Deadline Effects in Regulatory Drug Review: A Methodological and Empirical Analysis

Daniel Carpenter[+]
Jake Bowers[#]
Justin Grimmer[+]
Susan Moffitt[#]
Clayton Nall[+]
Evan James Zucker[@]

Version 2.2
July 2007

NOTE: THIS PAPER SERVES AS A METHODOLOGICAL COMPANION FOR ANOTHER PAPER. CERTAIN FIGURES, TABLES AND DATA THAT OVERLAP WITH THAT PAPER (WHICH IS UNDER SUBMISSION) HAVE BEEN WITHDRAWN FROM THIS VERSION.

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, dynamic duration models of time to decision are elaborated, 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 for observational analysis of whether the deadlines in question influence the "quality" of the reviews or the likelihood of error. In particular, we examine post-marketing regulatory events (PMREs) such as safety-based withdrawals and labeling changes, and we assess whether PMREs occur at a higher rate for drugs approved just before the deadline, compared to control sets of drugs approved at other times in the review cycle. 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, we find that these "pre-deadline" approvals are subject to substantially different post market experiences than drugs approved either after the deadlines or "very early" in the approval process.

+ Department of Government, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts.
# Robert Wood Johnson Foundation Scholar in Health Policy, Harvard University, Cambridge, Massachusetts.
@ Harvard Medical School, Boston, Massachusetts.

Prepared for presentation at the "Strengthening FDA" Workshop, February 22, 2007, at Program on Scientific Knowledge and Public Policy (SKAPP), George Washington University, Washington, D.C. We thank Jerry Avorn, Einer Elhauge, Gary King, Bruce Psaty, Meredith Rosenthal, Katherine Swartz and Heidi Williams for critical remarks and helpful discussions. 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 2004, Whitford 2006), there has been very little analysis of how political and administrative institutions may affect the timing and character of regulatory decisions. To the extent that institutions have been represented at all in such research, they have entered 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 the influence of deadlines upon 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 were 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. In both cases, we are not creating new statistical methods, but we are 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 and 2002, imposed quite specific and 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), and 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).

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.

This paper serves as a technical companion to a shorter paper, Carpenter and Zucker (2006), in which statistical evidence is presented for highly abrupt changes in FDA drug approval probabilities in the months just before and just after statutory deadlines occur. In addition to replicating and extending the results of that paper – showing that drugs approved just before the deadlines have substantially different post-marketing experiences than those approved just after – we also augment generalized linear models with other forms of observational analysis (including 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 and 2002. 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. We then turn at least to generalized linear models for assessing the influence of the deadlines upon the quality of FDA choices, and examine the correlation of deadlines with post-marketing regulatory events such as drug withdrawals, relabeling, package and manufacturing revisions and other observable post-approval events.

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 new drugs were brought under its provisions starting September 1, 1992. User-fee reauthorization came in an act that made other procedural changes to the FDA, in the Food and Drug Administration Modernization Act (FDAMA) of 1997. New drugs were subsequently governed by that act if they were submitted on or after October 1, 1997. Most recently, the user-fee program was re-authorized in 2002 as part of an omnibus package related to bioterrorism. Because its provisions expire every five years unless explicitly re-authorized, U.S. policymakers will be revisiting the user-fee program in 2007 at the latest. The three user-fee laws have made many changes at the FDA, and among the most important of these is the provision 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.[1]

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.[2] 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% or more) of new molecular entities (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 [8]. The clock differed according to whether the new drug application was designated "priority" or "standard," as follows.[3]

---

[1] 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).

[2] 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."

[3] 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/fdamaplnresponse/rptgoalsFY99.html (accessed October 16, 2005).

PDUFA, 1992 (began 9/1/1992): by 1997, review and act upon 90% of standard NDAs in 12 months, 90% of priority NMEs in 6 months.

FDAMA, 1997 (began 10/1/1997): by FY 1999, 30% of standard NDAs in 10 months, by FY 2002 90% of standard NDAs in 10 months; same as PDUFA for priority NMEs.[4]

"PDUFA III," 2002 (began 10/1/2002): For standard and priority NDAs, same deadline months as in 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 and 2002 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.

---

[4] Indeed FDA officials trumpeted the fact that the agency 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%, meaning that over half of NDAs were meeting the 10-month deadline goal. FDA Commissioner's Office, "Performance on FY 99 FDAMA Goals," http://www.fda.gov/oc/fdama/fdamaplnresponse/rptgoalsFY99.html (accessed October 16, 2005).

II. Statistical Decomposition of the FDA Review Cycle, before and after PDUFA.

In order to test our hypotheses, we must retrieve statistical estimates of the conditional probability of approval at each point of the FDA 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 drug 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.[5]

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.

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.

$$\mathrm{AHR}_{\tau=t} = \frac{h_{\tau=t}(\beta' X)}{h_{\tau=1}(\beta' X)} \qquad \textbf{(1)}$$

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 the different institutional changes, we can also examine

[5] 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.

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:

$$\mathrm{AHR}_{\tau=t} = \frac{h_{\tau=t}^{REFORM=1}\left(\beta' X\right)}{h_{\tau=t}^{REFORM=0}\left(\beta' X\right)}$$

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

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

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

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

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

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

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

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

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 $y_i(\beta,t)=\exp\left[X_i(t)\beta\right]$, and let *Y*(*s*) be the aggregate number of drugs still under review (or at "risk" of getting approved) at time *s*. Then

$$\bar{x}(\beta,s)=\frac{\sum_{j=0}^{Y(s)}R_i(s)\,y_i(s)X_i(s)}{\sum_{j=0}^{Y(s)}R_i(s)\,y_i(s)}$$

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

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

For various reasons of information and computational feasibility, the inverse of the observed information matrix $I^{-1}(\beta)$ is used in lieu of the inverse of the *expected* information matrix $\left\{E_{nt}I(\beta)\right\}^{-1}$, even though the latter is analytically kosher (Therneau and Grambsch 2000: 40-41). The maximum partial likelihood estimator is obtained by solving the partial likelihood equation $U\left(\hat{\beta}\right)=0$. This is done via a Newton-Raphson algorithm, which iteratively computes $\hat{\beta}^{(n+1)}=\hat{\beta}^{(n)}+I^{-1}\left(\hat{\beta}^{(n)}\right)U\left(\hat{\beta}^{(n)}\right)$ until convergence is reached. With dynamic Cox estimation, order months in *X* such that the τ^{th} month is the τ^{th} element of β. 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

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

Let υ and ω be any two months in the review cycle, such that $\upsilon > 0, \omega > 0, \upsilon \neq \omega$. Then we can represent restrictions by $q\left(\hat{\beta}\right) = g$ such that $\hat{\beta}_{\tau=\omega} = \hat{\beta}_{\tau=\upsilon}$, 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.

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

In Appendix A we display sample estimates from Cox estimation using S-Plus output. The code is repeated as are the estimates.

An Application to FDA Review Times under Deadlines.

In Figure 1 we plot monthly approval hazard ratios and their standard errors 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 and twelfth months was not statistically differentiable from that in the *third* month ($\chi^2 = 0.46$; $p = 0.4987$), the *sixth* month ($\chi^2 = 0.08$; $p = 0.7733$), the *tenth* month ($\chi^2 = 0.15$; $p = 0.6957$), or the *eighteenth* month ($\chi^2 = 0.05$; $p = 0.8154$). Moreover, there was no statistically detectable difference between the approval hazard of the 11th and 12th months of review, and the approval hazard in the 13th and 14th months of review ($\chi^2 = 0.00$; $p = 0.9962$).

Our analyses suggest that the PDUFA clocks have dramatically changed the behavioral structure of the FDA 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 and twelfth month of review compared to the same months before PDUFA ($\chi^2 = 6.91$; $p = 0.006$). Moreover, as hypothesized, approval hazards fall off appreciably for the two months after the review clock deadline ($\chi^2 = 6.93$; $p = 0.0085$). 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 two months before the ten-month FDAMA clock deadline is *twelve* times greater than the approval hazard in the month after the review deadline has elapsed ($\chi^2 = 11.58$; $p = 0.0006$).

Figure 2 shows that these patterns apply to priority drugs as well, where the deadline has been stable at six months over the 1993-2004 period.

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. 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.[6] While some analysts (including the FDA itself) have examined whether the *overall* rate of drug safety problems has risen or fallen since 1993,[7] 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 area. 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 six measures of post-marketing regulatory events. The first four (two measures of safety-related labeling revisions and black-box warnings, and two measures of safety-based withdrawals) are directly concerned with safety issues. Three other variables (significant labeling revisions, labeling revisions for changes in patient population, and manufacturing changes) are often but not always concerned with issues of safety. 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 and labeling revisions. We use two separate measures of postmarketing safety-related labeling revision. Our first measure is whether or not the approved drug has received a black-box warning for a 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.

---

[6] 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.

[7] 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.

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.

Our second measure is a list of post-approval drug safety warnings compiled by physicians and epidemiologists at the Kansas University Medical Center (KUMC).[8] The KUMC list is larger than the Lasser list, most likely because the KUMC list retrieves information directly from the Food and Drug Administration, whereas Lasser and co-authors require a change in the *Physician's Desk Reference* in order to code a drug as having been given a black box warning for a newly recognized ADR.

Safety-based withdrawals. We examine safety-related market withdrawals from the global market. Our data on withdrawals are from two sources. First, from SCRIPS reports and Pharmaprojects, we identified 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 just one country.

Second, we also examine safety-related withdrawals from Canada since 1963. Lexchin (2002) 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 examine a wider sample of withdrawals rather 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."[9] 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 U.S. withdrawals are.

Changes in patient population on officially approved labeling. From the Drugs@FDA database (accessed December 2004), we calculated the number of official alterations in patient population for each new molecular entity, and divided these by the number of years

---

[8] 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.

[9] 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.

that the drug had been on the U.S. market. It is important to remember that not all these changes are safety-related. But clearly some important ones are. For instance, CDER recently made effective a patient population change for Avandia (rosiglitazone), based upon poor efficacy results for a sub-population. The official announcement reads: "Provides for changes to the labeling describing the results of a study comparing the effects of Avandia to those of metformin in children with type 2 diabetes mellitus, aged 10-17 years. An indication for the use of Avandia in this population is not supported by the results of the study."

Manufacturing process revisions – From the Drugs@FDA database (accessed December 2004), we calculated the number of official revisions in manufacturing process for each new molecular entity, and divided these by the number of years that the drug had been on the U.S. market.

Product (dosage form) discontinuation. What does discontinuation imply? From 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.[10] Discontinuation is a code ("3") in the Product Market Status of approved products, as tracked by CDER, and available at Drugs@FDA website. It tracks discontinuation of particular dosage and administration versions of an NDA. We note that such discontinuations are not explicitly or officially safety or efficacy-related. They are, however, positively correlated with global market withdrawals [probit coefficient = 2.63 (0.98); $p = 0.008$]. 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.

Another possibility would be to examine adverse event reports (AERs), which have received some study in recent years (Olson 2004). 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.

[Table 1 about here.]

[10] 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).

We report summary statistics for these measures in Table 1. As anticipated, all of the measures are characterized by rarity, though the converted annual rate data are characterized by greater continuity.

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 two vectors of parameters for estimation – a set of terms for each primary indication (modeled either as a fixed or a random effect) and a set of indicator variables for the firm sponsoring the application. The immediate result of this estimation strategy is that hundreds of terms and parameters are added to the models we estimate.

