firm submissions. Under optimal matching with two control observations matched to each treated observation, this results in 86 treated and 172 control observations, permitting inferences about causal effects of “looming deadline” approval on drugs that were approved under a looming deadline. Coarsened exact matching yields a smaller sample of 45 treated observations and 80 control-group observations.