

Drug Safety and Deadlines

Causal Inference About A Regulatory Regime Change at the FDA

Jake Bowers ¹ Dan Carpenter ² Ben Hansen ³

¹Harvard RWJ Cohort 12, Harvard University and
Political Science & NCSA University of Illinois, Champaign-Urbana

²Michigan RWJ Cohort 9, and Government, Harvard University

³Statistics & SRC, University of Michigan

1 June 2007 / RWJ @ Aspen

What do these drugs have in common?

Baycol Tasmar Duract Famvir Posicor Prelay Rezulin Raxar Provan

All withdrawn from the market since 1992 because of safety problems.

All submitted, reviewed, and approved after the Prescription Drug User Fee Act (PDUFA) was passed and implemented (October 29, 1992)

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

2

Today I want to tell you about how we've been trying to answer the tricky question of whether the Prescription Drug User Fee Act caused drug safety problems.

This is joint work with Dan Carpenter of Harvard University dept of Govt (point) and Ben Hansen in the Statistics Dept at the University of Michigan. It is part of a larger project on the FDA which includes Susan Moffitt (RWJ Cohort 13 at Harvard) (point) and Chris Afendulis (Harvard Univ Medical School, Dept of Healthcare Policy).

So, by now you've had time to answer the question posed by this slide.

What do these drugs have in common?

Baycol Tasmar Duract Famvir Posicor Prelay Rezulin Raxar Provan

All withdrawn from the market since 1992 because of safety problems.

All submitted, reviewed, and approved after the Prescription Drug User Fee Act (PDUFA) was passed and implemented (October 29, 1992)

These drugs have two things in common: For one thing, they were withdrawn from either the us, the canadian, or the global market for safety reasons. For those of you following the FDA related news, Rezulin is a treatment for diabetes in the same general class of drugs as Avandia --- but it was withdrawn because of liver toxicity, not heart problems.

What do these drugs have in common?

Baycol Tasmar Duract Famvir Posicor Prelay Rezulin Raxar Provan

All withdrawn from the market since 1992 because of safety problems.

All submitted, reviewed, and approved after the Prescription Drug User Fee Act (PDUFA) was passed and implemented (October 29, 1992)

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

4

These drugs have more in common than just that they were pulled from the market for safety reasons. All were submitted and approved after the The Prescription Drug User Fee Act of 1992 was enacted; an act which required the FDA to do its work under a deadline, and an act which made the funding of the FDA directly dependent on contributions from drug companies.

The fact that these drugs were approved under a deadline, where possibly safety could be sacrificed to speed, or where an organizational culture could be changed from a preoccupation with preventing safety problem to one focused on pleasing corporate clients raises the question which orients this talk, which represents a very preliminary bit of analysis:

We want to know how many of them would not have had these problems if PDUFA has not been in effect when they were approved?

Now, most social scientists would look at these 9 events and say, ``What? You want to explain why 9 drugs were withdrawn out of the hundreds approved and sold? You are trying to explain what is essentially noise." And, as you'll see we run into problems with this and have a couple of ideas for solutions --- including using outcomes that we think are much better than withdrawals. That said, these 9 withdrawals are not just a small bit of social science data. They have caused a lot of attention to be focused on the FDA. (next slide)

Rare events with public policy consequences

$9 > 5$ but $9/154(.058) \approx 5/98(.051)$

“Since the passage of PDUFA in 1992 there have been an unprecedented number of drugs approved and then withdrawn for safety reasons. Nine of the drugs . . . approved . . . from 1993 to 2000 have been withdrawn because of safety concerns. By comparison, only **five** of the drugs approved . . . from 1985 through 1992 later had to be withdrawn. . . . **At least five of these nine drugs . . . were approved despite known safety problems and the availability of multiple treatment options in other, older (and safer) drugs approved for the same medical uses.**” (Lurie and Sasich, N.d., from Public Citizen)

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

5

This slide shows an example of the kinds of concerns raised by these withdrawals and other withdrawals since 1992 by Public Citizen, among other groups concerned with drug safety. They note that more drugs were withdrawn for safety reasons after PDUFA than were before. Since 9 is nearly twice as many as 5, this ought to be compelling evidence about the troubles PDUFA has caused for the FDA. However, if you take into account the number of drugs approved in those periods, you discover that there was virtually no change in the rate of withdrawals.

