
RWJF Scholars in Health Policy Research Program

Deadline Effects in Regulatory Review: A Methodological and Empirical Analysis

**Daniel Carpenter
Jacqueline Chattopadhyay
Susan Moffitt
Justin Grimmer
Jake Bowers
Clayton Nall
Evan James Zucker**

October 2009
Working Paper Series

WP-45

Preparation and dissemination of this working
paper series was assisted by a grant from
The Robert Wood Johnson Foundation
Princeton, New Jersey

Please contact Daniel Carpenter with questions and/or comments:
dcarpenter@latte.harvard.edu

Deadline Effects in Regulatory Review: A Methodological and Empirical Analysis

Daniel Carpenter⁺
Jacqueline Chattopadhyay⁺
Susan Moffitt*
Justin Grimmer⁺
Jake Bowers#
Clayton Nall⁺
Evan James Zucker@

Version 3.1
August 2009

ABSTRACT

This paper elaborates two sets of statistical models for the analysis of regulatory review deadlines or “review time goals” and their influence upon regulatory decisions. In the first set, we elaborate dynamic duration models of time to decision, with particular focus on semi-parametric methods from which the behavioral structure of a regulatory review can be induced or “backed out.” In the second set, we consider different estimators (generalized linear models with random effects, and two different matching estimators) for observational analysis of whether the deadlines in question influence the “quality” of the reviews or the likelihood of error. We apply these methods to the imposition of review-time goals by the Prescription Drug User-Fee Act (PDUFA) and its amended successors upon new drug review by the U.S. Food and Drug Administration (FDA). Using the first set of methods, we find broad evidence that the deadlines have induced a piling of approvals right before the deadline elapses. Using the second set of methods, and controlling for various correlates of drug quality and conditions of the review, we find that these “just-before-deadline” approvals are subject to substantially different post market experiences than drugs approved at other times in the review process. While our empirical analysis is focused upon drug approval in the United States, applications in other national settings and in many other regulatory policies are possible.

⁺ Department of Government, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts.

* Department of Political Science, Brown University, Providence, RI.

Department of Political Science, University of Illinois, Champaign-Urbana, IL.

@ Harvard Medical School, Boston, Massachusetts.

First prepared for presentation at the “Strengthening FDA” Workshop, February 22, 2007, at the Program on Scientific Knowledge and Public Policy (SKAPP), George Washington University, Washington, D.C. We thank Jerry Avorn, Ernst Berndt, Einer Elhauge, Gary King, Bruce Psaty, Meredith Rosenthal, Katherine Swartz and Heidi Williams for critical remarks and helpful discussions. Janna King provided excellent research assistance. Carpenter acknowledges a Robert

Wood Johnson Foundation Investigator Award in Health Policy Research, the National Science Foundation (SES-0076452), and Harvard University (the Center for American Political Studies and Institute for Quantitative Social Science) for support. Professor Carpenter neither seeks nor accepts research funding or any other form of compensation from the FDA, from private entities that sponsor product applications to the FDA, from patient advocacy groups or from public interest research groups (PIRGs). In addition, Professor Carpenter does not directly own stock or any other form of asset in any company regulated by the FDA. All arguments, errors and omissions are the responsibility of the primary author.

The duration of choice remains a crucial dimension of government decision making. Many agencies and regulators possess discretion and power not merely in their ultimate choices, but also in the question of when those choices are made. While an emerging literature in institutional political science, economics and sociology has begun to tackle these questions (Spence 1999; Olson 1997, 2002; Ando 1999; Carpenter 2002; Kosnik 2006; Whitford 2005; Yackee and Yackee 2009), there has been very little analysis of how political and administrative institutions such as deadlines affect the timing and character of regulatory decisions. Some scholarship suggests that Congress uses deadlines to influence the amount and kind of information agencies receive before making regulatory decisions (Bawn 1995, 65). Yet, surveys find bureaucrats perceive deadlines to be a weak form of legislative oversight (Furlong 1998). When deadline institutions have been represented in such research, they often enter statistical analyses as indicator or “dummy” variables, leaving too much unknown about the mechanics with which these institutions change regulatory behavior.

In this paper we elaborate some statistical methods that permit analysts to examine the influence of deadlines upon regulatory decision making. Deadline institutions impose a penalty (explicit or implicit) for the endurance of a decision process beyond a specified timepoint (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. In other cases, the deadline penalty is smaller, such that the deadline becomes one of plural factors to influence the timing of regulatory behavior. On these and other questions, the scholarly literature is all but silent. We are unaware of any literature examining how deadlines influence agency decision making, and very little on how deadlines influence organizational learning and delay more generally.

The Timing Effect and the Quality Effect. 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. Consider an example from the scholastic realm. If for instance the dean of a university were to institute a new rule requiring professors to spend no more than 30 minutes with a term paper before grading it, then observers would be interested in at least two questions: (1) does the time limit on grading actually influence the distribution of professors’ work? and (2) does the time limit affect the grading patterns and other observable

features of the grades? If professors spent no more than 15 minutes grading papers before the half-hour limit was imposed, then the deadline would be expected to have little if any effect. If on the other hand the usual professor was accustomed to spending one hour with each term paper before grading it, we would be interested in whether the new deadline really did shorten the grading time, and we would also be interested in questions such as whether students felt that they were shortchanged by the deadline. We might ask, then, not only whether the timing of grading changed, but whether the distribution of grades was changed by the imposition of a deadline, and perhaps whether student grade challenges and protests increased because of it.

In this paper, we develop statistical methods to address these two dimensions of regulatory choice. We first elaborate a statistical model – the dynamic Cox model with time-varying covariates – that allows the researcher to “peer inside” the regulatory decision and to assess whether the deadline influenced its ultimate timing. We then elaborate generalized linear models that examine, observationally and with statistical controls, whether the deadlines have influenced the results of the ultimate decision: whether drugs approved shortly before a deadline experienced post-marketing safety issues at a greater rate than drugs approved at other points in the drug review process. We then supplement these models with matching analyses. In each case, we are not creating new statistical methods, but rather adapting relatively novel statistical methods for particular use in institutional political science and the social scientific study of regulation and organizations more generally.

We apply these methods 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 revision in 1997, 2002, and 2007 imposed quite specific if somewhat flexible review time goals upon the FDA. We assess whether these review time goals (“PDUFA clocks”) – which can be fruitfully viewed as deadlines with implicit penalties for their violation – have influenced the timing of FDA decision making. We also ask whether drugs approved just before these PDUFA deadlines are subject to different post-market regulatory experiences than other drugs approved by the FDA. While the application here is quite specific, the methods have broader application, including to other products and applications reviewed by the FDA (devices, vaccines, supplements), to other

pharmaceutical regulators (the European Medicines Evaluation Agency of the E.U., or Health Canada), to other regulatory situations where delay is an important feature of the administrative landscape (dam license renewal by the Federal Energy Regulatory Commission, for instance), and to other agencies facing deadlines.

While our empirical analysis is focused upon drug approval in the United States, applications in other national settings and in many other regulatory policies are possible. In particular, drug regulators in the European Union also operate under a system of review time deadlines, and many other FDA decision processes not analyzed here (the regulation of medical devices, biologic drugs and generic drugs, among others) are governed by deadlines. In addition, a wide variety of administrative decisions is characterized by timing phenomena, including licensing and rulemaking, and in many of these cases implicit or explicit deadlines may apply.

A crucial feature of the present analysis is that the two statistical analyses are tied to one another in a flexible manner. The dynamic Cox hazard estimation offers semi-parametric statistical evidence that the PDUFA clock deadlines induce changes in FDA decision making, and the results of these hazard analyses inform the statistical analysis of post-marketing regulatory events. Using flexible statistical techniques to refine and yield the most comparable and controlled samples, we use this “before versus after” comparison to shed light upon the safety consequences of PDUFA deadlines.