We employ the generalized linear model (GLM) framework (McCullagh and Nelder 1992) for panel data and mixed effects models. Recall that for any drug *i*, its primary indication is indexed by ψ and its sponsor by κ. Recall, too, that S_κ and D_ψ 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

$$y^{PMRE}_{\psi\kappa i} = f\left(\alpha^S S_\kappa + \gamma' Z_{\psi\kappa i} + u_\psi + e_{\psi\kappa i}\right) \quad \textbf{(5a)}$$

or

$$y^{PMRE}_{\psi\kappa i} = f\left(\alpha^S S_\kappa + \alpha^D D_\psi + \gamma' Z_{\psi\kappa i} + e_{\psi\kappa i}\right), \quad \textbf{(5b)}$$

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 and α^D are fixed effects coefficients, and where *e* is a model disturbance. Notice that primary-indication-specific effects are modeled either as random effects or as fixed effects.

For "mixed-effects" models which embed a random-effect term for the primary indication and a hierarchical structure, we employ methods described by Gelman and Hill (2007) for the analysis of our data. In particular, we analyze our models by Markov Chain Monte Carlo posterior distribution sampling. This technique allows for diagnostic analysis of the convergence of the maximum likelihood algorithm and also permits an analysis of overfitting and parameter estimate stability. (All linear mixed-effects models are estimated in R (or S-Plus) using the `lmer` functions authored by Bates and the `mcsamp` post-estimation commands of Gelman.) We do not report these estimates here but they are available on-line at http://people.hmdc.harvard.edu/~dcarpent/pdufaclock-mcsamp-run20070701.pdf (accessed July 23, 2007).

We note here one important feature of our modeling strategy, namely its assignment of sponsor-specific and indication-specific terms to each drug. Our indexation of primary indications is much more refined than that used by other analysts of FDA drug approval (e.g., Lasser 2002, Olson 1997, Carpenter 2002, Olson 2004). 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) and often for continuously valued firm-level covariates, but not for particular primary indications and not for firm-specific fixed effects. Because drugs are assigned to divisions based primarily upon their primary indication, the primary indication index ψ 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.

We acknowledge, of course, that our modeling strategy also embeds the disadvantage that too many terms may be introduced into estimation. We have estimated the models with coarser categories and these models generally produce larger coefficients and smaller standard errors for our variables of interest (indicators for pre-deadline approvals). In this sense, then, the estimates we report are "conservative" in the sense that larger and more robust estimates are possible.

For rare events such as safety-based withdrawals, we supplement our models with extreme value regressions which account for the infrequency of these events. In this case the link function of the GLM is a Gompertz or Gumbel extreme-value link function. Where we can retrieve standard errors from these regressions, we do so. We then assess differences in event rates from pre-deadline approvals to post-deadline approvals. One advantage of the multivariate analyses is that we can control statistically for a wide variety of indication- and therapy-specific indicator variables, as well as for effects of the sponsoring firm. We can also directly compare the rate of postmarketing regulatory events from before and after the various user fee acts, whereas cross-tabulations require separate analyses.

For each of the regulatory event variables, we regress the regulatory event variable on a battery of indicator variables, the submission year (to capture the time trend) and selected other continuous measures. Because we estimate coefficients for the deadline terms, our tests for the user-fee deadline effects are embedded within a larger statistical specification. We 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 one 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 one.

2. Post-Deadline Approval. Similarly, for any deadline month $\tau^{deadline}$, we construct a "post-deadline" approval indicator Z^{POST} scored one if the drug in question was approved in $\tau^{deadline}+1$ or $\tau^{deadline}+2$, and 0 otherwise. Where the deadline is 12 months, for instance, then approvals in the 13th and 14th month are coded one.

3. Month Indicators. We also create indicator variables for approvals during given months as a form of control. When for instance the deadline is at 12 months, we also seek to identify those drugs that were approved in the eleventh and twelfth months of drug review before PDUFA. This permits a comparison of (a) drugs approved in the 11th and 12th months when these months were pre-deadline months to (b) drugs approved in the 11th and 12th months where no deadlines were attached (before September 1992).

4. "Early Approval" Indicators. We also seek to compare post-market regulatory event rates for drugs that are approved far earlier than the review time goals. For standard NMEs, we construct an "early approval" indicator Z^{EARLY} scored one if the drug in question was approved in the first six months after submission and zero otherwise. For priority NMEs the same variable is scored one if the drug was approved in the first four months after submission.

5. Cox-Weighted Deadline Indicators. Because our statistical analysis of FDA review times allows for empirical estimation of approval hazard ratios, we can also weight the deadline variables by the hazard ratio appropriate to the particular month of approval. This results in a continuously-valued deadline assignment measure, which is computed as

$$Z^{\text{PRE-COX}} = \frac{\hat{h}_{\tau^{deadline}}(X,\beta)}{h_{\tau=1}(X,\beta)} 1\left[\tau^{deadline}\right] + \frac{\hat{h}_{\tau^{deadline}-1}(X,\beta)}{h_{\tau=1}(X,\beta)} 1\left[\tau^{deadline}-1\right] \qquad \textbf{(6a)}$$

Where $1[\cdot]$ is the indicator function scored 1 if the drug in question was approved in the pre-deadline months in $\tau^{deadline}$ or $\tau^{deadline}-1$ and scored zero otherwise. Similarly, the post-deadline condition can be weighted as

$$Z^{\text{POST-COX}} = \frac{\hat{h}_{\tau^{deadline}+1}(X,\beta)}{h_{\tau=1}(X,\beta)} 1\left[\tau^{deadline}+1\right] + \frac{\hat{h}_{\tau^{deadline}+2}(X,\beta)}{h_{\tau=1}(X,\beta)} 1\left[\tau^{deadline}+2\right] \textbf{(6b)}$$

Before versus After Deadline Comparisons. Our most common comparison is between NMEs approved in the two months before the deadline and NMEs approved in the two months afterwards. For instance, for NMEs submitted to CDER under the first PFUFA Act (September 1, 2002 to September 30, 1997), we can code non-priority drugs approved in the eleventh and twelfth months of the review cycle as "pre-deadline" approvals, and we can code drugs approved in the thirteenth and fourteenth months of review as "post-deadline" approvals. Because the semi-parametric analyses in the previous section point to the two months before the deadline as the time when relative hazards are heightened, our comparison is based upon the empirical and statistical findings in the previous section. However, we have also examined the immediate month before and month after the deadline and have arrived at similar results.[11]

[11] Another possibility is measuring "time-to-deadline" in weeks and days, but given that the user-fee legislation encourages the FDA (and companies) to think in terms of months, we have avoided this measure. Estimating week-specific parameters leaves us with too little data to estimate parameters for every week or every month of the review cycle, especially for drugs submitted after 1993.

Using the Z^{PRE} and Z^{POST} terms, we can compute effect estimates γ^{PRE} and γ^{POST} in the GLM framework. Retrieving these estimates and associated information from the estimated covariance matrix, we can test for the equality of γ^{PRE} and γ^{POST} by computing the following Wald-like statistic.

$$W = U'\left(q_{\left[\hat{\gamma}^{PRE}=\hat{\gamma}^{POST}\right]}(\hat{\gamma})-\hat{\gamma}\right) I\left(q_{\left[\hat{\gamma}^{PRE}=\hat{\gamma}^{POST}\right]}(\hat{\gamma})-\hat{\gamma}\right)^{-1} U\left(q_{\left[\hat{\gamma}^{PRE}=\hat{\gamma}^{POST}\right]}(\hat{\gamma})-\hat{\gamma}\right) \quad \textbf{(7a)}$$

where U is the "discrepancy" vector, q is a set of restrictions, and I is the information matrix or the covariance matrix. Letting $\mathbf{C}\hat{\boldsymbol{\gamma}} = \mathbf{q}$ be the set of linear restrictions in which γ^{PRE} and γ^{POST} are held identical, the discrepancy vector is $\mathbf{C}\hat{\boldsymbol{\gamma}} - \mathbf{q}$.

Deadline Month Approval versus "Early" Approval Comparisons. A similar test statistic can be computed for comparing whether the PMRE rate is higher for pre-deadline approvals than among even earlier approvals, as follows:

$$W = U'\left(q_{\left[\hat{\gamma}^{PRE}=\hat{\gamma}^{EARLY}\right]}(\hat{\gamma})-\hat{\gamma}\right) I\left(q_{\left[\hat{\gamma}^{PRE}=\hat{\gamma}^{EARLY}\right]}(\hat{\gamma})-\hat{\gamma}\right)^{-1} U\left(q_{\left[\hat{\gamma}^{PRE}=\hat{\gamma}^{EARLY}\right]}(\hat{\gamma})-\hat{\gamma}\right) \quad \textbf{(7b)}$$

Deadline Month versus pre-User-Fee (same month) Comparisons. One possibility, of course, is that some unobserved factor that just happens to be associated with deadline-month approvals would produce these results. As this is an observational study, we cannot eliminate this possibility. We can, however, perform several checks on the data. First, we can conduct the same "before-and-after" comparison *as if* the deadlines were in place before 1992, as a check upon our analyses. In other words, we can code approvals in the 11th and 12th months *before* the user-fee era as "pre-deadline" approvals, even though no user-fee review clock was in place, and we can compare postmarketing regulatory events for these drugs and drugs approved in the 13th and 14th months.[12]

Accounting for Separation Bias: Alternative Procedures using Firth's Bias Correction, Efficient Monte-Carlo Estimation, and Exact Logistic Regression for smaller samples. In the case of withdrawals and

---

[12] In other papers we consider additional statistical analyses to deal with this issue. In particular, in Bowers, Carpenter, et al. (2007) we conduct a different non-parametric matching analyses using robust and genetic matching techniques, where drugs are matched on epidemiological characteristics of their primary indication (incidence of primary indication, hospitalization rate of primary indication, age-adjusted death rate of primary indication, and various indicators of broadcast media coverage given to primary indication, and the size of CDER's staff at the time of the NME's submission).

Finally, for illustrative purposes we cross-tabulate the regulatory event variables in question – withdrawals, warnings, discontinuations – with a binary variable coded "1" if the drug in question was approved before the deadline and scored "0" if a "control" condition was met. The control condition can be "all other approvals" or it can be a "post-deadline" approval. We then employ an exact probability test to test the hypothesis of a difference in regulatory event rates before and after the deadline approvals. For some of the variables we also employ extreme-value regressions because the number of "positive" events is so small.

black-box warnings, our outcome measures are binary, and because our independent variables are also binary, this can lead to separation bias for datasets without balance (King and Ryan 2002, Mehta, Patel and Senchaudhuri 2000). In separation bias, the prefect prediction of a subset of the binary response by a vector or subvector of the independent variables (the **Z**'s, in our GLMs), leads to estimates of parameter effect that are biased away from zero (see Firth 1993, King and Ryan 2002). We use three methods to address this problem. The first two of these methods employ larger samples and avoid exact methods, while the third uses exact logistic regression.

The first method – the bias correction introduced by Firth (1993) – is to employ an estimator for larger samples where exact logistic regression is infeasible and where efficient Monte Carlo estimation (Mehta, Patel and Sanchaudhuri 2000) is also computationally infeasible. We classify either of these methods as computationally "infeasible" when the estimated computation time was greater than 240 hours (ten days).[13] These models are binary logistic regression models, and the methods and penalized likelihood function are as reported in Firth (1993).

Second, when possible we use the efficient Monte Carlo method for logistic regression developed by Mehta, Patel and Senchaudhuri (2000). This estimator has the advantage of having been used for observational data (see the examples using fraudulent automobile insurance claims in Mehta, Patel and Senchaudhuri (2000)). This method becomes computationally infeasible in the presence of many covariates, so we employ it only for models with few covariates. By necessity, this excludes the fixed-effects and random-effects models estimated in the GLM framework, and noted above. For details of the computational method, see Mehta, Patel and Senchaudhuri (2000: 99-103).