Adding a denominator in this way does not mean that Public Citizen is wrong to be concerned. Maybe some of those nine drugs would not have been approved if their Stage III clinical trials had ended before the new regulatory regime began. That is, maybe PDUFA caused some of those 9 safety problems.

Over the past year we've developed a set of tools to try to answer this question (and, as you'll see, we realize that the data are insufficient to answer the question about withdrawals per se). However, before I jump into methods and results, I figure that I should provide a brief explanation about why PDUFA is such a big deal at the FDA.

So, I've been talking a bit about PDUFA. And many of you know this acronym. Others have figured out that whatever it is, it must provide people who care about drug safety cause for concern. In one slide, let me quickly bring everyone up to speed on what PDUFA means, and why it might cause concern. (next slide)

The Changing FDA Mandate

Oversight → Gatekeeper → Industry Partner

- 1906: Interdict and Penalize False Marketing
- 1938: Pre-Market Approval and Safety
- 1962: Pre-Market Approval, Safety, and Effectiveness FDA approval required before drugs can be sold. No deadline for approval.
"At all costs prevent another Thalidomide disaster."
- 1992: ...and Speed The Prescription Drug User Fee Act (PDUFA I) required the FDA to act on 90% of standard drugs within 12 months.
"Don't allow HIV sufferers to die waiting for life saving drugs."
- 1997 ...and more Speed (PDUFA II) required the FDA to act on 90% of standard drugs within 10 months.
- 2002 ...and more Safety After Approval (PDUAFA III) added funding for post-marketing surveillance.

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

6

Here is a very quick overview:

The FDA didn't have a lot of regulatory teeth until 1938 --- when it could pull unsafe drugs from market, whether or not firms intended to deceive or hurt consumers --- and where firms had to notify the FDA that their drugs were safe before marketing them.

In 1962, however, it got long, sharp teeth and strong jaws. No drugs could be sold without explicit approval of the FDA and the FDA had to approve clinical trials, too. And eventually the FDA came to determine how clinical trials were run and analyzed.

However, even then, critics of the FDA worried about ``drug-lag" and a small literature sprang up pointing out that people in Germany had access to life-saving (and money making) medicines months and years before folks in the USA.

More voices were added to the group of those worried about the overly conservative and slow aspects of the FDA approval of new drugs when HIV hit the nation --- by the end of 1980s groups like Act Up were engaging in civil disobedience against drug companies claiming that FDA for withholding life-saving drugs from the market.

Clearly, preventing bad drugs from coming to market is only one way to improve public health, obviously another way is to speed the process by which new and better drugs get to the market. And the PDUFA acts (starting in 1992) were mainly designed with this speed in mind.

The PDUFA acts heralded what FDA observers might say is a big change in the prevailing ethos at the FDA that I've indicated here with a couple of exhortations:

(advance)

The Changing FDA Mandate

Oversight → Gatekeeper → Industry Partner

1906: Interdict and Penalize False Marketing

1938: Pre-Market Approval and Safety

1962: Pre-Market Approval, Safety, and Effectiveness FDA approval required before drugs can be sold. No deadline for approval.
“At all costs prevent another Thalidomide disaster.”

1992: ... *and Speed* The Prescription Drug User Fee Act (PDUFA I) required the FDA to act on 90% of standard drugs within 12 months.
“Don’t allow HIV sufferers to die waiting for life saving drugs.”

1997 ... *and more Speed* (PDUFA II) required the FDA to act on 90% of standard drugs within 10 months.

2002 ... *and more Safety After Approval* (PDUAFA III) added funding for post-marketing surveillance.

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

7

While preventing safety problems was the predominant concern especially after the vivid Thalidomide disaster in Europe. What has been especially vivid in the current environment is the image of potentially life saving drugs existing just out of reach of dying patients.

So, in 1992 the Prescription Drug User Fee Act was passed which aimed squarely at adding speed to the FDA mandate. In order to fund the increased work, PDUFA required drug companies to pay a fee for each drug they submit --- currently, as I understand it, around \ \$200k.

What happened? FDA nearly doubled in the number of reviewers and other workers it hired, and it successfully met its deadline speed goals.

The act was reauthorized in 1997 and 2002 and is currently up for reauthorization --- any day now.

After PDUFA I was approved attention soon turned back from speed to safety after a series of high profile drug withdrawal problems. And withdrawals were not the only issue. In general, there are two main theoretical criticisms of PDUFA. And I'll use some quotes from a recent open letter to the members of congress who are currently involved in the debate about reauthorizing the Act to articulate these two criticisms. (next slide)

Criticisms of PDUFA

Speed “In fact, **premature approval of a drug with safety problems**, or an inadequate means of detecting problems that emerge after marketing (as occurred with Vioxx) actually cost the taxpayers far more than the Treasury appears to save through user fees.”