Some preliminary support for our hypotheses appears in Carpenter, Zucker and Avorn (2008), in which very brute statistical evidence (contingency tables and exact logistic regressions with one treatment and one control variable) displays an association between “just-before-deadline” approvals and postmarket safety problems. In several ways, the analyses here are appreciably different and much more comprehensive than in Carpenter, Zucker and Avorn (2008). First, our sample includes a much more extensive range of drugs, usually from January 1980 to December 2008, though in some cases a full range of drugs submitted from 1950 to 2008. (Carpenter, Zucker and Avorn focus on a 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), which is important as observers, including FDA officials themselves, have previously argued that priority drugs account for both quicker reviews and greater

postmarket safety issues (see, for example, Berlin 2009). Third, compared to previous studies and to Berlin (2009) 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 forms of observational analysis (namely optimal and nearest-neighbor matching) to assess the robustness of causal inferences concerning deadline institutions and regulatory behavior.

We begin empirically by describing the user-fee law of 1992 and its amended versions passed in 1997, 2002, and 2007. We then elaborate a dynamic Cox model for estimating the effect of deadlines upon regulatory review timing, and apply these methods to FDA review times for new molecular entities (NMEs). We then turn to generalized linear models to assess the influence of the deadlines upon the quality of FDA choices, examining the correlation of deadlines with observable post-market regulatory events such as drug withdrawals, additional black-box drug warnings, changes to drug labels, drug safety alerts, and dosage form discontinuations.

I. The User-Fee Program and Review Clocks: A Description

Drug review and marketing approval by the FDA is one of the most consequential (and controversial) regulatory policies of our time. The FDA's drug review practices have been criticized from numerous perspectives. The Administration 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 things. First, much (though not all) 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.

The first PDUFA legislation was passed in 1992 and was reauthorized in 1997, 2002 and 2007,¹ The four user-fee laws have made many changes at the FDA, among the most important of which is the provision

¹ The 1997 reauthorization occurred as part of the Food and Drug Administration Modernization Act (FDAMA) of 1997.

of new staff resources to the Center for Drug Evaluation and Research, whose total employment has risen from 1,041 in 1981 to 2,395 in 2005.² The essence of the bargain struck under PDUFA is that the agency gets needed staff while the pharmaceutical industry and concerned disease advocates get quicker approvals.³ There were many ways to achieve the aim of quicker approvals, but the PDUFA legislation did so in a specific way: the introduction of a review clock. From the date of first NDA submission, a drug's "review clock" begins ticking. The legislation then embedded goals such that a large percentage (usually 90 percent or more) of NMEs would be reviewed by a certain date. The embedded incentive provision in PDUFA was that, if the FDA failed to meet the review time goals, the user-fee program would not be renewed. The clock differed according to whether the new drug application was designated "priority" or "standard," as follows.⁴

PDUFA, 1992 (signed into law 10/29/1992): by 1997, review and act upon 90% of standard NDAs in 12 months, 90 percent of priority NMEs in 6 months.

FDAMA, 1997 "PDUFA II" (signed into law 11/21/1997): by FY 1999, 30 percent of standard NDAs in 10 months, by FY 2002 90 percent of standard NDAs in 10 months; same as PDUFA for priority NMEs.⁵

² 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). Data from earlier years (including 1981) are taken from the CDER historical database, FDA Project Archive, Department of Government, Harvard University. See also P. B. Hutt and S. White, *A Statistical History of the Food and Drug Administration* (unpublished, FDA History Office, 1992).

³ 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."

⁴ 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).

⁵ In fact the FDA trumpeted the fact that it was ahead of congressional statute in meeting these deadline goals. For instance, by June 2000, "51 percent were within the time period for review," whereas the statutory goal was 30 percent, meaning that over half of NDAs were meeting the 10-month deadline goal. FDA

PDUFA III, 2002 (signed into law 6/12/2002): For standard and priority NDAs, same deadline months as in FDAMA.

PDUFA IV, 2007 (signed into law 9/27/07): For standard and priority NDAs, same deadline months as FDAMA

Among the many notable features of the user-fee program is the absolute nature of the “PDUFA clock” deadline. If the deadline is 12 months, then once the 12th month has elapsed, CDER has far less incentive to hurry the drug, as it is simply impossible that the drug can count as one meeting the annual review time goals. Because the review time goals were structured upon deadlines, an absolute priority is given to the deadline as opposed to an “average” or “median” review time. Put differently, the user-fee laws of 1992, 1997, 2002 and 2007 accelerated the FDA *in specific ways*. The user-fee law did not ask for 2- and 3-month reviews, and it did not ask that the occasional two-year review disappear entirely. 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 review behavior.

Statistical Implications of Deadlines: Hypotheses. We hypothesize that the PDUFA clock deadlines introduce a temporal discontinuity into drug review. 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 has been reached, then the incentives for completing NDA review in the next two months are quite high, because near-term completion will mean that the agency has met the review clock for this drug. However, if the agency fails to meet the review time goals, and the thirteenth month of the review clock has been reached, then there should be much less incentive for the agency to approve the drug in the next two months. Hence we should observe a high proportion of approvals concentrated or “piled up” in the months and weeks just before the deadline, and relatively few concentrated just after the deadline. The same logic should obtain for a six month review. When the fifth month of review has been reached, incentives to approve in the next

month or two are quite high. When the seventh month of the review has been reached, however, there is much less incentive to approve in the next two months.

II. Statistical Decomposition of the Regulatory Review Cycle.

In order to test our hypotheses, we must retrieve statistical estimates of the conditional probability of approval at each point of the regulatory review cycle. In other words, we seek to address the question: at each month of the review cycle, what is the relative hazard rate of approval in *this* month, given that the product or case has not yet been approved? We conduct likelihood-based hazard analyses of FDA review times and retrieve month-specific hazard estimates that allow us to construct a statistical portrait of the FDA review cycle. To minimize dependence upon parametric statistical assumptions we employ Cox proportional hazard models.⁶

Fundamental Terms and Indexation. We begin formal elaboration of our statistical models by defining terms and indices that will be used throughout the following analysis. For any new drug application submitted to the FDA, identification by at least four indices is possible. All of the drugs in a selected sample can be individually indexed i – the assignment of unique NDA numbers to drug applications is one example – which can serve as an encompassing index. We use Greek letters to denote the other three indices (sponsor, primary indication and time submitted). The drug will be submitted by a sponsor κ in year ζ for primary indication ψ . Only the index i is alone sufficient to identify all drugs, as knowing the sponsor (κ), the primary indication (ψ) and the year submitted (ζ) may often leave more than one drug in a category. We further create two sets of indicator variables for the sponsor and primary indication of a drug. We will denote the indicator variable for a particular sponsor as S_{κ} , equal to one if sponsor κ has submitted drug i for FDA review (and 0 otherwise), and we will denote the indicator variable for a primary indication as D_{ψ} , equal to one if the drug i is intended to treat disease ψ , 0 otherwise.

⁶ We have employed other models that embed parametric assumptions, with results that are substantively identical to those that we display here. Estimates are available from authors upon request.