Third and finally, we occasionally draw up small samples for exemplary use and use exact logistic regression (Mehta and Patel 1995; Tritchler 1984). We performed all three separation-adjustment methods – (1) MLE with Firth's penalty-adjusted likelihood bias correction, (2) efficient Monte Carlo estimation of logistic regression, and (3) exact logistic regression – with LogXact7™ Software (Cambridge, Massachusetts, Cytel Corporation, 2006).

Results: GLM Estimation

[Tables 2A, 2B, 2C, 2D, and 2E about here.]

We report the generalized linear model results in Tables 2 and 3. Tables 2A through 2E report estimated GLM models for post-marketing regulatory events for non-priority ("standard") NMEs, while Tables 3A through 3D report models for priority NMEs. The GLM estimates are, in general, rather stable across specifications. Across the replications for

[13] Estimated computation times are reported by LogXact software. Mehta, Cytel, and Senchaudhuri (2000) report that their applications "were performed on a Pentium 200 PC" and that "Computation times for network building plus Monte Carlo sampling ranged between 1 and 20 minutes" (2000: 103) Hence our cutoff for "feasibility" lies well above the most arduous models computed by Mehta, Cytel, and Senchaudhuri (2000). For most of the models we chose as computationally infeasible, LogXact reported estimated computation times of 1000 hours or more.

standard NMEs, coefficient estimates change signs across models only four times (out of 104 estimates).

Beyond this, the coefficient estimates for pre-deadline indicators are also consistent in their behavior. Coefficient estimates for pre-deadline approvals are positive in 27 of 34 estimates for standard NMEs, are positive and statistically significant in 16 of these 34 estimates, and are negative in 7 estimates and negative and significant in three. (All three of these occur in one regression and for one term: the pre-deadline FDAMA variable for postmarket manufacturing revisions.) Coefficient estimates for pre-deadline indicators are positive in 14 of the 16 estimates for priority drugs, and seven of these estimates are statistically significant. Across both categories, then, 41 of 50 estimated pre-deadline effects are positive, 30 of 50 are positive and statistically differentiable from zero and 3 are negative and statistically differentiable from zero.

Comparison of Pre-Deadline with Post-Deadline Approvals: Results from GLM Models. Comparison of pre-deadline and post-deadline approvals is conducted by computing the relevant test statistic. For each model, this statistic is reported in the second portion of the table ("Parameter Estimate Comparisons") with its associated probability below it. For standard NMEs, pre-deadline approvals are significantly more likely to experience dosage-form discontinuation (particularly for FDAMA deadlines), postmarket patient population label changes (for PDUFA deadlines), postmarket manufacturing revisions (for PDUFA deadlines), and Canadian and global safety-based withdrawals.[14]

Comparison of 11th-month and 12th-month Approvals under PDUFA and 11th-month and 12th-month Approvals Later Regimes. Another quasi experiment is possible if we examine the set of drugs approved in the eleventh or twelfth months of the review cycle since the enactment of PDUFA in 1992. Under the first user-fee act (September 1992 to September 1997) these were "pre-deadline" approvals. With FDAMA and subsequent legislation (October 1997 and after), however, these are "post-deadline" approvals. Our hypothesis suggests that 11th and 12th month approvals under PDUFA I should be more likely to experience post-market regulatory events than 11th and 12th month approvals in the years since.

If one finding emerges most consistently from our estimates, it is that the PMRE rate for 11th and 12th month approvals under the first user-fee act is significantly higher than the PMRE rate for 11th and 12th month approvals under the user-fee acts since September 1997. Eleventh and twelfth month approvals are more likely to experience dosage-form discontinuation, patient population shifts, postmarket manufacturing revisions, and safety-based withdrawals ($p < 0.03$ in nine of ten tests). While we cannot rule out other changes between PDUFA I and PDUFA II, these results are highly suggestive that pre- versus post-deadline approval status is correlated with post-market issues.

[14] For postmarket patient population labeling changes, the differential is statistically significant in the model with fixed effects ($p = 0.0081$) but lies just above the $p = 0.05$ threshold for the random effects model ($p = 0.0507$). The statistical significance of these differentials is also model dependent for safety-based withdrawals in Tables 2D and 2E.

Comparison of Pre-Deadline with Pre-PDUFA approvals of the same cycle month: Were these differential patterns of post-marketing regulatory events in evidence before the user-fee acts? In other words, were there abrupt changes in the rate of postmarketing regulatory events depending on whether a drug was approved in (1) the 9th and 10th months, (2) the 11th and 12th months, or (3) the 13th and 14th months? Our analyses for non-priority drugs approved before September 1992 suggest not. These tests appear in the third and fifth rows of the "Parameter Estimate Comparisons portion of Tables 2A through 2E. In only two cases (of thirty nine possible) do the relevant categorical variables reach statistical significance ($p > 0.2$ in all other regressions), and PMRE rate differences among 9th and 10th month approvals, 11th and 12th month approvals, and 13th and 14th month approvals are not evident. For priority drugs, similar patterns are evident, as 5th- and 6th-month approvals were, before the user-fee acts, if anything less likely to experience postmarket regulatory issues than approvals in other months.

Comparison of Pre-Deadline with "Early" Approvals. The possibility remains that the behavioral sorting of drugs into pre-deadline approval versus post-deadline approval is based on higher perceived efficacy, and in fact that drugs with higher efficacy are characterized by greater post-marketing regulatory events. Again we cannot rule out this possibility, though we note that the greater incidence of dosage-form discontinuations among pre-deadline approvals sheds doubt upon this rival explanation. If pre-deadline approvals are being discontinued from clinical utilization and the "marketplace" at higher rates, it is more difficult to sustain the hypothesis that pre-deadline approvals have higher therapeutic value than those approved afterwards, at least in the short term.

A more compelling test, perhaps, is to examine those drugs that are approved "very early" – within the first six months for standard NMEs, and within the first four months for priority NMEs – and to ask whether these "quick approvals" are also characterized by postmarketing issues. These tests appear in the bottom rows of the tables, which report F-statistics and chi-squared statistics for the equivalence of γ^{PRE} and γ^{EARLY}. For standard NMEs, equivalence of "early" and "pre-deadline" approval PMRE rates can be rejected for dosage-form discontinuations, for postmarket manufacturing revisions (PDUFA deadlines) for Canadian safety withdrawals (PDUFA deadlines). For priority drugs, inference is more complicated because in no case did an approval in the first four months of the review cycle for a priority drug in the last twenty years experience a global or Canadian market withdrawal. From maximum likelihood models with categorical data, the relevant test statistic is infinite and no reference probability can be retrieved. Using a linear model, the equivalence of γ^{PRE} and γ^{EARLY} can be rejected in global safety withdrawals. Equivalence of pre-deadline and early approval PMRE rates can also be rejected for black-box warnings as coded by Lasser and colleagues (2002).

More broadly, there are only two cases – both black-box warnings, as coded by Lasser and colleagues (2002) and KUMC, respectively – where the value of the early-approval coefficient rests numerically above the deadline coefficients. In neither of these cases is the difference statistically significant. Put differently, we can in numerous cases accept the null of equivalence in favor of the alternative hypothesis that the pre-deadline approval PMRE rates are higher than "early" approval PMRE rates. We cannot once reject the null of

equivalence in favor of the alternative hypothesis that the early approval PMRE rates are higher than pre-deadline approval PMRE rates.

[Tables 4A-4D about here.]

Results from Efficient Logistic Regression and Firth-Corrected Logistic Regression. As an additional check on our results, we estimate reduced models using methods for addressing separation bias and small-sample inference problems. In Tables 4A through 4D, we report the results of logistic regressions where either Firth-corrected maximum likelihood estimation or efficient Monte Carlo estimation is used. In one case (Canadian safety withdrawals) we are able to stratify the sample so as to produce "informative" strata (with a smaller sample size) and we report an additional set of results for that estimation (Table 4A-2).

In all of these tables, we have combined all standard and priority NMEs and have generated one "pre-deadline approval" variable, scored one if (a) the drug is a standard NME and was approved in the 11^{th} or 12^{th} month of the review cycle under PDUFA, or if (b) the drug is a standard NME and was approved in the 9^{th} or 10^{th} month of the review cycle under FDAMA or later user-fee laws, or if (c) the drug is a priority NME and was approved in the 5^{th} or 6^{th} month of the review cycle under any of the user-fee laws. (The "pre-deadline approval" variable is scored zero if none of the conditions (a)-(c) hold.) We add the year of NME submission as a control variable to these regressions.

The combination of the different deadlines into a single variable assumes no heterogeneity in the different deadline effects. In other words, the twelve-month deadline and the ten-month deadline are assumed equal in effect. This assumption does not appear to be unwarranted. The coefficients for the different deadline measures in Tables 2A through 2E can rarely be differentiated statistically. Only for Canadian withdrawals are the different deadlines statistically distinct as assessed by Wald test.

In Tables 4A-1 and 4A-2, the regressions show a significant positive association between pre-deadline approvals and safety withdrawals in Canada, using a sample of NMEs approved in the U.S. from 1962 to the present. The sample size is smaller for the stratified estimation as "noninformative" strata are deleted. In Table 4B, similar results are shown for global safety withdrawals since 1980. Here the Firth penalized likelihood method results in an estimate on the pre-deadline approval variable that is statistically differentiable from zero at the $p < 0.05$ level, but the Monte Carlo method results in an estimate that is statistically non-zero only at the $p < 0.10$ level.

For the black-box and safety warning variables, we observe positive but statistically insignificant coefficients for the analyses of warnings as coded by Lasser and colleagues (2002) (Table 4C), and positive and statistically significant coefficients for warnings as coded on the KUMC list.

Two features of the estimations in Table 4A through 4D are worth noting. First, notice that, as claimed by Mehta and Patel, analysis of rather large samples is possible with efficient Monte Carlo logistic regression. These are sample sizes that are infeasible for exact logistic

regression. Second, notice that Monte Carlo (CMLE) standard errors are larger and corresponding *p*-values are smaller than the respective estimates for Firth-corrected MLE.

[Table 4E about here.]

Results from Exact Logistic Regression. Finally, for subsamples where the sample size allows for estimation of an exact logistic regression, we have created a single "pre-deadline" approval variable for all standard NMEs and have restricted our attention to NMEs submitted since January 1993. In Table 4E, we report results from exact logistic regression for safety-based withdrawals (Canadian, 1962-present, and global, 1981-present) and dosage-form discontinuation. For all three binary variables, a pre-deadline approval is positively associated with the probability of withdrawal, even as each form of withdrawal has become less likely in terms of the overall time trend. Exact *p*-values are larger than asymptotic *p*-values, but all are below the 0.05 level for these three withdrawal measures. We used Cytel LogXact7 software to produce these estimates, and the output for these estimations is available upon request.

We recognize that the reported exact and efficient logistic regressions are rather brute in that one deadline variable is used and there is only one covariate. Monte Carlo efficient and exact logistic estimators compel these sorts of restrictions unless the analyst is willing to allow hundreds or thousands of hours for estimation of the models to converge. We recognize that more covariates are possible, and for analyses with many more covariates (hundreds more, in some cases) we simply direct the reader to Tables 2 and 3.

Statistical Summary, Discussion and Conclusion

[Table 5 about here.]