Capture “**This negotiated arrangement [between the FDA and PhRMA]**, which has explicitly limited its ability to conduct post-marketing drug safety surveillance and other critical activities, **clearly diminishes the capacity of FDA to do its work on behalf of the nation.**”

Both quotes from “An Open Letter to: Chairman Edward Kennedy and Senator Mike Enzi, Chairman John Dingell and Representative Joe Barton Members of the Senate Health, Education, Labor and Pensions Committee Members of House Energy and Commerce Committee” by Angell, et al., March 14, 2007

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

8

These quotes are a very clear and measured articulation of the potential problems that the current regulatory structure could cause for drug safety related outcomes.

They come from an Open Letter sent recently by 22 leading scholars of the FDA (including Dan Carpenter) as illustration of the two main criticisms of the FDA --- which amount to a theory about what could be causing problems, and that the problems are.

The two mechanisms are (1) Speed which causes insufficient attention to drug approval, ``premature approval" is crucial here --- not precisely that certain drugs shouldn't be approved, but that FDA approval involves things like requirements about warnings and other cautionary activities along with the mere fact of approval plus learning among members of the clinical community about how to use the drug.

The second possible mechanism by which the new regime could cause drug safety problems is by capture --- that is, the fact that about a 1/3 of the FDA budget is now provided by industry that the FDA is supposed to regulate, and that industry has some say in how the FDA spends this money, can easily put the people at the FDA in a situation where they might pay more attention to the needs or desires of drug companies than to the demands of drug safety. (After all, it is not common for the regulatee to have a lot of say over the budget of the regulator precisely for these kinds of reasons).

The empirical question is how we would detect whether these particular mechanisms are causing problems with drug safety. For this talk, we are starting at the beginning of this process, by asking a prior question: were drug safety related outcomes related to the change in regulatory regime at all. If we can satisfy ourselves that we have a clear and unconfounded inference for this change, then we can take the next step to look at mechanisms. And this theory does imply that we should see more drug safety problems after PDUFA, whether the mechanism by which safety problems happen is speed or capture.

So, to be clear we are not testing whether speed or capture per se are having an effect --- PDUFA made many many changes occur at the FDA, and we are looking at the overall effect of all of these changes. We hope and plan to parse out the mechanism for such changes after we have more confidence in the results from merely looking at the effect of the change in law.

\textbf{aside} I have not found comparable counter-theories --- which state that drug safety has been enhanced by PDUFA, so although I have no ideological stake in this argument, I will not show a slide detailing an alternative explanation for drug safety problems.

When I say ``drug safety related problems", so far I've referred to withdrawals and used numbers that reflect what Public Citizen had to say, now let me be a bit more careful and tell you about the outcomes I use in this talk. (next slide)

Our Drug Safety Outcomes

Safety based withdrawals occurred 7 times for PDUFA I drugs vs. 4 for pre-PDUFA drugs. ($7 > 4$)

(but $7/114$ (.06) $\approx$ $4/85$ (.05) $\rightarrow p = .8$)

Changes in black box warnings after approval occurred 31 PDUFA I drugs vs. 23 for pre-PDUFA drugs. ($31 > 23$)

(but $31/114$ (.27) $\approx$ $23/85$ (.27) $\rightarrow p = 1$)

Adverse drug events after approval occurred (in our data) about 37,080 times for PDUFA I drugs vs. 21,046 for pre-PDUFA I drugs. ($37,080 > 21,046$)

(and $37,080/58$ (639) $\approx$ $21,046/30$ (702) $\rightarrow p = .8$)

PDUFA I drugs $\equiv$ “submitted and approved during PDUFA I”

pre-PDUFA I $\equiv$ “submitted and approved in the 4 years before PDUFA I”

Here I'll show the outcomes we examined --- the raw comparisons of which are just as provocative as the withdrawal information displayed by Public Citizen. Here the numbers are a bit different because I'm restricting the analysis to drugs both submitted and approved during PDUFA I (not just approved, since the approved numbers [which the GAO and Public Citizen use] may reflect drugs submitted under previous regimes and thus subject to different rules.