We now turn to estimation of the approval hazard function. Intuitively, relative to a baseline month (this is set as the first month of the review cycle in all of our analyses), we seek to retrieve the ratio of (a) the hazard rate in the month under consideration to (b) the hazard rate for the baseline month, controlling for all relevant and feasible covariates. Setting the baseline month as the first month of review (the denominator hazard rate), estimates of the hazard for all subsequent months can be used to construct an ‘‘approval hazard ratio’’ (AHR) or the ratio of the hazard (h) in month $\tau = t$ to the hazard in month one ($\tau = 1$), holding all other variables (X) and parameters (β) constant.

$$AHR_{\tau=t} = \frac{h_{\tau=t}(\beta' X)}{h_{\tau=1}(\beta' X)} \quad (1)$$

Once approval hazard ratios for each month are calculated, they can be compared to determine whether some months present a higher or lower probability of approval than others. Using data from before and after the different institutional changes, we can also examine whether the review cycle has changed after the introduction of these changes, and whether approvals are concentrated in certain months of the review cycle after the reform. Let REFORM be a binary variable measuring whether the drug was submitted before (0) or after (1) a certain institutional reform. Holding all other observed variables constant, we can compare the approval hazard ratio for any month t (say the tenth month of the review cycle) before and after the user-fee act, as follows:

$$AHR_{\tau=t} = \frac{h_{\tau=t}^{REFORM=1}(\beta' X)}{h_{\tau=t}^{REFORM=0}(\beta' X)}$$

We can then test whether discontinuities exist by comparing approval hazard ratios within regimes, as follows

$$\hat{\xi}_{t,t+1} = AHR_{t+1} - AHR_t = \frac{h_{\tau=t+1}^{REFORM=1}(\beta' X)}{h_{\tau=t+1}^{REFORM=0}(\beta' X)} - \frac{h_{\tau=t}^{REFORM=1}(\beta' X)}{h_{\tau=t}^{REFORM=0}(\beta' X)} \quad (2)$$

The statistical significance of $\hat{\xi}_{t,t+1}$ must be determined not from a t -test or z -score (where the null hypothesis posits a single and known value), but from a restricted test (such as a Wald test or score test W) that compares the values of two or more stochastic coefficients.

We begin by retrieving a portrait of the behavioral review cycle of CDER by estimating several forms (parametric and semi-parametric) of dynamic (time-varying covariate) duration models. We begin with the

semi-parametric version, which is a form of the Cox model. We begin with representation of the review process in counting process notation. For any drug i , let

$$R_i(t) = \begin{cases} 1 & \text{if drug } i \text{ is under review at time } t \\ 0 & \text{otherwise} \end{cases}$$

We then seek estimates of a k -element parameter vector β , where β contributes to the partial likelihood function

$$PL(\beta) = \prod_{i=1}^n \prod_{t_i \geq 0} \prod_{s=0}^{t_m} \left\{ \frac{R_i(s) e^{X_i(s)\beta}}{\sum_j R_j(s) e^{X_j(s)\beta}} \right\}^{dN_i(s)} \quad (3)$$

Here s is a variable of integration that varies within spells of total length t_i [the review time for drug i] and m (“month”) is an arbitrary index for s which can, without loss of generality, be represented as a discrete, connected and dense partition of t_i into increments t_m such that $\sum_m t_m = t_i$. The sample size is n and N is a counting process (which can always be modeled as locally Poisson (Therneau and Grambsch 2000: 11)) that offers a generalized characterization of the number of events in $[0, t_m)$. The log-partial likelihood is then

$$l(\beta) = \sum_{i=1}^n \sum_{t_m=1}^{t_i} \int_0^{t_m} \left[R_i(s) X_i(s) \beta - \ln \left(\sum_j R_j(s) e^{X_j(s)\beta} \right) \right] dN_i(s)$$

which can be differentiated to generate a k by 1 score vector, as follows

$$U(\beta) = \sum_{i=1}^n \sum_{t_m=1}^{t_i} \int_0^{t_m} \left[X_i(s) - \bar{x}(\beta, s) \right] dN_i(s)$$

Where $\bar{x}(\beta, s)$ is a weighted mean of X over those drugs still under review at time s . To represent this quantity, let y_i be the *approval score* for the i th drug, that is $y_i(\beta, t) = \exp[X_i(t)\beta]$, and let $Y(s)$ be the aggregate number of drugs still under review (or at “risk” of getting approved) at time s . Then

$$\bar{x}(\beta, s) = \frac{\sum_{j=0}^{Y(s)} R_j(s) y_j(s) X_j(s)}{\sum_{j=0}^{Y(s)} R_j(s) y_j(s)}$$

Notice that the scored observations $R_i(s)y_i(s)$ function as weights for the independent variables. The negative second derivative can be formed from the k -by- k information matrix

$$I(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{t_m=1}^{t_i} \int_0^{t_m} \frac{\sum_i R_i(s)y_i(s) [X_i(s) - \bar{x}(\boldsymbol{\beta}, s)] [X_i(s) - \bar{x}(\boldsymbol{\beta}, s)]}{\sum_i R_i(s)y_i(s)} dN(s)$$

For various reasons of information and computational feasibility, the inverse of the observed information matrix $I^{-1}(\boldsymbol{\beta})$ is used in lieu of the inverse of the *expected* information matrix $\{E_{nt} I(\boldsymbol{\beta})\}^{-1}$, even though the latter is analytically appropriate (Therneau and Grambsch: 40-41). The maximum partial likelihood estimator is obtained by solving the partial likelihood equation $U(\hat{\boldsymbol{\beta}}) = 0$. This is done via a Newton-Raphson algorithm, which iteratively computes $\hat{\boldsymbol{\beta}}^{(n+1)} = \hat{\boldsymbol{\beta}}^{(n)} + I^{-1}(\hat{\boldsymbol{\beta}}^{(n)})U(\hat{\boldsymbol{\beta}}^{(n)})$ until convergence is reached. With dynamic Cox estimation, order months in X such that the τ th month is the τ th element of $\boldsymbol{\beta}$. With the first month estimated as a baseline, the AHR estimate for month τ is $\hat{\eta}_\tau = e^{\hat{\beta}_\tau}$. Because the month-specific estimates are embedded in the coefficient vector of the dynamic Cox model, tests of differences between any two months in the review cycle can be executed by means of a score test statistic, which can be computed using the first iteration of the Newton-Raphson algorithm. This is

$$U'(\boldsymbol{\beta}^{(0)})I(\boldsymbol{\beta}^{(0)})^{-1}U(\boldsymbol{\beta}^{(0)})$$

Let ν and ω be any two months in the review cycle, such that $\nu > 0, \omega > 0, \nu \neq \omega$. Then we can represent restrictions by $q(\hat{\boldsymbol{\beta}}) = g$ such that $\hat{\beta}_{\tau=\omega} = \hat{\beta}_{\tau=\nu}$, and the following test score statistic (which is distributed χ^2) will represent an appropriate test of the identity of the approval hazard across the two cycle months in question.

$$W = U' \left(q_{[\hat{\beta}_{\tau=\omega} = \hat{\beta}_{\tau=\nu}]}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}} \right) I \left(q_{[\hat{\beta}_{\tau=\omega} = \hat{\beta}_{\tau=\nu}]}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}} \right)^{-1} U \left(q_{[\hat{\beta}_{\tau=\omega} = \hat{\beta}_{\tau=\nu}]}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}} \right) \quad (4)$$

An Application to FDA Review Times under Deadlines.

In Figure 1, we plot monthly approval hazard ratios retrieved from dynamic Cox estimation for the first 24 months of the review cycle where these can be estimated. Figure 1 shows that for drugs submitted before 1993, no discontinuity was in evidence at the tenth or twelfth months of review. Indeed, for molecules submitted before the user-fee act governed them, the approval hazard ratio for drugs in the eleventh month were not statistically differentiable from that in the *seventh* month ($\chi^2 = 0.00$; $p = 0.974$), the *ninth* month ($\chi^2 = 2.38$; $p = 0.1229$), or the *seventeenth* month ($\chi^2 = 2.05$; $p = 0.1521$).

[Figure 1 here]

Our analyses suggest that the PDUFA clocks have dramatically changed the statistically observed behavioral structure of the FDA drug review cycle. 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 eleventh month of review compared to the same month before PDUFA ($\chi^2 = 6.39$; $p = 0.012$). Moreover, as hypothesized, approval hazards fall off appreciably for the month after the review clock deadline ($\chi^2 = 12.06$; $p = 0.0005$). 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 ninth and tenth month of the review cycle, and again a corresponding decline in the eleventh and twelfth month. Specifically, the approval hazard in the month before the ten-month FDAMA clock deadline is *eighteen* times greater than the approval hazard in the month after the review deadline has elapsed ($\chi^2 = 18.02$; $p < 0.0001$).