A Summary. We now turn to summarize the analyses reported in Tables 2A-2E and 3A-3D. We do so in Table 4, which collects summary results for standard drugs. We first calculate differentials in rates of regulatory events. We do so by computing model coefficient differences, subtracting the "control" condition coefficient from the pre-deadline approval coefficient, and then dividing this differential by the mean of the relevant PMRE variable. Specifically, we compute the following differential for pre-deadline approvals versus post-deadline approvals

$$Diff^{PMRE}_{PRE\prime POST} = \frac{\hat{\gamma}^{PRE} - \hat{\gamma}^{POST}}{\overline{y}^{PMRE}_{\psi\kappa i}}$$

We have already tested the hypothesis of difference between γ^{PRE} and γ^{POST} by computing statistics of the form *W* as described above (equations 7a and 7b). Using the $Diff^{PMRE}_{PRE\prime POST}$ statistic, we comparing [1] the 11th and 12th month (pre-deadline) approvals to the 13th and 14th month (post-deadline) approvals for PDUFA (1993-1997), [2] the 11th and 12th month (pre-deadline) approvals for PDUFA to the 11th and 12th month (post-deadline) approvals for FDAMA, [3] the 11th and 12th month (pre-deadline) approvals for PDUFA to the 11th

and 12th month approvals before any user-fee act (before 1993), and [4] the 9th and 10th month approvals under FDAMA to the 11th and 12th month approvals under FDAMA.

Results from these computations appear in Table 5.

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, especially those approved in the months just following the elapsing of the deadline. For non-priority molecules, pre-deadline approvals are associated with three to five times the rate of safety-based withdrawal from the global market and Canadian markets. Pre-deadline approvals have two to three times as patient population labeling changes per year of marketing ($p < 0.05$) and, for drugs approved since FDAMA, over five times the rate of product discontinuations per year ($p < 0.0001$).

Again, some of the most compelling results come from the second row of differential estimates in Table 4, where the event rates for 11th- and 12th-month approvals are compared from the first PDUFA law, when these months indicated pre-deadline approvals, to the second FDAMA law (after 1997), when the 11th and 12th month approvals fall after the deadline. In other words, when the eleventh and twelfth months of the review cycle are *pre*-deadline months (1993-1997), they are associated with a higher rate of postmarketing issues. However, when the same eleventh and twelfth months of review are *post*-deadline months (1998-present), they are associated with much *lower* rates of postmarketing regulatory events. ($p < 0.0001$ for global withdrawals, dosage-form discontinuation and manufacturing changes; $p < 0.01$ for Canadian withdrawals and patient population changes). Here the differentials are positive and strikingly large in all eight measures and statistically significant in five of eight measures.[15]

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 almost always positive partial correlations. That is, we observe very few statistically significant negative relationships between deadline approvals and post-marketing regulatory events (PMREs). We *do* observe a large number of statistically significant positive relationships between deadline approvals and PMRE rates, and these hold across statistical specifications.

It also merits remark that the positive partial correlations between pre-deadline approvals and PMRE rates appear stronger and more robust for standard approvals than for priority approvals. This is not to say that there are not still some appreciable patterns for priority approvals. Such patterns are in evidence. Yet the statistical results are generally more consistent and larger for standard approvals. If this statistical difference reflects an underlying difference in decisions across the two categories of NDAs, it may be that the FDA's acceleration for priority review was less difficult and less drastic at CDER than for standard drugs. By this line of reasoning, the FDA was in the late 1980s and early 1990s

[15] We also estimated random-effects models, as well as generalized estimating equations models. The results are substantively identical to those reported here, though in some cases estimation of more non-linear models standard errors cannot be retrieved.

already moving rather quickly on priority molecules, hence the institutional disruption of the six-month deadline for priority drugs was less than that caused by the deadlines for standard drugs. Testing this claim is beyond the scope of the present paper and would require data that we do not have.

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 in ways that be 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, to be conducting not 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, 2004) 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.

Conclusion. Analysis of new molecular entity data from the last fifty years suggests that the deadlines of PDUFA and FDAMA have introduced immense 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 the 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 have employed statistical methods to account for this. We end, then, with a note of caution about using methods such as these. 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.

Avorn, Jerome. 2004. *Powerful Medicines* (New York: Knopf).

Bowers, Jake, Daniel Carpenter, and Ben Hansen. 2007. "The Effect of PDUFA on Drug Safety: Causal Inference for a Discrete Policy Change," Department of Political Science and National Supercomputing Laboratory, University of Illinois.

Carpenter, Daniel. 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval," *American Journal of Political Science* **46** (July 2002)(3): 490-505.

Carpenter, Daniel. 2004. "Protection without Capture: Product Approval by a Politically Responsive, Learning Regulator," *American Political Science Review* **98**(4):613-31.

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] December 17, 2003, W3-618-624.

Carpenter, Daniel, and A. Mark Fendrick. 2004. "Accelerating Approval Times for New Drugs in the United States," *The Regulatory Affairs Journal – Pharma*, 15 (6) (June): 411-417.

Carpenter, Daniel, and Justin Grimmer. 2007. "Dynamic Stochastic Learning in Regulatory Optimization: Deadlines and Error with Continuous and Discrete Evidence," manuscript, Department of Government, Harvard University.

Carpenter, Daniel, and Michael M. Ting. 2007. "Regulatory Errors with Endogenous Agendas," *American Journal of Political Science* 51 (4) (October 2007) 835-53. Earlier distributed as "Regulatory Errors under Two-Sided Uncertainty," RWJF Scholars in Health Policy Working Paper.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates," *Biometrika* 80 (1993) (1): 27-38.

Fontanarosa, P. B., D. Rennie, and C. D. De Angelis. 2004. "Postmarketing Surveillance -- Lack of Vigilance, Lack of Trust," *Journal of the American Medical Association*, 292(21) (December 1, 2004) 2647-2650.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Analytical Methods for Social Research) (New York: Cambridge University Press).

Lasser, K. E., P. D. Allen, S.J. Woolhandler, D. U. Himmelstein, S. M. Wolfe, D. H. Bor, "Timing of New Black Box Warnings and Withdrawals for Prescription Medications." *JAMA* 287 (17) (May 1, 2002):2215-2220.

Lexchin, J. "Drug withdrawals from the Canadian market for safety reasons, 1963–2004," *CMAJ* 172 (6) (March 15, 2005).

King, Elizabeth N., and Thomas P. Ryan. 2002. "A Preliminary Investigation of Maximum Likelihood Logistic Regression versus Exact Logistic Regression," *The American Statistician* 56 (3) (August 2002) 163-170.

Kosnik, Lea-Rachel D. 2005. "Sources of Bureaucratic Delay: A Case Study of FERC Dam Relicensing," *Journal of Law, Economics and Organization* **22**(1) 258-88.

McCullagh, Peter and J.A. Nelder. 1989. *Generalized Linear Models* (London: Chapman and Hall).

Mehta, Cyrus R., Nitin R. Patel, and Pratay Senchaudhuri. 2000. "Efficient Monte Carlo Methods for Conditional Logistic Regression," *Journal of the American Statistical Association* 95 (449) (March 2000) 99-108.

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, M.K. Managing Delegation in the FDA: Reducing Delay in New-Drug Review. *Journal of Health Politics, Policy, and Law* 29(3): 397-430, 2004.

Olson, Mary K. 2004. "Are Novel Drugs More Risky for Patients than Less Novel Drugs?" *Journal of Health Economics*, 23(6): 1135-1158, 2004.

Pinheiro, Jose C., and Douglas M. Bates. 2000. *Mixed Effects Models in S and S-PLUS*. (New York: Springer-Verlag).

Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model* (New York: Springer-Verlag).

Appendix A-1: Sample S commands and output for estimation of the dynamic Cox model

```
> coxclock3 <- coxph(Surv(.t0, .t, .d) ~ stafcder + subyear + month1pdufa + month2pdufa +
month3pdufa + month4pdufa + month5pdufa + month6pdufa + month7pdufa + month8pdufa +
month9pdufa + month10pdufa + month11pdufa + month12pdufa + month13pdufa + month14pdufa +
month15pdufa + month16pdufa + month17pdufa + month18pdufa + month19pdufa + month20pdufa +
month21pdufa + month22pdufa + month23pdufa + month24pdufa + month7fdama + month8fdama +
month9fdama + month10fdama + month11fdama + month12fdama + month13fdama + month14fdama +
month15fdama + month16fdama + month17fdama + month18fdama + frailty(discode), data =
approved.drugdata.st.20051108.subset.TVC, subset = priority == 0, na.action = na.exclude,
eps = 0.0001, iter.max = 10, method = "efron")

> summary(coxclock3)
  n=34536 (935 observations deleted due to missing values)
                 coef se(coef)      se2 Chisq   DF        p
    stafcder -0.00108 0.00024  0.000235 20.22  1.0 6.9e-006
     subyear -0.00573 0.00787  0.007710  0.53  1.0 4.7e-001
 month1pdufa -1.02380 0.72041  0.719984  2.02  1.0 1.6e-001
 month2pdufa -1.55817 1.01074  1.010447  2.38  1.0 1.2e-001
 month3pdufa -0.20724 0.59993  0.599459  0.12  1.0 7.3e-001
 month4pdufa  0.69101 0.48651  0.485945  2.02  1.0 1.6e-001
 month5pdufa  0.62091 0.41303  0.412373  2.26  1.0 1.3e-001
 month6pdufa  1.70436 0.31887  0.318054 28.57  1.0 9.0e-008
 month7pdufa  0.31006 0.73189  0.731496  0.18  1.0 6.7e-001
 month8pdufa  0.92331 0.61396  0.613506  2.26  1.0 1.3e-001
 month9pdufa -0.48555 1.01701  1.016734  0.23  1.0 6.3e-001
month10pdufa  0.67020 0.60857  0.608100  1.21  1.0 2.7e-001
month11pdufa  2.50317 0.28998  0.289013 74.52  1.0 0.0e+000
month12pdufa  1.17080 0.53658  0.535998  4.76  1.0 2.9e-002
month13pdufa  0.83626 0.60842  0.607893  1.89  1.0 1.7e-001
month14pdufa  1.85710 0.36168  0.360723 26.36  1.0 2.8e-007
month15pdufa  1.64749 0.45109  0.450239 13.34  1.0 2.6e-004
month16pdufa  1.46416 0.54276  0.542001  7.28  1.0 7.0e-003
month17pdufa  0.79674 0.60418  0.603447  1.74  1.0 1.9e-001
month18pdufa  1.77107 0.49637  0.495440 12.73  1.0 3.6e-004
month19pdufa  1.32192 0.50264  0.501990  6.92  1.0 8.5e-003
month20pdufa  0.89923 0.54175  0.541128  2.76  1.0 9.7e-002
month21pdufa  1.87828 0.39961  0.398721 22.09  1.0 2.6e-006
month22pdufa -0.01345 1.03147  1.031109  0.00  1.0 9.9e-001
month23pdufa  1.67523 0.38815  0.387197 18.63  1.0 1.6e-005
month24pdufa  1.29086 0.49516  0.494341  6.80  1.0 9.1e-003
 month7fdama -0.50379 1.22514  1.224939  0.17  1.0 6.8e-001
 month8fdama  0.19153 0.81710  0.816793  0.05  1.0 8.1e-001
 month9fdama  2.15835 1.06951  1.069274  4.07  1.0 4.4e-002
month10fdama  1.96435 0.63331  0.632863  9.62  1.0 1.9e-003
month11fdama -0.87136 0.46675  0.466074  3.49  1.0 6.2e-002
month12fdama  0.78953 0.62791  0.627265  1.58  1.0 2.1e-001
month13fdama  0.79765 0.73126  0.730675  1.19  1.0 2.8e-001
month14fdama -2.05782 1.04953  1.049091  3.84  1.0 5.0e-002
month15fdama -0.06193 0.60692  0.606058  0.01  1.0 9.2e-001
month16fdama  0.55665 0.64701  0.646049  0.74  1.0 3.9e-001
month17fdama -0.20794 0.91401  0.913264  0.05  1.0 8.2e-001
month18fdama -0.36465 0.73182  0.730816  0.25  1.0 6.2e-001
frailty(discode)                          93.85 45.9 3.8e-005

Iterations: 8 outer, 18 Newton-Raphson
     Variance of random effect= 0.0676   I-likelihood = -9696.3
Degrees of freedom for terms=  …
Rsquare= 0.015   (max possible= 0.436 )
Likelihood ratio test= 518  on 83.7 df,   p=0
Wald test            = 354  on 83.7 df,   p=0
```