However, we see more withdrawals for drugs submitted and approved during PDUFA I than for drugs submitted and approved in the previous four years [I'll be happy to talk more about exactly why we chose 4 years previous as the control group --- suffice it to say now that going back beyond 4 years threatens to make the two groups of drugs look very different in terms of covariates like diseases treated, amount of medline articles, etc...] (advance)

Our Drug Safety Outcomes

Safety based withdrawals occurred 7 times for PDUFA I drugs vs. 4 for pre-PDUFA drugs. ($7 > 4$)

(but $7/114$ (.06) $\approx 4/85$ (.05) $\rightarrow p = .8$)

Changes in black box warnings after approval occurred 31 PDUFA I drugs vs. 23 for pre-PDUFA drugs. ($31 > 23$)

(but $31/114$ (.27) $\approx 23/85$ (.27) $\rightarrow p = 1$)

Adverse drug events after approval occurred (in our data) about 37,080 times for PDUFA I drugs vs. 21,046 for pre-PDUFA I drugs. ($37,080 > 21,046$)

(and $37,080/58$ (639) $\approx 21,046/30$ (702) $\rightarrow p = .8$)

PDUFA I drugs $\equiv$ “submitted and approved during PDUFA I”

pre-PDUFA I $\equiv$ “submitted and approved in the 4 years before PDUFA I”

Our Drug Safety Outcomes

Safety based withdrawals occurred 7 times for PDUFA I drugs vs. 4 for pre-PDUFA drugs. ($7 > 4$)

(but $7/114$ (.06) $\approx 4/85$ (.05) $\rightarrow p = .8$)

Changes in black box warnings after approval occurred 31 PDUFA I drugs vs. 23 for pre-PDUFA drugs. ($31 > 23$)

(but $31/114$ (.27) $\approx 23/85$ (.27) $\rightarrow p = 1$)

Adverse drug events after approval occurred (in our data) about 37,080 times for PDUFA I drugs vs. 21,046 for pre-PDUFA I drugs. ($37,080 > 21,046$)

(and $37,080/58$ (639) $\approx 21,046/30$ (702) $\rightarrow p = .8$)

PDUFA I drugs $\equiv$ “submitted and approved during PDUFA I”

pre-PDUFA I $\equiv$ “submitted and approved in the 4 years before PDUFA I”

And, the adverse drug reactions that we can observe were also more common for drugs submitted during PDUFA I than for drugs submitted just before the new regime was enacted. Let me explain briefly what these are: the FDA requires manufacturers to submit reports of drug problems, and doctors voluntarily submit these kinds of reports both directly to the FDA but also (and mainly, as I understand it) to the manufacturers themselves. There are thousands of such reports every year. And so there is a lot of variation to explain. (advance)

Our Drug Safety Outcomes

Safety based withdrawals occurred 7 times for PDUFA I drugs vs. 4 for pre-PDUFA drugs. ($7 > 4$)

(but $7/114 (.06) \approx 4/85 (.05) \rightarrow p = .8$)

Changes in black box warnings after approval occurred 31 PDUFA I drugs vs. 23 for pre-PDUFA drugs. ($31 > 23$)

(but $31/114 (.27) \approx 23/85 (.27) \rightarrow p = 1$)

Adverse drug events after approval occurred (in our data) about 37,080 times for PDUFA I drugs vs. 21,046 for pre-PDUFA I drugs. ($37,080 > 21,046$)

(and $37,080/58 (639) \approx 21,046/30 (702) \rightarrow p = .8$)

PDUFA I drugs $\equiv$ “submitted and approved during PDUFA I”

pre-PDUFA I $\equiv$ “submitted and approved in the 4 years before PDUFA I”

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

12

But, if we adjust these raw numbers by the numbers of drugs submitted and approved during those two periods, PDUFA qua PDUFA no longer seems as strongly related to big problems. But, to believe that this is now evidence that PDUFA had no effect would also be misleading. What if the kinds of drugs submitted after PDUFA tend to address more difficult diseases with possibly worse side effects; side effects which can be easily interpreted as adverse drug effects if they are bad enough. Results like these made it clear to us that one would have to do a lot more adjustment before we believed results claiming to show either that PDUFA had no effect or that it had a detrimental effect.

Some other scholars have done some related analyses of the Adverse Events data, but even with complex linear models these authors disagree with each other: with Olson showing using drugs approved more quickly tend to have more adverse drug reaction reports (using a sample of drugs mostly submitted before PDUFA), and Grabowski and Wang showing that this relationship disappears when they use a sample of drugs mostly from after PDUFA and a different specification. So, there is an empirical puzzle among scholars who have attempted to go beyond these raw comparisons, too.