Figure 2 shows that these patterns apply to priority drugs as well. For priority drugs submitted after the introduction of 6 month PDUFA deadlines, we observe a sizable increase in approval hazards for the sixth month of the review. The hazard rate drops considerably in the eighth month, after the deadline has passed.

[Figure 2 here]

III. Deadline Institutions and Post-marketing Regulatory Events: Data and Measures.

The available review time evidence, then, generally supports that hypothesis that the PDUFA clocks have influenced FDA review behavior in quite particular and observable ways. We now turn to investigate the second hypothesis, namely that the review clock institutions have exercised an influence upon not just the timing but also the quality of the FDA's decision, including and especially the dimensions of drug safety and postmarketing regulatory issues.⁷ While some analysts (including the FDA itself) have examined whether the *overall* rate of drug safety problems has risen or fallen since 1993,⁸ we conduct a different, more focused comparison. We compare the incidence of postmarketing and regulatory issues for pre-deadline approvals to post-deadline approvals during the user-fee regimes. In other words, we compare the postmarketing experiences of drugs approved in the months before the PDUFA clock deadline to the postmarketing experiences for drugs approved in the months after the deadline.

Measures of Postmarketing Regulatory Events (PMREs). To assess whether pre-deadline approvals are associated with postmarketing issues at a greater rate than post-deadline approvals, we examine five measures of post-marketing regulatory events. The first four (black box warnings, safety-related label changes, safety-based withdrawals, and safety alerts) are directly concerned with safety issues. 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.

Black box warnings. One measure of postmarketing safety we examine is a dichotomous indicator of whether or not the approved drug received a new post-market black-box warning on its label for a

⁷ We use the term "quality" of decisions sparingly here, and only as shorthand. The hypothesis in question is quite particular: do the PDUFA review time goals influence not only the timing of the FDA's review behavior but also other features of the drug's clinical profile that are observed in its postmarketing phase? Whether the quality of the FDA's judgments is affected is a much more difficult issue to address and is beyond the scope of this paper.

⁸ 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.

significant new adverse drug reaction (ADR), as identified and reported by Lasser and colleagues (2002). Lasser and colleagues relied upon changes to drug descriptions in the *Physician's Desk Reference* to compile their list. Using this measure has the disadvantage of excluding drugs very recently approved from the sample – Lasser and colleagues stopped their list in 2000 – making it difficult to test for FDAMA-related effects, which could only be observed for molecules submitted after September 1997.

To this measure we added a list of drugs receiving a new black-box warnings from 2000 through December 2008, compiled by searching the Physicians' Desk Reference and FDA Safety Alerts. In addition, the addition of black-box warnings can be tracked using a list of post-approval drug safety warnings compiled by physicians and epidemiologists at the Kansas University Medical Center (KUMC).⁹

Safety-Related Labeling Changes. We supplement the measure of black-box warning additions with a second measure of label changes, by modeling a count of the number of safety-related label changes that each NME experienced after it was introduced into the market. Using data retrieved from the FDA's MedWatch, we calculated the number of official safety-related alterations to each NME's label between 1996 and 2006. This includes changes to the sections of a drug's label describing warnings, precautions, contraindications and adverse events. Because these data are not systematically collected before 1996 from MedWatch, the analysis sample will be smaller when examining this outcome variable.

Safety-based withdrawals. We also model a dichotomous indicator of whether an NME faced a safety-based withdrawal. We gathered safety-related market withdrawals from the global market. Our data on withdrawals are from three sources. First, we gather data on safety-based withdrawals from Lasser and colleagues (2002). Second, we gathered data from SCRIPS reports and Pharmaprojects, identifying all NMEs that had been approved in the United States and then were withdrawn for safety reasons in at least one industrialized nation since 1980. In Pharmaprojects, this includes most all European nations as well as Japan, Australia, New Zealand, India and the United States. It is worth noting that very few drugs are withdrawn in

⁹ The KUMC maintains an information page for its formulary at <http://www.formularyproductions.com/kumc/> (accessed October 16, 2005), which has direct links to lists of drugs with black box warnings as well as FDA safety issues. The KUMC list was researched and compiled under the direction of Joyce Generali, MS, RPh, FASHP, Director of Drug Information and Clinical Professor, Kansas University Medical Center.

just one country. Third, we examined safety-related withdrawals from Canada since 1963. Lexchin (2005) tabulates safety-based drug withdrawals from the Canadian market, though he notes that Health Canada's drug withdrawal information is non systematic and that his list "might not be complete." 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.¹⁰

We examine a wider sample of withdrawals than U.S. withdrawals only because we would like a measure of withdrawals that is less dependent upon FDA decision making. As recent controversies might suggest, regulatory agencies like the FDA may be less willing to revisit their own decisions, which casts doubt upon the FDA's own drug withdrawal decisions as a measure for "regulatory error."¹¹ None of this suggests that adding non-U.S. withdrawals produces a better indicator, nor are global withdrawals fully independent of U.S. regulatory decisions. It is certainly plausible, however, that non-U.S. withdrawals are less dependent upon initial FDA drug review than are U.S. withdrawals.

Safety-Alerts. A fourth measure of Post-Marketing Regulatory Events comes from our model of 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 1950. 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.

This measure does not double-count multiple alerts for the same issue. In many instances, the FDA and/or a pharmaceutical firm issues the same alert to different constituencies and/or through a variety of

¹⁰ 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.

¹¹ For a recent suggestion that the FDA regulates postmarketing risks more loosely than in other countries, see M. Kaufman, S. Vedantam, Pregnant Women Warned by FDA to Avoid Paxil, *Washington Post* (December 9, 2005): A3. More generally, see D.P. Carpenter, M.M. Ting, "The political logic of regulatory error," *Nature Reviews – Drug Discovery* 4(10) (October 2005): 819-23.

mediums. For example, an alert about the same adverse event may be issued as a letter to pharmacists, a letter to physicians, and as a press release to the general public. We would count these letters and press release as one unique alert. Moreover, this measure does not include safety-related alerts that do not reflect inherent problems with the drug, such as drug shortages, manufacturing changes, and product confusion between two or more drugs with similar names.

Product (dosage form) discontinuation. A fifth measure models drug dosage form discontinuations. What does discontinuation imply?¹² From the FDA website, discontinuation “indicates drugs that have been discontinued from marketing or that have had their approvals withdrawn for *other than safety or efficacy reasons*,” [accessed July 28, 2005]. In many such cases, this is due to weak clinical demand.¹³ Very often, product dosage forms are discontinued due to poor utilization patterns (hence poor sales) on the market. Dosage-form discontinuation may, we note, signify safety issues as well. When less glaring safety problems have arisen and have been noticed by physicians and clinical specialists, clinical demand for the drug may be dampened as a result. This is a case where safety may be a reason for withdrawal of dosage-form but not of the molecule entirely, hence a lesser safety-related issue may be emerging here without being widely observed or publicized. This possibility deserves separate analysis and lies outside the confines of this paper. Nonetheless, we do model dosage form discontinuations here. While we note that such discontinuations are not explicitly or officially safety or efficacy-related, our indicator of discontinuations is *not* coded as 1 for discontinuations accompanied by a statement that there exists a Federal Register determination that the discontinuation occurred for reasons other than safety or efficacy. Twenty-two NMEs

¹² 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.

¹³ 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).

that have undergone dosage-form discontinuations were coded as 0 in our indicator as a result of having all of their discontinuations tagged with this qualifying statement.

Another possibility not undertaken here would have been to examine adverse event reports (AERs), which have received some study in recent years (Olson 2004a). For several reasons – mainly because adverse event reports are often inconsistent and are heavily dependent upon physicians’ reporting patterns – we leave analysis of these data to another paper. Our aim is instead to focus on actions that *the FDA and firms must take* to *revisit* approved drugs, and to leave actions that are more directly dependent upon physician reporting for other analyses. We report summary statistics for these measures in Table 1.