Appendix A-2: Sample S commands and output for estimation of the linear mixed effects model

```
> mixeff.discont.clockerror01 <- lme( discontperyear ~ subyear +
approve0910month + approve1112month + approve1314month + approve1112monthpdufa
+ approve1314monthpdufa + approve0910monthfdama + approve1112monthfdama, data =
approved.drugdata.xt.newnmes.matchcovar.20051108, random = ~ 1 | discode,
subset = ndaduplicate < 1 & priority < 1, na.action = na.exclude, control =
list(msVerbose = TRUE) )

Iteration:  0 ,  1  function calls, F=  849.2591
Parameters:
[1] 1.272172
Iteration:  1 ,  2  function calls, F=  848.877
Parameters:
[1] 1.445163
Iteration:  2 ,  3  function calls, F=  848.8632
Parameters:
[1] 1.48648

> summary(mixeff.discont.clockerror01)
Linear mixed-effects model fit by REML
 Data: approved.drugdata.xt.newnmes.matchcovar.20051108
  Subset: ndaduplicate < 1 & priority < 1
       AIC       BIC   logLik
  -3885.537 -3828.708 1953.769

Random effects:
 Formula:  ~ 1 | discode
        (Intercept)   Residual
StdDev:  0.01167518 0.05174127

Fixed effects: discontperyear ~ subyear + approve0910month + approve1112month +
approve1314month + approve1112monthpdufa +   approve1314monthpdufa +
approve0910monthfdama + approve1112monthfdama
                           Value Std.Error   DF   t-value p-value
          (Intercept)  0.2426683 0.1350362 1102  1.797062  0.0726
              subyear -0.0001111 0.0000685 1102 -1.623405  0.1048
     approve0910month -0.0013380 0.0076494 1102 -0.174912  0.8612
     approve1112month  0.0067223 0.0073661 1102  0.912597  0.3617
     approve1314month  0.0016860 0.0075225 1102  0.224131  0.8227
approve1112monthpdufa  0.0533005 0.0127216 1102  4.189769  <.0001
approve1314monthpdufa  0.0217312 0.0155111 1102  1.401012  0.1615
approve0910monthfdama  0.0783764 0.0141356 1102  5.544625  <.0001
approve1112monthfdama -0.0821309 0.0195235 1102 -4.206764  <.0001

Standardized Within-Group Residuals:
      Min         Q1        Med        Q3      Max
 -2.31866 -0.4362177 -0.3367561 0.2192182 10.88838

Number of Observations: 1304
Number of Groups: 194
```

Appendix A-3: R commands for linear mixed-effects model estimation of PMRE data using combined just-before-deadline indicator, including MCMC posterior sampling.

```
combowit.lmer.clockerror01 <- lmer(combowit ~ (1 | discode) + subyear +
predead, family = binomial, data = dandat, subset = subyear > 1962, method =
"Laplace", control = list(msVerbose = TRUE), na.action = na.exclude, model =
TRUE)

summary(combowit.lmer.clockerror01)

combowit.lmer.clockerror01.sim <- mcsamp(combowit.lmer.clockerror01)

print(combowit.lmer.clockerror01.sim)

combowit.lmer.clockerror02 <- lmer(combowit ~ (1 | discode) + subyrctr +
predead, family = binomial, data = dandat, subset = subyear > 1962, method =
"Laplace", control = list(msVerbose = TRUE), na.action = na.exclude, model =
TRUE)

summary(combowit.lmer.clockerror02)

combowit.lmer.clockerror02.sim <- mcsamp(combowit.lmer.clockerror02)

print(combowit.lmer.clockerror02.sim)

combobbw.lmer.clockerror01 <- lmer(combobbw ~ (1 | discode) + subyrctr +
predead, family = binomial, data = dandat, subset = subyear > 1974, method =
"Laplace", control = list(msVerbose = TRUE), na.action = na.exclude, model =
TRUE)

summary(combobbw.lmer.clockerror03)

combobbw.lmer.clockerror03.sim <- mcsamp(combobbw.lmer.clockerror03)

print(combobbw.lmer.clockerror03.sim)

witorbbw.lmer.clockerror01 <- lmer(combobbw ~ (1 | discode) + subyrctr +
predead, family = binomial, data = dandat, subset = subyear > 1974, method =
"Laplace", control = list(msVerbose = TRUE), na.action = na.exclude, model =
TRUE)

summary(witorbbw.lmer.clockerror01)

witorbbw.lmer.clockerror01.sim <- mcsamp(witorbbw.lmer.clockerror01)

print(witorbbw.lmer.clockerror01.sim)


discont01.lmer.clockerror01 <- lmer(discont01 ~ (1 | discode) + subyrctr +
predead, family = binomial, data = dandat, subset = subyear > 1949, method =
"Laplace", control = list(msVerbose = TRUE), na.action = na.exclude, model =
TRUE)

summary(discont01.lmer.clockerror01)

discont01.lmer.clockerror01.sim <- mcsamp(discont01.lmer.clockerror01)

print(discont01.lmer.clockerror01.sim)
```

Figure 1 withdrawn from this version.

Figure 2 withdrawn from this version

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

STANDARD NMEs

Variable	Valid NMEs	Mean	Std. Dev.	Minimum	Maximum
Manufacturing Revisions per Marketing Year	1241	0.1331	0.1731	0	1.43
Patient Population Changes per Marketing Year	1241	0.0042	0.0219	0	0.24
Dosage-Form Discontinuations per Marketing Year	1314	0.0272	0.0540	0	0.67
Major Labeling Revision (Lasser et al (2002))	789	0.0330	0.1786	0	1
Major Labeling Revision (KUMC)	1137	0.1618	0.3685	0	1
Canadian Safety Withdrawal (1962-present)	1137	0.0193	0.1378	0	1
Global Safety Withdrawal (1981-present)	650	0.0277	0.1642	0	1

PRIORITY NMEs

Variable	Valid NMEs	Mean	Std. Dev.	Minimum	Maximum
Manufacturing Revisions per Marketing Year	151	0.2402	0.2554	0	1.33
Patient Population Changes per Marketing Year	151	0.0115	0.0395	0	0.33
Dosage-Form Discontinuations per Marketing Year	154	0.0257	0.0676	0	0.43
Major Labeling Revision (Lasser et al (2002))	150	0.0667	0.2503	0	1
Major Labeling Revision (KUMC)	155	0.2968	0.4583	0	1
Canadian Safety Withdrawal (1962-present)	155	0.0194	0.1382	0	1
Global Safety Withdrawal (1981-present)	140	0.0429	0.2033	0	1

Table 2A: **GLM Analysis of Postmarket Dosage-Form Discontinuation for Standard NMEs** (Standard Errors in Parentheses)			
Variables	Linear Mixed Effects Model	Linear All Fixed Effects Model	Linear Fixed Effects Model (post-1962)
Year of Submission	-0.0000 (0.0001)	-0.0000 (0.0001)	-0.0003 (0.0002)
Approval in First Six Months	-----	-----	-0.0005 (0.0084)
Approval in 9th or 10th Month, pre-PDUFA	-0.0011 (0.0075)	0.0003 (0.0080)	-----
Approval in 11th or 12th Month, pre-PDUFA	0.0065 (0.0072)	0.0067 (0.0076)	-----
Approval in 13th or 14th month, pre-PDUFA	0.0021 (0.0074)	0.0015 (0.0078)	-----
Approval in 11th or 12th Month, PDUFA [pre-deadline]	0.0641 (0.0132)	0.0643 (0.0143)	0.0430 (0.0119)
Approval in 13th or 14th Month, PDUFA [post-deadline]	0.0298 (0.0160)	0.0332 (0.0169)	-----
Approval in 9th or 10th Month, FDAMA [pre-deadline]	0.0852 (0.0151)	0.0769 (0.0173)	0.0633 (0.0162)
Approval in 11th or 12th Month, FDAMA [post-deadline]	-0.1035 (0.0199)	-0.1033 (0.0214)	-----
Number of Indicator Variables for Sponsors	49	49	0
Number of Effects Terms for Primary Indication	194 [random]	194 [fixed]	178 [fixed]
NMEs	1,304	1,304	938
R-squared	0.1336	0.1230	0.0430
Corr (u_i, $\boldsymbol{Z\gamma}$)	Assumed = 0	-0.0205	0.0424
Joint Significance of Primary Indication Terms	-----	F(193, 1054) = 1.41 (p = 0.0005)	F (177,756) = 1.36 (p = 0.0033)
Joint Significance of Firm Indicators	Chi-sq = 102.29 (p < 0.0001)	F (48, 1054) = 1.88 (p = 0.0003)	-----

Table 2A [GLM Analysis of Postmarket Dosage-Form Discontinuation for Standard NMEs] Continued			
Parameter Estimate Comparisons	F-Statistic	F-Statistic	F-Statistic
Pre-Deadline versus Post-Deadline [PDUFA] 11th or 12th month approval (PDUFA) versus 13th or 14th month approval (PDUFA)	2.94 ($p = 0.0864$)	2.10 ($p = 0.1473$)	-----
Pre-Deadline versus Post-Deadline, across regimes 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (FDAMA)	34.95 ($p < 0.0001$)	29.79 ($p < 0.0001$)	-----
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (pre-PDUFA)	10.17 ($p = 0.0014$)	8.88 ($p = 0.0030$)	-----
Pre-Deadline versus Post-Deadline [FDAMA] 9th or 10th month approval (FDAMA) versus 11th or 12th month approval (FDAMA)	21.94 ($p < 0.0001$)	42.36 ($p < 0.0001$)	-----
Pre-Deadline [FDAMA] versus same months, pre-FDAMA 9th or 10th month approval (FDAMA) versus 9th or 10th month approval (pre-PDUFA)	18.92 ($p < 0.0001$)	12.00 ($p = 0.0006$)	-----
Pre-Deadline (PDUFA) versus Early Approval 11th or 12th month approval (PDUFA) versus approval in first six months	-----	-----	9.40 ($p = 0.0023$)
Pre-Deadline (FDAMA) versus Early Approval 9th or 10th month approval (FDAMA) versus approval in first six months	-----	-----	13.10 ($p = 0.0003$)
Note: Comparisons of deadline approvals with "early" approvals (less than six months) are restricted to post-1962 approvals because of the very rapid pace of approvals in the 1950s.			