Dan and the FDA team have chased down and focused some of these questions about approval time, using withdrawals and black box changes as outcomes --- focusing specifically on PDUFA and a whole set of preliminary analyses with lots of different methods of adjustment, have been done which actually do suggest that drugs which were approved just before the deadline (seemingly rushed) were more likely to have negative outcomes on withdrawals and black box warnings than drugs approved just after the deadline or just after submission.

One thing I should make sure to note here is that, even if we discover that there were discernibly more drug safety problems occurring after PDUFA than before it (using carefully constructed comparisons rather than these unadjusted figures), our work would NOT imply that PDUFA as a regulatory regime is somehow bad because so far we have not estimated (either roughly or carefully) the numbers of lives saved because of speedy approval of drugs. Many of the drugs the FDA approves are in essence unsafe drugs, but they are approved for use with lots of cautionary labeling and guidance because they save lives and are better than extant drugs. This means, by the way, when we think about withdrawals, we have to be aware that some of those withdrawals only occurred because alternative and safer drugs came on the market --- not necessarily because the withdrawn drug was so unsafe that it was not worth using at all.

Ok. So with those caveats. Now I'm going to give you a rough overview of our methodological approach. Then I'll highlight the particular methodological problems that we faced and list our current proposed solutions. This project has been particularly exciting because it has provided the opportunity for a lot of methodological development. However, rather than drag you through all the exciting details, I'll just announce what we did and then show you some different sets of results. I'll be more than happy to nerd-out in the Q\&A or anytime for the rest of the day. (next slide)

Definitions

Intervention Change in policy mandate. (written Z where $Z \in \{0, 1\}$)

Outcomes Withdrawals because of safety problems, changes in black box warnings, and adverse drug events. (written as potential outcomes (y_T, y_C)).

Sample New molecular entities submitted and approved both shortly before AND shortly after PDUFA I was implemented on Oct 29, 1992 (and also Nov 11, 1997 for PDUFA II).

The Causal Effect How many of these 7 withdrawals/31 blackbox changes/37,080 adverse events might have occurred even if PDUFA had not been passed? ($A = \sum Z_i(y_{Ti} - y_{Ci})$, where we assume $y_{Ti} \geq y_{Ci}$)

Bowers, Carpenter, Hansen

 $FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

13

We want to estimate the effect of PDUFA on safety based market withdrawals, changes in black box warnings, and adverse drug events.

First, we are going to ask a simpler question than the previous researchers. We just want to know the effect of the change in regulatory regime. Perhaps part of this effect will be the effect of time pressure on drug reviewers, but other parts might be the effect of hiring new drug reviewers as part of the PDUFA expansion, or other factors. In later analyses we hope to we'll look at the effect of time to approval, that is, time pressure. But we're starting for now just looking at the effect of PDUFA per se.

Second, we cast our question in the potential outcomes framework --- so we posit that each drug has an fixed outcome that we would observe if it were submitted during the PDUFA regime and another outcome that we would observe if it were submitted before the PDUFA regime.

Third, if at all possible we'd like to estimate this effect with simple comparisons so that we don't get into arguments about negative binomial models or beta priors or what have you --- we'd like to base our inference as closely as possible on a simple research design and to choose what we observe so as to closely approximate what we would see if we had randomly assigned drugs to be reviewed before and after the PDUFA regime change. We're going to make this simple comparison by comparing drugs submitted, reviewed, and approved after (w/ deadline) to before. We feel confident making use of the discontinuity induced by the law because (1) the law was debated on and passed over only a 3 month or so period just before it was implemented in 1992 and we know that firms would have trouble rushing or holding back drugs in Stage III trials since those timeframes are agreed upon in advance with the FDA and (2) that in equilibrium firms will rush to submit drugs as fast as possible and this new law doesn't change that behavior and (3) we can check for balance on observed covariates (which we did and found that 4 years is about right).

Fourth, we define the treatment effect at the level of the individual drug (τ_i) as the comparison of its potential outcomes, and our overall causal estimand is just the number of potential outcomes in reaction to the regulatory regime. This assumption that Rosenbaum calls non-negativity here is something I can talk about in the discussion session --- now, I'll just note that nobody is claiming that PDUFA enhanced safety outcomes, only that it didn't hurt safety outcomes. If this is wrong, we'd know it and can estimate effects with a different assumption. Turns out that we have no evidence arguing against non-negativity.