[Table 1 about here.]

IV. Deadline Institutions and Post-marketing Regulatory Events: Specification and Estimation of Generalized Linear Models.

Drugs vary in numerous ways that are unobservable to the statistical 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 standard 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).¹⁴ The immediate result of this estimation strategy 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., Lasser 2002, Olson 1997, Carpenter 2002, Olson 2004b). Other analysts control for generic therapeutic category terms (for example, a binary indicator for all anti-neoplastic drugs or for all central nervous system (CNS) drugs), but not for particular primary indications. Because drugs are assigned to divisions based primarily upon their primary

¹⁴ In robustness checks, we also modeled disease category as a fixed effect; doing so does not substantively change the results as compared to those found using random effects.

indication, the primary indication index y is a sufficient index for CDER divisions, so any static factors associated with the division level are captured by this set of hundreds of terms.

Robustness checks have also included fixed effects for firms that sponsored a large number of drug reviews. Doing so 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 here, we therefore simply include a measure of the number of previous drug approvals the drug’s sponsoring firm received from the FDA.

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. Recall that for any drug i , its primary indication is indexed by 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^{PMRE} , and we estimate models of the form

$$y_{\psi ki}^{PMRE} = f\left(\alpha^S S_k + \gamma' Z_{\psi ki} + u_{\psi} + e_{\psi ki}\right) \quad (5)$$

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 α^S represents firm-specific coefficients, and where e is a model disturbance. Notice that primary-indication-specific effects are modeled as random effects.

For each of the five post-market event variables (black box warning indicator, safety-based withdrawal indicator, label change count, 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. Pre-Deadline Approval. In particular, for any deadline month and its preceding month ($\tau^{deadline}, \tau^{deadline} - 1$), we construct a “pre-deadline” approval indicator Z^{PRE} scored 1 if the drug in question was approved in $\tau^{deadline}$ or $\tau^{deadline} - 1$, and 0 otherwise. 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 2 through 6 as “approved within 2 months of deadline.”

2. Count of Previous Firm NME Approvals. To control for firm effects on drugs' post-marketing experiences, we include a measure of firm experience, derived from a count of the number of NMEs that a sponsoring firm has had previously approved.

3. NME Novelty. Log Order of Drug Entry. Prevailing scholarship demonstrates that the order of a drug's entry into the market (i.e. whether a drug is the 1st, 2nd ... nth drug to treat a particular disease) bears significantly on the timing of a drug's approval. Drug approval times are an increasing function of the drug's order of market entry: early entrants enjoy shorter review times (Carpenter 2002; Carpenter et. al. 2010). Not only does the order of market entry reflect patient demand for new therapies, but it also captures potential uncertainty about how the therapy will perform in patient populations outside of experimental conditions: more post-marketing regulatory events may accompany more novel drug therapies. We derived this measure by taking the log of the number of drugs previously approved to treat the drug's primary indication.

4. Priority Review. Our models also control for whether the FDA designated the drug as a priority or standard drug review. The FDA identifies drugs that represent significant innovation over existing therapies for priority review. Such innovation holds potential for greater drug safety, greater drug efficacy and yet greater uncertainty as new formulations are introduced into the market for the first time. This variable is an indicator, coded as 1 if the drug experienced a priority rather than standard review.

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 2, 3, 4 5 and 6 about here.]

Results: GLM Estimation. We report the generalized linear model results in Tables 2 through 6. We observe that 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 dosage form discontinuations at the 0.1 level at the least, and more often at the 0.05 level or better. Although not reported, these results held up well with modeling disease category using fixed rather than random effects, and when running extreme value regressions.

The results in the first columns of each regression results in tables 2 through 6 are coefficients rather than readily interpreted quantities of interest. In tables 2, 3 and 6, however, the second column of each regression result 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 3.73 times greater than the probability of a drug approved at some other time receiving a new black box warning post-market, according to the baseline model for drugs approved between 1950 and 2008, 3.48 times greater according to the full model for NMEs approved between 1980 and 2008, and 3.04 times greater according to the full model specified in the NMEs approved after 1992. These differences are statistically significant at above the 0.05 level.

The results for safety-based withdrawals are even more dramatic. For NMEs approved between 1950 and 2008, according to the baseline model, the probability of a safety-based withdrawal is 3.32 times greater for drugs approved in the two months before a deadline than for drugs approved at other times. According to the full model for NMEs approved between 1980 and 2008, the probability of a post-market withdrawal is 4.6 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 suggests that the probability of a safety-based withdrawal is a stark 5.12 times greater for a drug approved in the two months leading up to its approval deadline than for drugs approved at other times. It is interesting to note that 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. 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 drug approval under a looming deadline and adverse post-market events.

While less striking, the odds ratios for dosage form discontinuations also support this conclusion. For NMEs approved after 1979, the probability of a dosage form discontinuation is 1.71 times greater for drugs approved in the two months leading up to their approval deadlines than for drugs approved at other times. This difference is 1.82 for drugs approved from 1993 to 2008.

The exception to the story of post-market regulatory events associated with looming deadline approvals occurs in the case of label changes. There appears to be no consistent relationship between pre-deadline approval and the number of label changes a drug undergoes post-market. Nonetheless, the results on balance align with the literature that has criticized review deadlines for their potential to yield drugs with poorer post-market performance.

Results from Matching Analyses. We also seek to minimize the possibility that there are non-“ignorable” differences between the drugs in our implicit treatment group (“looming deadline” approval) and our implicit control group (not a “looming deadline” approval). For a variety of reasons related to non-random assignment, these groups may not be equivalent in all relevant aspects but for the just-before-deadline approval.

There is, to be sure, no way of eliminating non-random assignment as an explanation for our results, which are generated from a purely observational research design. To reduce the likelihood that our results are due to unmeasured covariates or to an imbalance of non-ignorable factors distinguishing the treatment and control groups, we resorted to matching analyses (e.g., Hansen 2004). To perform matching, it was necessary to remove all cases which had missing data for one or more of the variables used in the analyses with generalized linear models. After removing cases which lacked data for any of the 11 variables used in the

models,¹⁵ there were 993 cases left, which had submission dates ranging from 1943 through 2006. It was possible to perform nearest neighbor matching and optimal matching (Hansen 2004) on this full set of data.¹⁶

Performing nearest-neighbor matching of the “treatment” variable (pre-deadline approval using a two-month window) on the matching variables – year of NME submission, Priority review, log of order of disease-niche entry, and number of previous firm submissions – results in 86 “treatment” cases and 86 control cases, for a matched dataset of 172 cases. Performing optimal matching results in a matched dataset of 258 cases, 86 of which are treated and 172 of which are control cases.

[Table 7 about here.]

Basic results from nearest-neighbor matching and optimal matching appear in table 7. The results from nearest-neighbor matching support the original findings nearly without fail. The baseline model of safety-based withdrawals run on this matched data using a generalized estimating equation for logistic regression nearly supports the original finding of a positive, statistically significant relationship between safety-based withdrawals and approval within the two months leading up to an approval deadline ($Z = 1.80$). Running this same sort of regression on the nearest-neighbor matched data on the baseline model of black box warnings likewise supports the original finding of a strong, positive relationship between black box warnings and approval within two months of an impending deadline ($Z = 2.59$). Running this regression on the nearest-neighbor matched data for the baseline model of dosage form discontinuations supports the original finding of no statistically significant relationship between discontinuations and approval within two months of a deadline ($Z = 0.33$). Running generalized estimating equation for Poisson regression on the nearest-neighbor matched data for the baseline model of the total count of safety alerts supports the original finding of a positive but at best marginally significant relationship between safety alerts and approval time ($Z = 1.55$). Running a simple negative binomial model on the matched data for the baseline model of safety

¹⁵ The variables were disease code, priority review, year of NME submission, total firm submissions at the time of NME submission, safety-based withdrawal, black box warnings, number of label changes, total safety alerts, dosage-form discontinuation, log of disease-niche order of entry, and deadline approval indicator.