Table 2B:
GLM Analysis of Postmarket Patient Population Changes for Standard NMEs
(Standard Errors in Parentheses)

Variables	Linear Mixed Effects Model	Linear All Fixed Effects Model	Linear Fixed Effects Model (post-1962)
Year of Submission	0.0001 (0.0000)	0.0001 (0.0000)	0.0004 (0.0001)
Approval in First Six Months	-----	-----	-0.0006 (0.0040)
Approval in 9th or 10th Month, pre-PDUFA	0.0013 (0.0031)	0.0011 (0.0035)	-----
Approval in 11th or 12th Month, pre-PDUFA	0.0006 (0.0032)	0.0003 (0.0035)	-----
Approval in 13th or 14th month, pre-PDUFA	0.0015 (0.0034)	0.0022 (0.0037)	-----
Approval in 11th or 12th Month, PDUFA [pre-deadline]	0.0033 (0.0055)	0.0060 (0.0064)	0.0055 (0.0054)
Approval in 13th or 14th Month, PDUFA [post-deadline]	-0.0100 (0.0068)	-0.0134 (0.0074)	-----
Approval in 9th or 10th Month, FDAMA [pre-deadline]	-0.0075 (0.0060)	-0.0078 (0.0074)	-0.0139 (0.0073)
Approval in 11th or 12th Month, FDAMA [post-deadline]	-0.0261 (0.0083)	-0.0283 (0.0091)	-----
Number of Indicator Variables for Sponsors	48	48	0
Number of Effects Terms for Primary Indication	186 [random]	186 [fixed]	174 [fixed]
NMEs	1,232	1,232	902
R-squared	0.1125	0.1005	0.0199
Corr (u_i, $\boldsymbol{Z\gamma}$)	Assumed = 0	-0.3050	-0.0861
Joint Significance of Primary Indication Terms	-----	F (185,990) = 0.89 (p = 0.8473)	F (173,724) = 0.67 (p = 0.9992)
Joint Significance of Firm Indicators (chi-sq)	Chi-sq = 275.64 (p < 0.0001)	F (48, 990) = 2.71 (p < 0.0001)	-----

Table 2B [GLM Analysis of Postmarket Patient Population Changes for Standard NMEs] Continued			
Parameter Estimate Comparisons	F-Statistic	F-Statistic	F-Statistic
Pre-Deadline versus Post-Deadline [PDUFA] 11th or 12th month approval (PDUFA) versus 13th or 14th month approval (PDUFA)	2.39 (p = 0.1220)	4.23 (p = 0.0400)	-----
Pre-Deadline versus Post-Deadline, across regimes 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (FDAMA)	6.26 (p = 0.0123)	6.58 (p = 0.0105)	-----
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (pre-PDUFA)	0.12 (p = 0.7314)	0.42 (p = 0.5177)	-----
Pre-Deadline versus Post-Deadline [FDAMA] 9th or 10th month approval (FDAMA) versus 11th or 12th month approval (FDAMA)	3.37 (p = 0.0664)	2.97 (p = 0.0853)	-----
Pre-Deadline [FDAMA] versus same months, pre-FDAMA 9th or 10th month approval (FDAMA) versus 9th or 10th month approval (pre-PDUFA)	1.20 (p = 0.2732)	0.88 (p = 0.3473)	-----
Pre-Deadline (PDUFA) versus Early Approval 11th or 12th month approval (PDUFA) versus approval in first six months	-----	-----	0.89 (p = 0.3455)
Pre-Deadline (FDAMA) versus Early Approval 9th or 10th month approval (FDAMA) versus approval in first six months	-----	-----	2.72 (p = 0.0997)
Note: Comparisons of deadline approvals with "early" approvals (less than six months) are restricted to post-1962 approvals because of the very rapid pace of approvals in the 1950s.			

Table 2C: **GLM Analysis of Postmarket Manufacturing Revisions for Standard NMEs** (Standard Errors in Parentheses)			
Variables	Mixed Effects Model	All Fixed Effects Model	Disease Fixed Effects (post-1962)
Year of Submission	0.0014 (0.0002)	0.0015 (0.0002)	0.0040 (0.0008)
Approval in First 6 months	-----	-----	-0.1045 (0.0265)
Approval in 9th or 10th Month, pre-PDUFA	-0.0091 (0.0220)	0.0001 (0.0236)	-----
Approval in 11th or 12th Month, pre-PDUFA	-0.0142 (0.0226)	-0.0132 (0.0240)	-----
Approval in 13th or 14th month, pre-PDUFA	0.0485 (0.0239)	0.0551 (0.0250)	-----
Approval in 11th or 12th Month, PDUFA [pre-deadline]	0.1576 (0.0392)	0.2243 (0.0434)	0.1608 (0.0364)
Approval in 13th or 14th Month, PDUFA [post-deadline]	0.0393 (0.0483)	0.0546 (0.0502)	-----
Approval in 9th or 10th Month, FDAMA [pre-deadline]	-0.1305 (0.0430)	-0.1426 (0.0503)	-0.1501 (0.0489)
Approval in 11th or 12th Month, FDAMA [post-deadline]	0.0877 (0.0584)	-0.1066 (0.0623)	-----
Number of Indicator Variables for Sponsors	47	47	-----
Number of Effects Terms for Primary Indication	186 [random]	186 [fixed]	174 [fixed]
NMEs	1,232	1,232	902
R-squared	0.2732	0.2409	0.0763
Corr (u_i, $\boldsymbol{Z\gamma}$)	Assumed = 0	-0.1951	-0.1370
Joint Significance of Primary Indication Terms	-----	F(185, 990) = 1.59 ($p < 0.0001$)	F (173,724) = 1.54 ($p = 0.0001$)
Joint Significance of Firm Indicators	Chi-sq = 275.64 ($p < 0.0001$)	F(48, 990) = 5.36 ($p < 0.0001$)	-----

Table 2C [GLM Analysis of Postmarket Manufacturing Revisions for Standard NMEs] Continued			
Parameter Estimate Comparisons	F-Statistic	F-Statistic	F-Statistic
Pre-Deadline versus Post-Deadline [PDUFA] 11th or 12th month approval (PDUFA) versus 13th or 14th month approval (PDUFA)	3.82 (p = 0.0507)	7.04 (p = 0.0081)	-----
Pre-Deadline versus Post-Deadline, across regimes 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (FDAMA)	8.67 (p = 0.0032)	13.29 (p = 0.0003)	-----
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (pre-PDUFA)	9.73 (p = 0.0018)	15.81 (p = 0.0001)	-----
Pre-Deadline versus Post-Deadline [FDAMA] 9th or 10th month approval (FDAMA) versus 11th or 12th month approval (FDAMA)	0.35 (p = 0.5550)	0.20 (p = 0.6557)	-----
Pre-Deadline [FDAMA] versus same months, pre-FDAMA 9th or 10th month approval (FDAMA) versus 9th or 10th month approval (pre-PDUFA)	4.51 (p = 0.0338)	4.95 (p = 0.0263)	-----
Pre-Deadline (PDUFA) versus Early Approval 11th or 12th month approval (PDUFA) versus approval in first six months	-----	-----	36.56 (p < 0.0001)
Pre-Deadline (FDAMA) versus Early Approval 9th or 10th month approval (FDAMA) versus approval in first six months	-----	-----	0.72 (p = 0.3956)
Note: Comparisons of deadline approvals with "early" approvals (less than six months) are restricted to post-1962 approvals because of the very rapid pace of approvals in the 1950s.			

Table 2D: GLM Analysis of Safety-Based Withdrawals for Standard NMEs (Canada, 1962-present)				
Variables	Linear Mixed Effects Model	Linear All Fixed Effects Model	Linear Mixed Effects Model	Extreme Value Regression
Year of Submission	-0.0019 (0.0013)	-0.0023 (0.0015)	-0.0021 (0.0013)	-0.0861 (0.0911)
Approval in First 6 months	-----	-----	-0.0263 (0.0191)	Estimate unobtainable
Approval in 9th or 10th Month, pre-PDUFA	0.0495 (0.0256)	0.0694 (0.0283)	-----	-----
Approval in 11th or 12th Month, pre-PDUFA	-0.0217 (0.0212)	-0.0220 (0.0230)	-----	-----
Approval in 13th or 14th month, pre-PDUFA	-0.0195 (0.0214)	-0.0156 (0.0236)	-----	-----
Approval in 11th or 12th Month, PDUFA [pre-deadline]	0.1738 (0.0390)	0.1735 (0.0465)	0.1236 (0.0310)	3.2544 (1.2003)
Approval in 13th or 14th Month, PDUFA [post-deadline]	0.0402 (0.0535)	0.0576 (0.0569)	-----	-----
Approval in 9th or 10th Month, FDAMA [pre-deadline]	0.0637 (0.1556)	0.0645 (0.1710)	0.0829 (0.1536)	Estimate unobtainable
Approval in 11th or 12th Month, FDAMA [post-deadline]	-0.1617 (0.0804)	-0.1493 (0.0873)	-----	-----
Number of Indicator Variables for Sponsors	46	46	46	46 [38 dropped]
Number of Effects Terms for Primary Indication	152 [random]	152 [fixed]	152 [random]	0
NMEs	951	951	951	583
R-squared [model Wald stat for Extreme Value model]	0.0516	0.0322	0.0444	Wald = 97.72 ($p < 0.0001$)
Corr (u_i, $\boldsymbol{Z\gamma}$)	Assumed = 0	-0.1833	Assumed = 0	-----
Joint Significance of Primary Indication Terms	-----	F(151, 746) = 1.34 ($p = 0.0078$)	-----	-----
Joint Significance of Firm Indicators	Chi-sq = 33.17 ($p = 0.9216$)	F(46, 746) = 0.72 ($p = 0.9134$)	Chi-sq = 33.01 ($p = 0.9246$)	Chi-sq = 26.61 ($p = 0.0008$)

Table 2D [GLM Analysis of Safety-Based Withdrawals for Standard NMEs (Canada, 1962-present)] Continued				
Parameter Estimate Comparisons	F-Statistic	F-Statistic	Chi-Squared	F-Statistic
Pre-Deadline versus Post-Deadline [PDUFA] 11th or 12th month approval (PDUFA) versus 13th or 14th month approval (PDUFA)	4.73 (p = 0.0297)	2.95 (p = 0.0863)	-----	-----
Pre-Deadline versus Post-Deadline, across regimes 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (FDAMA)	11.44 (p = 0.0007)	8.29 (p = 0.0041)	-----	-----
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (pre-PDUFA)	13.48 (p = 0.0002)	10.32 (p = 0.0014)	-----	-----
Pre-Deadline versus Post-Deadline [FDAMA] 9th or 10th month approval (FDAMA) versus 11th or 12th month approval (FDAMA)	1.60 (p = 0.2057)	1.20 (p = 0.2744)	-----	-----
Pre-Deadline [FDAMA] versus same months, pre-FDAMA 9th or 10th month approval (FDAMA) versus 9th or 10th month approval (pre-PDUFA)	0.01 (p = 0.9300)	0.00 (p = 0.9777)	-----	-----
Pre-Deadline (PDUFA) versus Early Approval 11th or 12th month approval (PDUFA) versus approval in first six months	-----	-----	18.08 (p < 0.0001)	Infinite [p unobtainable]
Pre-Deadline (FDAMA) versus Early Approval 9th or 10th month approval (FDAMA) versus approval in first six months	-----	-----	0.50 (p = 0.4803)	Infinite [p unobtainable]
Note: Models also include total CDER staff in year of submission (varies over time), and order of entry for NME into therapeutic class (varies over time and across NMEs.				