Now, I want spend two slides quickly showing you the particular challenges we've faced in estimating causal effects of PDUFA and the particular solutions that we've come up with to address these challenges before I show you some results. (next slide)

Major Challenges to Our Causal Inference

Problem of Ignorability: Rubin et al. have shown most causal comparisons assume $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$, $0 < p(Z = 1 | \mathbf{X}) < 1$ (either explicitly or tacitly) .

Absence of unmeasured confounding $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$?

Presence of adequate comparisons $0 < p(Z = 1 | \mathbf{X}) < 1$?

Problem of Probability Pretense In order test hypotheses, we must pretend we know

the likelihood: $y = f(\mathbf{X}, Z)$ and DGP for y (i.e.

$y \sim N(\mu, \sigma^2)$ or $y \sim \text{Bernoulli}(\pi)$) **OR**

the assignment process: that our study was randomized within matched set (i.e. DGP for $Z | \mathbf{X}$).

Most of you are familiar with the struggles necessary to justify claims about ignorability. Here, I just want to point out that there are two pieces to this claim.

So first we face the overall problem of justifying why we think that it is ok to compare the, say, black box warning changes on one drug submitted and approved after PDUFA I with those which have occurred on another drug submitted and approved before PDUFA I. This overall problem has two parts: (advance)

Major Challenges to Our Causal Inference

Problem of Ignorability: Rubin et al. have shown most causal comparisons assume $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$, $0 < p(Z = 1 | \mathbf{X}) < 1$ (either explicitly or tacitly) .

Absence of unmeasured confounding $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$?

Presence of adequate comparisons $0 < p(Z = 1 | \mathbf{X}) < 1$?

Problem of Probability Pretense In order test hypotheses, we must pretend we know

the likelihood: $y = f(\mathbf{X}, Z)$ and DGP for y (i.e.

$y \sim N(\mu, \sigma^2)$ or $y \sim \text{Bernoulli}(\pi)$) **OR**

the assignment process: that our study was randomized within matched set (i.e. DGP for $Z | \mathbf{X}$).

Major Challenges to Our Causal Inference

Problem of Ignorability: Rubin et al. have shown most causal comparisons assume $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$, $0 < p(Z = 1 | \mathbf{X}) < 1$ (either explicitly or tacitly) .

Absence of unmeasured confounding $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$?

Presence of adequate comparisons $0 < p(Z = 1 | \mathbf{X}) < 1$?

Problem of Probability Pretense In order test hypotheses, we must pretend we know

the likelihood: $y = f(\mathbf{X}, Z)$ and DGP for y (i.e.

$y \sim N(\mu, \sigma^2)$ or $y \sim \text{Bernoulli}(\pi)$) OR

the assignment process: that our study was randomized within matched set (i.e. DGP for $Z | \mathbf{X}$).

Then, we need to be able to claim that some of the PDUFA drugs could have been submitted before PDUFA and some of the pre-PDUFA drugs could have been submitted after PDUFA. For example, we might be unable to make this claim if we had some drug from 1950 which would never be approved today in our control group --- the probability that a drug with it's covariates would be submitted after PDUFA is essentially zero.

As an aside, I was very lucky here because Dan and a team of graduate students have worked incredibly hard to assemble a bunch of covariates ("bold X") which we use throughout our analysis here --- they include things like dummies for disease treated, number of people hospitalized with the primary indication, 4 different content analysis coding of mentions of the drug from medline [two of which focus specifically on safety in clinical trials], media coverage of the clinical trials about the drug, whether the firm submitted more than one drug at this time, etc...
(advance)

Major Challenges to Our Causal Inference

Problem of Ignorability: Rubin et al. have shown most causal comparisons assume $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$, $0 < p(Z = 1 | \mathbf{X}) < 1$ (either explicitly or tacitly) .

Absence of unmeasured confounding $\mathbf{y}_T, \mathbf{y}_C \perp Z | \mathbf{X}$?

Presence of adequate comparisons $0 < p(Z = 1 | \mathbf{X}) < 1$?

Problem of Probability Pretense In order test hypotheses, we must pretend we know

the likelihood: $y = f(\mathbf{X}, Z)$ and DGP for y (i.e.

$y \sim N(\mu, \sigma^2)$ or $y \sim \text{Bernoulli}(\pi)$) OR

the assignment process: that our study was randomized within matched set (i.e. DGP for $Z | \mathbf{X}$).