¹⁶ It was not possible to perform either of these two sorts of matching on the subset of those 993 cases with submission years after 1979, or for the subset of those cases with submission years after 1992. However, most of the drugs that were retained in the treatment and control groups constructed by the matching estimators are of recent (post-1980) introduction.

alerts yields the same conclusion ($\Pr(> |z|) = 0.087$). Running the generalized estimating equation for Poisson regression on the matched data for the baseline model of the number of post-market label changes a drug saw supports the finding of a positive but statistically insignificant relationship between label changes and approval time ($Z = 1.28$). Running a simple negative binomial model on the matched data for the baseline model of label changes yields the same conclusion ($\Pr(> |z|) = 0.33$). The results from nearest-neighbor matching thus support the results from all of the baseline linear statistical models obtained using non-matched data.

The results from optimal matching actually suggest a possibly stronger and more robust relationship between post-market events and just-before-deadline approvals than do either the original results reported for regressions on unmatched data or the nearest-neighbor matched data. Running the baseline model of safety-based withdrawals on the optimally matched data using the generalized estimating equation for logistic regression supports the existence of a positive relationship between safety withdrawals and approval within the two months leading to a deadline, and suggests, like the results from the nearest-neighbor matched data do, that the relationship is at least marginally statistically significant ($Z = 1.77$). The positive relationship found between black-box warnings and approval within two months leading up to a deadline is remarkably strong if we are to believe the results from the optimally matched data ($Z = 2.99$). The relationship between discontinuations and approval before a deadline is, though insignificant when run on unmatched data, significant when run on the optimally matched data using the generalized estimating equation or logistic regression ($Z = 2.03$). Running a generalized estimating equation for Poisson regression on the optimally matched data for the baseline model of the total count of safety alerts, in contrast to the results from the genetically matched data, supports the significance of the positive relationship between safety alerts and approval time, more significant than when run on unmatched data ($Z = 2.65$). Running a negative binomial model on the optimally matched data for the baseline model of safety alerts likewise suggests that the relationship is significant ($\Pr(> |z|) = 0.008$). Running the generalized estimating equation for Poisson regression on the optimally matched data for the baseline model of the number of post-market label changes a drug saw supports the finding of a positive and, in this case, highly significant relationship between label

changes and approval time ($Z = 2.25$). Running a negative binomial model on the matched data for the baseline model of label changes yields the same conclusion ($\Pr(>|z|) = 0.076$). According to the results seen under optimally matched data, then, the post-market events modeled here are *uniformly positively and significantly related* to approval within two months of a deadline. This set of results run on “optimally matched” data thus offers stronger support for the hypothesis of a positive relationship between adverse post-market events and approval under a deadline than do the originally reported results run on unmatched data.

Discussion and Conclusion

We note that our results are not uniform, and that some of them vary by specification of the model (hence our preference for reporting multiple specifications of the GLMs). In the main, however, it is worth noting that the statistically significant results are always positive partial correlations. That is, 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. We *do* observe a large number of statistically significant positive relationships between deadline approvals and PMRE rates, and these hold across statistical specifications.

For four of the five regulatory events modeled, we observe a significant, positive correlation 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 results for safety-based withdrawals in particular stand in contrast to a 2005 study from the Tufts Center for the Study of Drug Development, which reported drug safety withdrawals within the United States unrelated to the speed of FDA drug approval. While some psychology literature (e.g. Huber and Kunz, 2007) could give rise to an expectation that deadlines should lead to stronger scrutiny of drugs, the findings here suggest that this hypothesis is incorrect. The existence of deadlines in drug review does not appear to bring about a review process that more effectively guards against the realization of undesirable post-market risks. In fact, it appears correlated with just the opposite pattern.

Methodological Implications. Our results have several methodological implications for those examining regulatory decision making and its policy implications. First, our analyses point to particular institutional forms – deadlines – as having effects when reliably and rigorously measured. Yet in order to uncover such effects, analysts must employ methods different from (and more refined than) the simplistic comparisons of means and least squares regressions that have dominated regulatory analysis in the past. As concerns the FDA, our investigations suggest that numerous analysts of post-marketing safety, including the Administration itself, may be looking in the wrong place for policy effects of the user-fee law. Analysts ought not, we think, conduct generic comparisons of drugs approved before and after the user-fee act. Instead, analyses of the laws' effects should be targeted to the specific features of the law, of which the review clock deadlines are the most notable and most measurable.

Second, as it concerns the user-fee law, our findings represent something of a middle ground between those who believe that PDUFA's acceleration of drug review times was a result of its institutional features (Olson 2000, 2004b) and those who claim that it was a product of more staff (Carpenter et al., 2003). While we find support for the hypothesis that additional resources have accelerated review, we find a weaker relationship between resources and review times than do previous analyses by Carpenter and colleagues. We are able to test different hypotheses about the cause of drug review acceleration jointly, and we find in some respects that *both* resources and incentives influence molecular approval times.

Implications for Future Research. We have focused the present analysis upon new molecular drug approval in the United States, in part because it has recently become a well-studied phenomenon with plentiful data available. Yet it is important to understand that the models and methods elaborated here can be applied in other national settings and in many other regulatory policies. 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, including licensing and rulemaking. In many licensing and rulemaking situations, deadline institutions are operative (O'Connell and

Gerson 2009). Moreover, wherever deadlines apply to administrative timing processes in ways that are implicit or explicit, our models will be of potential analytic value.

The correlation between drug approval deadlines and adverse post-market events moreover suggests that political scientists should dedicate more attention to the sorts of time-related tools that legislatures use to influence agency behavior, and to how thoroughly elected officials themselves understand and analyze those tools. By imposing drug review deadlines, Congress sought to bring prescription drugs to market faster. In an effort to find a feasible means to do so, lawmakers perhaps paid insufficient attention to the possibility that this blunt instrument of leverage over the bureaucracy could have upshots other than speeding drugs to market. Had there been a bulwark of well-publicized evidence in place in 1992 that review deadlines might be significantly related to higher rates of adverse drug events, Congress and the President might have been more hesitant to embrace deadlines as a means to fast-track drug approval, either out of ethical or electoral concerns. Such analyses would be a worthy addition to the existing literature on policy feedbacks. In addition, these analyses show the policy relevance (and possibly the health relevance) of methodological analyses in institutional political science.

Conclusion

Analysis of new molecular entity review data from the last fifty years suggests that the deadlines of PDUFA and FDAMA have introduced large temporal discontinuities into FDA decision making, and that pre-deadline approvals are associated with substantially different post-marketing regulatory experiences than are other approvals, especially approvals that closely follow the elapsing of the deadlines.

Any analysis of this sort – no matter how sophisticated its methods – possesses all of the limitations of observational studies. Clearly there is no randomized or blinded assignment of drugs to the “treatment” of approval “before” a deadline. Furthermore, the measurement of postmarketing events and “issues” – withdrawals, warnings, and other postmarketing indicators – is far from an exact science. Such events are rare, although we employed statistical methods to account for this in robustness checks. While we believe that the methods elaborated here improve substantially upon those used by other analysts, it is not clear that

there is a single “best” empirical strategy for the analysis of deadlines and regulatory decision making, and even if there were, we do not claim to have found it.