Table 2E: GLM Analysis of Global Safety-Based Withdrawals for Standard NMEs (1980-present)				
Variables	Linear Mixed Effects Model	Linear All Fixed Effects Model	Linear Mixed Effects Model	Extreme Value Regression
Year of Submission	0.0004 (0.0011)	0.0012 (0.0016)	0.0008 (0.0013)	0.0306 (0.0360)
Approval in First 6 months	-----	-----	-0.0202 (0.0255)	-----
Approval in 9th or 10th Month, pre-PDUFA	-0.0659 (0.0441)	-0.0486 (0.0574)	-----	-15.0984 (0.6079)
Approval in 11th or 12th Month, pre-PDUFA	-0.0583 (0.0443)	-0.0597 (0.0540)	-----	-15.2724 (0.9391)
Approval in 13th or 14th month, pre-PDUFA	0.0420 (0.0410)	0.0317 (0.0498)	-----	1.9233 (1.2785)
Approval in 11th or 12th Month, PDUFA [pre-deadline]	0.1177 (0.0566)	0.0683 (0.0691)	0.0264 (0.0310)	16.7486 (1.2590)
Approval in 13th or 14th Month, PDUFA [post-deadline]	-0.0020 (0.0599)	0.0229 (0.0724)	-----	-0.7420 (1.6393)
Approval in 9th or 10th Month, FDAMA [pre-deadline]	0.1013 (0.0601)	0.0520 (0.0775)	0.0280 (0.0411)	16.7770 (1.1333)
Approval in 11th or 12th Month, FDAMA [post-deadline]	-0.0972 (0.0626)	-0.0640 (0.0724)	-----	Estimate unobtainable
Number of Indicator Variables for Sponsors	47	47	47	46 [34 dropped]
Number of Effects Terms for Primary Indication	159 [random]	159 [fixed]	159 [random]	0
NMEs	632	632	632	453
R-squared [model Wald stat for Extreme Value model]	0.1667	0.1318	0.1551	Wald = 1150.04 ($p < 0.0001$)
Corr (u_i, $\boldsymbol{Z\gamma}$)	Assumed = 0	-0.3753	Assumed = 0	-----
Joint Significance of Primary Indication Terms	-----	F(158, 418) = 0.80 ($p = 0.9481$)	-----	-----
Joint Significance of Firm Indicators	Chi-sq = 109.69 ($p < 0.0001$)	F(47, 418) = 2.69 ($p < 0.0001$)	Chi-sq = 103.13 ($p < 0.0001$)	Chi-sq = 31.97 ($p = 0.0014$)

Table 2E [GLM Analysis of Global Safety-Based Withdrawals for Standard NMEs (1980-present)] Continued				
Parameter Estimate Comparisons	Chi-Squared	F-Statistic	Chi-Squared	Chi-Squared
Pre-Deadline versus Post-Deadline [PDUFA] 11th or 12th month approval (PDUFA) versus 13th or 14th month approval (PDUFA)	2.27 ($p = 0.1319$)	0.23 ($p = 0.6339$)	-----	82.20 ($p < 0.0001$)
Pre-Deadline versus Post-Deadline, across regimes 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (FDAMA)	4.92 ($p = 0.0266$)	1.29 ($p = 0.2569$)	-----	Infinite [p unobtainable]
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 11th or 12th month approval (PDUFA) versus 11th or 12th month approval (pre-PDUFA)	3.42 ($p = 0.0644$)	1.22 ($p = 0.2694$)	-----	246.33 ($p < 0.0001$)
Pre-Deadline versus Post-Deadline [FDAMA] 9th or 10th month approval (FDAMA) versus 11th or 12th month approval (FDAMA)	5.41 ($p = 0.0200$)	1.21 ($p = 0.2724$)	-----	Infinite [p unobtainable]
Pre-Deadline [FDAMA] versus same months, pre-FDAMA 9th or 10th month approval (FDAMA) versus 9th or 10th month approval (pre-PDUFA)	2.96 ($p = 0.0855$)	0.64 ($p = 0.4245$)	-----	427.37 ($p < 0.0001$)
Pre-Deadline (PDUFA) versus Early Approval 11th or 12th month approval (PDUFA) versus approval in first six months	-----	-----	1.56 ($p = 0.2115$)	-----
Pre-Deadline (FDAMA) versus Early Approval 9th or 10th month approval (FDAMA) versus approval in first six months	-----	-----	1.14 ($p = 0.2849$)	-----
Note: Models also include total CDER staff in year of submission (varies over time), and order of entry for NME into therapeutic class (varies over time and across NMEs.				

Table 3A:
GLM Analysis of Canadian Safety-Based Withdrawals for Priority NMEs (1980-present)
(Robust standard errors in parentheses)

Variables	Linear Model	Extreme Value Regression	Linear Fixed Effects Model	Extreme Value Regression (Fixed Eff)
Year of Submission	-0.0002 (0.0003)	-0.0115 (0.0274)	-0.0007 (0.0010)	0.1484 (0.1462)
Approval in First 4 months	-0.0148 (0.0116)	**	-0.0024 (0.0071)	**
Approval in 5th or 6th Month, pre-PDUFA	-0.0161 (0.0107)	-14.7742 (0.7529)	-0.0215 (0.0437)	-17.5971 (1.4550)
Approval in 5th or 6th Month, PDUFA [pre-deadline]	0.0624 (0.0445)	16.2405 (0.9649)	0.0450 (0.0437)	18.4802 (1.2480)
Number of Indicator Variables for Sponsors	0	0	28	28 [25 dropped]
NMEs	179	171	179	78
F-statistic [or Wald stat for Extreme Value model]	1.01 (p = 0.4028)	621.75 (p < 0.0001)	1.69 (p = 0.0119)	**
Joint Significance of Firm Indicators	-----	-----	0.10 (p = 1.0000)	1002.16 (p < 0.0001)
Parameter Estimate Comparisons	F-Statistic	Chi-Squared	F-Statistic	Chi-Squared
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 5th or 6th month priority approval (9/1992 - present) versus 5th or 6th month priority approval (pre-9/1992)	2.87 (p = 0.0920)	579.95 (p < 0.0001)	0.70 (p = 0.4040)	209.09 (p < 0.0001)
Pre-Deadline (PDUFA) versus Early Approval 5th or 6th month priority approval (9/1992 - present) versus approval in first four months	2.98 (p = 0.0862)	Infinite [p unobtainable]	1.14 (p = 0.2878)	Infinite [p unobtainable]

Notes: ** = Estimate unobtainable (for likelihood-based models where zero events occur in the relevant category). Robust standard error in extreme value regression is sandwich estimate. Where linear model F-statistic cannot be retrieved from robust covariance matrix estimation, an estimate is obtained from non-robust estimation. No withdrawals occur for 7th and 8th month approvals under PDUFA, hence parameter estimate comparisons are usually explosive (infinitely valued) and are excluded from this table.

Table 3B:
GLM Analysis of Global Safety-Based Withdrawals for Priority NMEs (1980-present)
(Robust standard errors in parentheses)

Variables	Linear Model	Extreme Value Regression	Linear Fixed Effects Model	Extreme Value Regression (Fixed Eff)
Year of Submission	-0.0013 (0.0034)	-0.0305 (0.0763)	-0.0006 (0.0050)	-0.0430 (0.0825)
Approval in First 4 months	-0.0390 (0.0200)	**	-0.0421 (0.0321)	**
Approval in 5th or 6th Month, pre-PDUFA	-0.0451 (0.0209)	-15.1716 (0.6250)	-0.0869 (0.0732)	0.5002 (1.2965)
Approval in 5th or 6th Month, PDUFA [pre-deadline]	0.0697 (0.0487)	15.7011 (1.0174)	0.1021 (0.0734)	-0.1151 (1.1736)
Number of Indicator Variables for Sponsors	0	0	27	27 [22 dropped]
NMEs	165	157	165	89
F-statistic [or Wald stat for Extreme Value model]	1.78 (p = 0.1358)	671.77 (p < 0.0001)	0.33 (p = 0.9999)	Wald = 5.67 (p = 0.6838)
Joint Significance of Firm Indicators	-----	-----	0.27 (p = 0.9999)	Chi-sq = 5.59 (p = 0.3480)
Parameter Estimate Comparisons	F-Statistic	Chi-Squared	F-Statistic	Chi-Squared
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 5th or 6th month priority approval (9/1992 - present) versus 5th or 6th month priority approval (pre-9/1992)	4.05 (p = 0.0459)	504.12 (p < 0.0001)	2.08 (p = 0.1519)	0.08 (p = 0.7770)
Pre-Deadline (PDUFA) versus Early Approval 5th or 6th month priority approval (9/1992 - present) versus approval in first four months	4.85 (p = 0.0291)	Infinite [p unobtainable]	2.95 (p = 0.0884)	Infinite [p unobtainable]

Notes: ** = Estimate unobtainable (for likelihood-based models where zero events occur in the relevant category). Robust standard error in extreme value regression is sandwich estimate. Where linear model F-statistic cannot be retrieved from robust covariance matrix estimation, an estimate is obtained from non-robust estimation. No global withdrawals occur for 7th and 8th month approvals before or after September 1992, hence parameters and relevant parameter estimate comparisons are excluded from this table.

Table 3C:
GLM Analysis of Major Safety-Based Label Changes, Lasser (*JAMA* 2002) List, NMEs 1975-2000
(Robust standard errors in parentheses)

Variables	Linear Model	Extreme Value Regression	Linear Fixed Effects Model	Extreme Value Regression (Fixed Effects)
Year of Submission	-0.0109 (0.0040)	-0.1400 (0.0383)	-0.0068 (0.0042)	-0.1216 (0.0686)
Approval in First 4 months	0.0959 (0.1116)	1.5424 (1.1019)	0.0976 (0.1365)	1.4529 (1.1543)
Approval in 5th or 6th Month, pre-PDUFA	-0.0969 (0.0349)	-15.5952 (0.5762)	-0.0745 (0.0524)	-14.3798 (1.9576)
Approval in 5th or 6th Month, PDUFA [pre-deadline]	0.1539 (0.0582)	16.6822 (0.9825)	0.0801 (0.0708)	14.6095 (1.7068)
Number of Indicator Variables for Sponsors	0	0	28	28 [21 dropped]
NMEs	175	175	175	91
F-statistic [or Wald stat for Extreme Value model]	2.48 ($p = 0.0461$)	807.35 ($p < 0.0001$)	1.79 ($p = 0.0066$)	478.74 ($p < 0.0001$)
Joint Significance of Firm Indicators	-----	-----	1.65 ($p = 0.0207$)	12.20 ($p = 0.0940$)
Parameter Estimate Comparisons	F-Statistic	Chi-Squared	F-Statistic	Chi-Squared
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 5th or 6th month priority approval (9/1992 - present) versus 5th or 6th month priority approval (pre-9/1992)	9.00 ($p = 0.0031$)	605.37 ($p < 0.0001$)	1.84 ($p = 0.1778$)	174.07 ($p < 0.0001$)
Pre-Deadline (PDUFA) versus Early Approval 5th or 6th month priority approval (9/1992 - present) versus approval in first four months	0.25 ($p = 0.6202$)	151.79 ($p < 0.0001$)	0.02 ($p = 0.8786$)	53.60 ($p < 0.0001$)

Notes: Robust standard error in extreme value regression is sandwich estimate. Where linear model F-statistic cannot be retrieved from robust covariance matrix estimation, an estimate is obtained from non-robust estimation. No withdrawals occur for 7th and 8th month approvals under PDUFA, hence parameter estimate comparisons are usually explosive (infinitely valued) and are excluded from this table.