Second, we not only wish to estimate treatment effects without worry about confounding or extrapolation, we also want to answer the question: "Could this effect be due to chance?" Bringing chance into the project requires us to play pretend a bit. Either we have to pretend that we know how the outcomes, covariates, and assignment to PDUFA are related structurally (like is y just a linear model of treatment plus covariates) PLUS something about the chance process leading to y (like is y just a bernoulli random variable or a negative-binomial or what). OR, we can answer this question about chance by pretending that, within matched set, the drugs were randomly assigned to PDUFA versus pre-PDUFA. So, I call this a problem because it requires us to assume something, and thus requires some special attention.

Now, I want to show you how we have addressed these different challenges. (next slide)

(for later if asked) We also need to assume SUTVA (that the odds of one drug getting submitted during PDUFA instead of before PDUFA was implemented are independent of the odds of another drug getting exposed to the PDUFA regime) among drugs with the same covariates. This assumption may be violated because drug companies are competing with one another --- so the fact that one company submits a diabetes drug to the FDA might easily mean that another company might soon submit a diabetes drug.

However, it is also my understanding that the equilibrium behavior of drug companies is just to rush rush rush their drugs to the FDA. So, although they are competing, what they are really doing is just moving as fast as possible --- this means that we don't feel that it is ridiculous to maintain the SUTVA assumption for now.

Proposed Solutions

...to the challenges facing causal inference about the effect of PDUFA on drug safety

- Strengthen the argument in favor of $y_T, y_C \perp Z | X$ by (1) **choosing** drugs submitted and approved close to the discontinuity in the law and (2) **adjusting** the data by matching on propensity scores.
→ Induces a **problem of precision** by discarding 1300 drugs.
- Bolster the argument in favor of $0 < p(Z = 1 | X) < 1$ by using **shrinkage** through (1) prognostic scores (Hansen 2007) and (2) Bayesian glm (Firth; Gelman).
- Address the **problem of precision** by using (1) optimal, full matching (using `optmatch` by Hansen 2004) and (2) prognostic scores (Hansen 2007).
- Estimate effects in two different frequentist **probability frameworks**:
(1) the likelihood based approach using mixed effects models and
(2) the randomization based approach (following and extending Rosenbaum 1999, 2001 using the `RIttools` software package (Bowers and Hansen, 2006)).

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

18

Our goal: Find the Experiment in the Observational Data.

So, the first and most important challenge to overcome involved comparing like with like without observed confounds. We addressed this challenge by using the discontinuity induced by the change in the law as a kind of assignment of treatment (to the PDUFA I regime, for example). In addition we matched drugs submitted and approved during PDUFA I to drugs submitted and approved before PDUFA to each other using propensity scores. Our balance assessments showed that we couldn't distinguish these matched sets from those which would be produced if assignment to PDUFA had occurred at the hands of an experimenter. And, our information about drug company behavior indicates that the equilibrium response of a drug company to learning about 2–3 months before PDUFA was implemented that it was coming was to not change their behavior --- but just to continue to submit drugs as fast as possible.

Of course, by limiting the set of drugs that we think are comparable to the drugs submitted during PDUFA I, we lose power --- this decision to solve the most important problem of comparability induces another problem with the precision of our estimates of causal effects. (advance)

Proposed Solutions

...to the challenges facing causal inference about the effect of PDUFA on drug safety

- Strengthen the argument in favor of $y_T, y_C \perp Z | X$ by (1) **choosing** drugs submitted and approved close to the discontinuity in the law and (2) **adjusting** the data by matching on propensity scores.
→ Induces a **problem of precision** by discarding 1300 drugs.
- Bolster the argument in favor of $0 < p(Z = 1 | X) < 1$ by using **shrinkage** through (1) prognostic scores (Hansen 2007) and (2) Bayesian glm (Firth;Gelman).
- Address the **problem of precision** by using (1) optimal, full matching (using `optmatch` by Hansen 2004) and (2) prognostic scores (Hansen 2007).
- Estimate effects in two different frequentist **probability frameworks**:
(1) the likelihood based approach using mixed effects models and
(2) the randomization based approach (following and extending Rosenbaum 1999, 2001 using the `RIttools` software package (Bowers and Hansen, 2006)).

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

19

The next piece of the problem of justifying the ignorability assumption was also tricky. If you just estimate propensity scores using simple logistic regressions you get a situation where big hunks of drugs are predicted to almost certainly NOT be submitted after PDUFA versus big groups of drugs predicted to almost certainly be submitted after PDUFA. This problem is called separation and is a well known problem with logistic regression.