References

- 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.
- 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.
- Carpenter, Daniel. 2002. “Groups, the Media, Agency Waiting Costs, and FDA Drug Approval.” *American Journal of Political Science*, 46(3): 490-505.
- Carpenter, Daniel, 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, and Michael M. Ting, “The political logic of regulatory error,” *Nature Reviews – Drug Discovery* 4(10) (October 2005): 819-23.
- Carpenter, Daniel, Evan James Zucker, Jerry Avorn. 2008. “Drug-Review Deadlines and Safety Problems.” *New England Journal of Medicine*, 358(13): 1354-1361.
- Carpenter, Daniel, Susan L. Moffitt, Colin D. Moore, Ryan T. Rynbrandt, Michael M. Ting, Ian Yohai, and Evan James Zucker. “Early Entrant Protection in Approval Regulation: Theory and Evidence from FDA Drug Review.” *Journal of Law, Economics and Organization*, forthcoming.
- 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.
- 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
- Huber, Oswald and Urs Kunz. 2007. “Time Pressure in Risky Decision-making: Effect on Risk Defusing.” *Psychology Science*, 49(4): 415-426.
- Kaufman, M., S. Vedantam. 2005. “Pregnant Women Warned by FDA to Avoid Paxil,” *Washington Post* December 9, 2005.

- 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.
- 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.
- Lexchin, Joel. 2005. "Drug Withdrawals from the Canadian Market for Safety Reasons, 1963–2004." *Canadian Medical Association Journal*, 172(6): 765-767..
- McCullagh, Peter and J.A. Nelder. 1989. *Generalized Linear Models*. London: Chapman and Hall.
- 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.
- Spence, David B. 1999. "Agency Discretion and the Dynamics of Procedural Reform." *Public Administration Review*, 59(5): 425-442.
- 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." Impact Report.
- 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: Semi-Parametric Approval Hazard Ratios, Standard Drugs

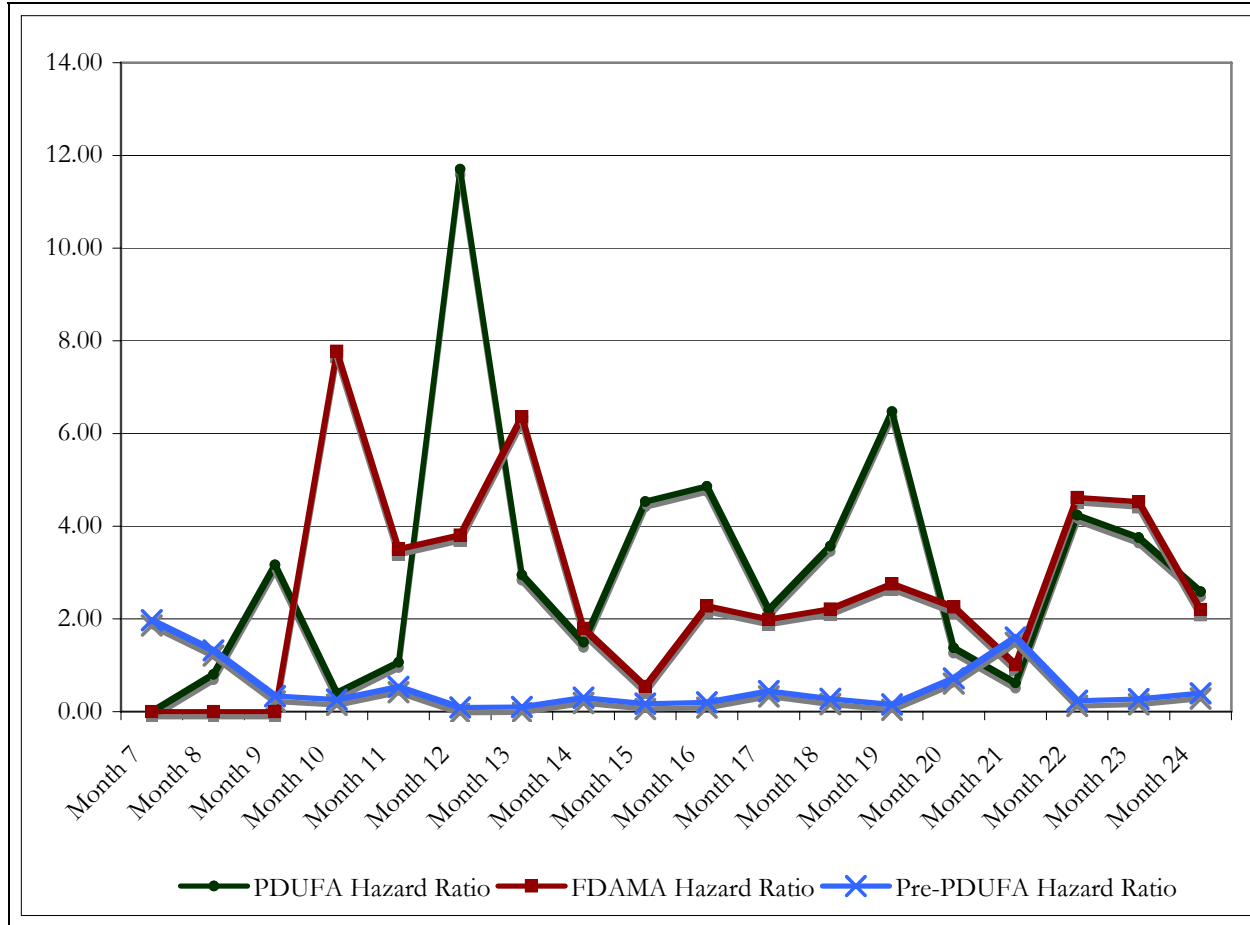


Figure 2: Approval Hazard Ratios for Priority NMEs before and after PDUFA

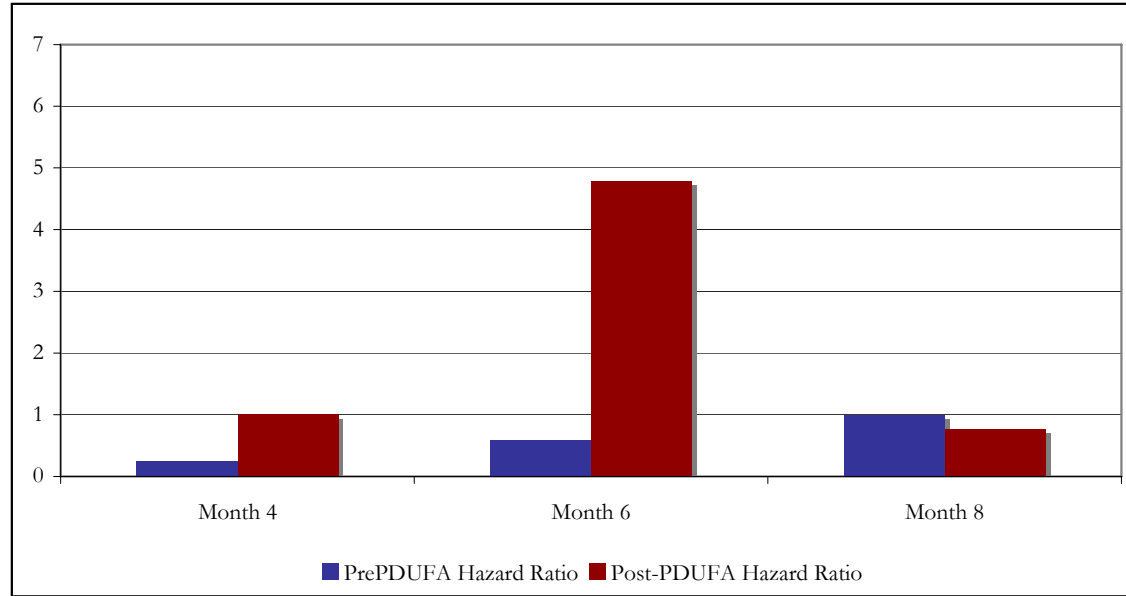


Table 1: Summary Statistics for Post-Marketing Regulatory Event (PMRE) Variables

STANDARD NMEs					
Variable	Valid NMEs	Mean	Std. Dev.	Minimum	Maximum
Safety-Based Withdrawal	581	0.024	0.153	0	1
Black-Box Warning Addition	563	0.020	0.139	0	1
Safety-Based Labeling Change (1996-2006)	581	1.439	2.673	0	19
Safety Alert (1996-2008)	581	0.396	1.124	0	10
Dosage Form Discontinuations	564	0.459	0.499	0	1
PRIORITY NMEs					
Variable	Valid NMEs	Mean	Std. Dev.	Minimum	Maximum
Safety-Based Withdrawal	518	0.015	0.123	0	1
Black-Box Warning Addition	511	0.022	0.145	0	1
Safety-Based Labeling Change (1996-2006)	518	1.693	3.311	0	17
Safety Alert (1996-2008)	518	0.475	1.217	0	8
Dosage Form Discontinuations	502	0.426	0.495	0	1