Table 3D:
GLM Analysis of Major Safety-Based Label Changes, KUMC List
(Robust standard errors in parentheses)

Variables	Linear Model	Extreme Value Regression	Linear Fixed Effects Model	Extreme Value Regression (Fixed Eff)
Year of Submission	-0.019 (0.0037)	-0.0064 (0.0102)	-0.0138 (0.0075)	-0.0655 (0.0296)
Approval in First 4 months	0.2328 (0.1832)	0.7392 (0.5398)	0.2087 (0.1646)	1.1162 (0.5344)
Approval in 5th or 6th Month, pre-PDUFA	0.1617 (0.1708)	0.2236 (0.3536)	0.0255 (0.2209)	0.0731 (0.7615)
Approval in 5th or 6th Month, PDUFA [pre-deadline]	-0.0910 (0.1890)	15.7011 (1.0174)	0.1275 (0.2327)	0.8291 (0.8511)
Number of Indicator Variables for Sponsors	0	0	27	27 [13 dropped]
NMEs	180	180	175	137
F-statistic [or Wald stat for Extreme Value model]	0.70 ($p = 0.5916$)	2.43 ($p = 0.6577$)	1.68 ($p = 0.0134$)	30.90 ($p = 0.0295$)
Joint Significance of Firm Indicators	-----	-----	182.58 ($p < 0.0001$)	25.62 ($p = 0.0290$)
Parameter Estimate Comparisons	F-Statistic	Chi-Squared	F-Statistic	Chi-Squared
Pre-Deadline [PDUFA] versus same months, pre-PDUFA 5th or 6th month priority approval (9/1992 - present) versus 5th or 6th month priority approval (pre-9/1992)	0.53 ($p = 0.4748$)	0.71 ($p = 0.4006$)	0.05 ($p = 0.8179$)	0.24 ($p = 0.6242$)
Pre-Deadline (PDUFA) versus Early Approval 5th or 6th month priority approval (9/1992 - present) versus approval in first four months	1.54 ($p = 0.2163$)	1.95 ($p = 0.1623$)	0.08 ($p = 0.7789$)	0.08 ($p = 0.7755$)

Note: Robust standard error in extreme value regression is sandwich estimate. Where linear model F-statistic cannot be retrieved from robust covariance matrix estimation, an estimate is obtained from non-robust estimation.

Table 4A-1: Combined NME Sample, Canadian Safety Withdrawals (1962-present) Firth-Corrected and Monte Carlo Efficient Regressions, Non-Stratified Sample

[Adapted from LogXact7 Tabular Output (Cytel 2006)]

regression (type=logit, model(canadian = predead subyear), estimate(predead subyear), method=monte, mle=firth, asiter=50, covmat);

Data file reduced-allNMEs4-20070226.cyd

Analysis type Estimate :: Monte Carlo

Number of NMEs in analysis: 1549

Summary Statistics

Statistics	Value	DF	P-Value
Deviance	64.9378	52	0.107366
Likelihood Ratio	1903.85	3	0.00

Parameter Estimates

	Point Estimate			Confidence Interval and P-Value for Beta			
					99%CI		P-Value	
Model Term	**Type**	**Beta**	**SE(Beta)**	**Type**	**Lower**	**Upper**	**2*1-sided**	**SE**
Constant	PMLE	110.8043948	42.59069	Asymptotic	27.328184	194.2806	0.009278	
Year of NME Submission	PMLE	-0.058094928	0.021543	Asymptotic	-0.100319	-0.01587	0.007004	
	CMLE	-0.058848902	0.022147	Monte Carlo	-0.103673	-0.01625	0.0068	0.001164
					(Seed = 27627, Samples = 10000)			
Pre-Deadline Approval (Standard & Priority)	PMLE	2.422677118	0.673313	Asymptotic	1.103008	3.742346	0.000321	
	CMLE	2.420301873	0.696211	Monte Carlo	0.8161904	4.041967	0.0016	0.000565
					(Seed = 27620, Samples = 10000)			

PMLE: Penalized MLE for bias correction (Firth's method)

Covariance Matrix

	Constant	Pre-Deadline Approval	Year of NME Submission
Constant	1813.97	19.49908057	-0.91754
Pre-Deadline Approval	19.4991	0.453350338	-0.00989
Year of NME Submission	-0.91754	-0.009887876	0.000464

Table 4A-2: Combined NME Sample, Canadian Safety Withdrawals (1962-present) Firth-Corrected and Monte Carlo Efficient Regressions, Stratified Sample

[Adapted from LogXact7 Tabular Output (Cytel 2006)]

regression (type=logit, model(canadian = predead subyear), stratum=discode, estimate(predead subyear), method=monte, mle=firth, asiter=50, covmat);

Data file	reduced-allNMEs4-20070226.cyd
Stratum variable	"discode" (Primary Indication of NME)
Informative strata	14
Analysis type	Estimate :: Monte Carlo
Number of NMEs in analysis: 370	
Number of groups	196

Summary Statistics

Statistics	Value	DF	P-Value
Likelihood Ratio	10.8488	2	0.0044077

Parameter Estimates

		Point Estimate		Confidence Interval and P-Value for Beta				
					99%CI		P-Value	
Model Term	**Type**	**Beta**	**SE(Beta)**	**Type**	**Lower**	**Upper**	**2*1-sided**	**SE**
Year of NME Submission	PMLE	-0.034397713	0.0232279	Asymptotic	-0.079924	0.011128	0.138639	
	CMLE	-0.034834886	0.0235903	Monte Carlo	-0.081485	0.010699	0.1388	0.005083
					(Seed = 28005, Samples = 10000)			
Pre-Deadline Approval (Standard & Priority)	PMLE	3.228001109	1.0170508	Asymptotic	1.2346182	5.221384	0.001504	
	CMLE	3.159580938	1.0048372	Monte Carlo	0.8892451	5.889337	0.0044	0.000937
					(Seed = 27986, Samples = 10000)			

PMLE: Penalized MLE for bias correction (Firth's method).

Covariance Matrix

	Pre-Deadline Approval	**subyear**
Pre-Deadline Approval	1.03439	-0.00961371
subyear	-0.00961	0.000539536

Table 4B: Combined NME Samples, Global Safety Withdrawals (1981-present) Firth-Corrected and Monte-Carlo Efficient Logistic Regression

[Adapted from LogXact7 Tabular Output (Cytel 2006)]

regression (type=logit, model(withdraw_p = subyear predead), estimate(subyear predead), method=monte, mle=firth, asiter=50, covmat);

Data file reduced-allNMEs5-20070226.csv
Analysis type Estimate :: Monte Carlo

Number of NMEs in analysis: 997

Summary Statistics

Statistics	Value	DF	P-Value
Deviance	48.5099	33	0.039922
Likelihood Ratio	1130.51	3	3.7E-09

Parameter Estimates

		Point Estimate		Confidence Interval and P-Value for Beta				
					99%CI		P-Value	
Model Term	**Type**	**Beta**	**SE(Beta)**	**Type**	**Lower**	**Upper**	**2*1-sided**	**SE**
Constant	PMLE	44.74434473	66.42083	Asymptotic	-85.43809	174.9268	0.500534	
Year of NME Submission	PMLE	-0.024323613	0.033399	Asymptotic	-0.089785	0.041137	0.466447	
	CMLE	-0.025870433	0.034374	Monte Carlo	-0.096268	0.038608	0.4642	0.008443
					(Seed = 1844, Samples = 10000)			
Pre-Deadline Approval	PMLE	1.175016836	0.563453	Asymptotic	0.0706701	2.279364	0.037034	
(Standard & Priority)	CMLE	1.159521119	0.56956	Monte Carlo	-0.159571	2.446721	0.095	0.004254
					(Seed = 1850, Samples = 10000)			

PMLE: Penalized MLE for bias correction (Firth's method).

Covariance Matrix

	Constant	Year of NME Submission	Pre-Deadline Approval
Constant	4411.73	-2.218383104	21.72235
Year of NME Submission	-2.21838	0.001115498	-0.01095
Pre-Deadline Approval	21.7224	-0.010945458	0.317479

Table 4C: Combined NME Sample (Standard & Priority), Lasser Relabeling Measure (1974-2000)
Firth-Corrected and Monte Carlo Efficient Maximum Likelihood Estimates

[Adapted from LogXact7 Tabular Output (Cytel 2006)]

regression (type=logit, model(lasser = predead subyear), estimate(predead subyear), method=monte, mle=firth, asiter=50, covmat);

Data file reduced-allNMEs5-20070226.csv

Analysis type Estimate :: Monte Carlo

Number of NMEs in analysis: 1176

Summary Statistics

Statistics	Value	DF	P-Value
Deviance	48.0919	31	0.0258
Likelihood Ratio	1263.3644	3	0.0000

Parameter Estimates

	Point Estimate			Confidence Interval and P-Value for Beta				
					99%CI		P-Value	
Model Term	**Type**	**Beta**	**SE(Beta)**	**Type**	**Lower**	**Upper**	**2*1-sided**	**SE**
Constant	PMLE	60.9635	46.2301	Asymptotic	-20.5246	151.5728	0.1873	
Year of NME Submission	PMLE	-0.0323	0.0233	Asymptotic	-0.0780	0.0133	0.1649	0.0054
	CMLE	-0.0332	0.0236	Monte Carlo	-0.0807	0.0133	0.1558	
Pre-Deadline Approval	PMLE	0.6315	0.6373	Asymptotic	0.0405034	1.016001	0.3217	
	CMLE	0.5293	0.6702	Monte Carlo	-1.2636	1.9349	0.6380	0.0093

PMLE: Penalized MLE for bias correction (Firth's method).

Covariance Matrix

	Constant	Pre-Deadline Approval	Year of NME Submission
Constant	2137.2220	12.7672	-1.0766
Pre-Deadline Approval	12.7672	0.4061	-0.0064
Year of NME Submission	-1.0766	-0.0064	0.0005

Table 4D: Combined NME Sample (Standard & Priority), KUMC Relabeling Measure Firth-Corrected Maximum Likelihood Estimates

[Adapted from LogXact7 Tabular Output (Cytel 2006)]

regression (type=logit, model(kumc = predead subyear), estimate(predead subyear), method=monte, mle=firth, asiter=50, covmat);

Data file reduced-allNMEs5-20070226.csv

Analysis type Estimate :: Monte Carlo

Number of NMEs in analysis: 2107

Summary Statistics

Statistics	Value	DF	P-Value
Deviance	88.8965	67	0.038045846
Likelihood Ratio	1116.6	3	3.05895E-09

Parameter Estimates

	Point Estimate			Confidence Interval and P-Value for Beta			
					99%CI		P-Value
Model Term	**Type**	**Beta**	**SE(Beta)**	**Type**	**Lower**	**Upper**	**2*1-sided**
Constant	PMLE	-9.094061997	5.832015706	Asymptotic	-20.5246	2.336479	0.118917
Year of NME Submission	PMLE	0.003729549	0.002954443	Asymptotic	-0.002061	0.00952	0.206822
	CMLE	Infeasible				Infeasible	
Pre-Deadline Approval	PMLE	0.528252138	0.248855981	Asymptotic	0.0405034	1.016001	0.033777
	CMLE	Infeasible				Infeasible	

PMLE: Penalized MLE for bias correction (Firth's method).

Covariance Matrix

	Constant	Pre-Deadline Approval	Year of NME Submission
Constant	34.0124	0.42396564	-0.01722936
Pre-Deadline Approval	0.42397	0.061929299	-0.00021677
Year of NME Submission	-0.01723	-0.000216773	8.72873E-06

[Table 4E withdrawn from this version.]

Table 5 withdrawn from this version.