Table 2: Black Box Warnings, Random Effects Logistic Regressions

Variable	NMEs Approved 1950-2008		NMEs Approved Post 1979		NMEs Approved Post 1992	
	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)
Time Year of Submission	0.11 (0.04)	1.12 (0.04)	0.10 (0.04)	1.11 (0.05)	-0.06 (0.07)	0.94 (0.07)
Pre-Deadline Approved within 2 Months of Deadline	1.32 (0.50)	3.73 (1.85)	1.25 (0.52)	3.48 (1.82)	1.11 (0.48)	3.04 (1.47)
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			0.41 (0.20)	1.51 (0.31)	0.43 (0.20)	1.54 (0.30)
Drug Priority Priority Review = 1			0.14 (0.50)	1.15 (0.58)	0.45 (0.50)	1.56 (0.77)
Constant	-233.07 (73.46)		-213.75 (87.62)		115.52 (138.16)	
Number of Primary Indications		230		189		138
NMEs (number of observations)		1025		647		342
Log Likelihood		-84.74		-80.31		-73.41
Note: Bold-type connotes statistical significance at the $p < .05$ level (all tests are two-tailed); + indicates significance at the $p < .1$ level						

Table 3: Safety-Based Withdrawals, Random Effects Logistic Regressions

Variable	NMEs Approved 1950-2008		NMEs Approved Post 1979		NMEs Approved Post 1992	
	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)
Time Year of Submission	0.03+ (0.02)	1.03+ (0.02)	-0.04 (0.04)	0.96 (0.04)	-0.22 (0.10)	0.80 (0.08)
Pre-Deadline Approved within 2 Months of Deadline	1.20 (0.56)	3.32 (1.87)	1.53 (0.61)	4.60 (2.81)	1.63 (0.67)	5.12 (3.43)
Firms Firms' Previous Drug Approvals			0.04 (0.02)	1.04 (0.02)	0.04 (0.02)	1.04 (0.02)
NME Novelty Log Order of Drug Market Entry			0.11 (0.20)	1.12 (0.23)	0.19 (0.32)	1.20 (0.38)
Drug Priority Priority Review = 1			-0.52 (0.52)	0.60 (0.31)	0.01 (0.72)	1.01 (0.72)
Constant	-71.94+ (40.66)		69.15 (74.94)		429.84 (207.20)	
Number of Primary Indications		238		198		147
NMEs (number of observations)		1047		669		364
Log Likelihood		-99.68		-83.18		-48.50
Note: Bold-type connotes statistical significance at the $p < .05$ level (all tests are two-tailed); + indicates significance at the $p < .1$ level.						

Table 4: Safety-Related Label Changes, Random Effects Negative Binomial Regressions

Variable	NMEs Approved 1950-2008	NMEs Approved Post 1979	NMEs Approved Post 1992
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Time Year of Submission	0.07 (0.00)	0.02 (0.01)	-0.17 (0.02)
Pre-Deadline Approved within 2 Months of Deadline	-0.15 (0.14)	0.03 (0.14)	0.18 (0.11)
Firms Firms' Previous Drug Approvals		0.02 (0.00)	0.02 (0.00)
NME Novelty Log Order of Drug Market Entry		0.10 (0.05)	0.09 (0.06)
Drug Priority Priority Review = 1		0.10 (0.10)	0.30 (0.12)
Constant	-147.23 (9.44)	-30.51 (12.87)	335.18 (32.33)
Number of Primary Indications	232	191	140
NMEs (number of observations)	1030	652	347
Log Likelihood	-1436.34	-1324.45	-677.14
Note: Bold-type connotes statistical significance at the $p < .05$ level (all tests are two-tailed); + indicates significance at the $p < .1$ level			

Table 5: Safety Alerts, Random Effects Negative Binomial Regressions

Variable	NMEs Approved 1950-2008	NMEs Approved Post 1979	NMEs Approved Post 1992
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Time Year of Submission	0.04 (0.01)	0.01 (0.01)	-0.08 (0.03)
Pre-Deadline Approved within 2 Months of Deadline	0.37+ (0.20)	0.39+ (0.20)	0.40 (0.19)
Firms Firms' Previous Drug Approvals		0.03 (0.01)	0.02 (0.01)
NME Novelty Log Order of Drug Market Entry		0.15 (0.10)	0.17 (0.14)
Drug Priority Priority Review = 1		0.19 (0.17)	0.22 (0.22)
Constant	-75.63 (11.73)	-27.88 (23.28)	160.21 (50.65)
Number of Primary Indications	238	198	147
NMEs (number of observations)	1047	669	364
Log Likelihood	-784.28	-626.37	-372.64
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: Dosage Form Discontinuations, Random Effects Logistic Regressions

Variable	NMEs Approved 1950-2008		NMEs Approved Post 1979		NMEs Approved Post 1992	
	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)	Coefficient (SE)	Odds Ratio (SE)
Time Year of Submission	-0.06 (0.01)	0.94 (0.01)	-0.13 (0.01)	0.88 (0.01)	-0.14 (0.04)	0.87 (0.04)
Pre-Deadline Approved within 2 Months of Deadline	0.09 (0.29)	1.09 (0.32)	0.54+ (0.31)	1.71+ (0.53)	0.60+ (0.32)	1.82+ (0.59)
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			-0.02 (0.08)	0.98 (0.08)	0.10 (0.12)	1.10 (0.13)
Drug Priority Priority Review = 1			-0.33 (0.20)	0.72 (0.15)	-0.10 (0.31)	0.90 (0.28)
Constant	123.19 (11.09)		256.33 (29.75)		276.91 (84.17)	
Number of Primary Indications		235		196		359
NMEs (number of observations)		1015		648		146
Log Likelihood		-599.91		-348.19		-155.36

Table 7: Nearest Neighbor and Optimal Matching
(Using General Estimating Equations for Logistic or for Poisson)

Variable	Black Box Warning		Safety Withdrawal		Label Change		Safety Alert		Discontinuation	
	Nearest-Neighbor Matching	Optimal Matching	Nearest-Neighbor Matching	Optimal Matching	Nearest-Neighbor Matching	Optimal Matching	Nearest-Neighbor Matching	Optimal Matching	Nearest-Neighbor Matching	Optimal Matching
Constant	62.37 (110.31)	37.16 (82.39)	291.50+ (156.95)	345.67+ (178.64)	277.93 (38.29)	292.73 (30.65)	135.99 (53.57)	128.31 (56.03)	285.20 (89.36)	287.30 (90.09)
Year of Submission	-0.03 (0.06)	-0.02 (0.04)	-0.15+ (0.08)	-0.17+ (0.09)	-0.14 (0.02)	-0.15 (0.02)	-0.07 (0.03)	-0.06 (0.03)	-0.14 (0.04)	-0.14 (0.05)
Approved within 2 Months of Deadline	1.83 (0.71)	1.41 (0.47)	2.11+ (1.17)	1.40+ (0.79)	0.20 (0.15)	0.27 (0.12)	0.44 (0.28)	0.62 (0.24)	0.12 (0.36)	0.65 (0.32)
Number of Primary Indications	88	111	88	111	88	111	88	111	88	111
NMEs (number of observations)	172	258	172	258	172	258	172	258	172	258

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

All reported standard errors are robust.