For us, separation causes the problem of finding adequate comparisons. This creates its own precision problem because, if you just used those propensity scores you'd want to exclude a lot of cases from analysis. This problem of logistic regression is well known, however, but in our case Ben realized that there was a way to take into account some of those 1300 drugs without biasing the careful comparisons that we created with matching, a way to do this which ends up preventing too the propensity score models from behaving badly. In addition, we used a couple of well known Bayesian methods for fixing this bad behavior of logistic regression. (advance)

Proposed Solutions

...to the challenges facing causal inference about the effect of PDUFA on drug safety

- Strengthen the argument in favor of $y_T, y_C \perp Z | X$ by (1) **choosing** drugs submitted and approved close to the discontinuity in the law and (2) **adjusting** the data by matching on propensity scores.
→ Induces a **problem of precision** by discarding 1300 drugs.
- Bolster the argument in favor of $0 < p(Z = 1 | X) < 1$ by using **shrinkage** through (1) prognostic scores (Hansen 2007) and (2) Bayesian glm (Firth;Gelman).
- Address the **problem of precision** by using (1) optimal, full matching (using optmatch by Hansen 2004) and (2) prognostic scores (Hansen 2007).
- Estimate effects in two different frequentist **probability frameworks**:
(1) the likelihood based approach using mixed effects models and
(2) the randomization based approach (following and extending Rosenbaum 1999, 2001 using the RIttools software package (Bowers and Hansen, 2006)).

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

20

Next, we had a sample with more treated units (like 200) than control units (like 100). And we were already worried about ultra large confidence intervals. To this end we eschewed pair matching (or matching with a fixed set size) and used full, optimal matching using software developed by Ben. It also turns out that prognostic scores have the potential to increase the precision of our estimates beyond providing shrinkage in our propensity score equations --- the general idea is similar to including predictors of y which have no relationship with X in an OLS model --- this decreases the variance in y left unexplained, and therefore tightens confidence intervals without affecting bias. Of course, we don't do exactly this since otherwise we'd be giving up on the benefits of using the discontinuity, sample choice, and matching, but the analogy gives the rough idea. (advance)

Proposed Solutions

...to the challenges facing causal inference about the effect of PDUFA on drug safety

- Strengthen the argument in favor of $y_T, y_C \perp Z | X$ by (1) **choosing** drugs submitted and approved close to the discontinuity in the law and (2) **adjusting** the data by matching on propensity scores.
→ Induces a **problem of precision** by discarding 1300 drugs.
- Bolster the argument in favor of $0 < p(Z = 1 | X) < 1$ by using **shrinkage** through (1) prognostic scores (Hansen 2007) and (2) Bayesian glm (Firth; Gelman).
- Address the **problem of precision** by using (1) optimal, full matching (using optmatch by Hansen 2004) and (2) prognostic scores (Hansen 2007).
- Estimate effects in two different frequentist **probability frameworks**:
(1) the likelihood based approach using mixed effects models and
(2) the randomization based approach (following and extending Rosenbaum 1999, 2001 using the RItools software package (Bowers and Hansen, 2006)).

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

21

Finally, we have to have some probabilistic framework within to make sense of statements like "Could these effects be due to chance?" We particularly like the randomization or neo-Fisherian framework (and we've been working this last year to implement it in software) since we are already pretending that, within sets, we can make comparisons as if they drugs had been randomly assigned to PDUFA vs. pre-PDUFA --- we don't have to make any extra burdensome assumptions to make the move to estimating effects and getting confidence intervals in this framework.

However, in this talk, I'll present the results of linear models (which requires the additional assumptions) mainly because everyone else here are using linear models and they are thus easy to understand and because our results using fewer and different assumptions were basically the same. So, we solve the problem of choosing a way to play pretend by not choosing but just doing it both ways.

So, now I will turn to the results. I'm happy to talk about any piece of our workflow in the discussion section.

So, I've already shown rough unadjusted results that suggest that PDUFA had no effect on either withdrawals, blackbox changes, or adverse events.

Here I will show the results after all of the adjustment and checking that I just alluded to. After I show these results I'll highlight some conceptual problems that we've encountered, especially with using withdrawals. But for now, I want to show some results so that I don't tax your patience anymore than I already have --- then I'll engage in self-criticism and use that self-criticism to motivate and raise some more substantive questions for discussion.

(next slide)

Predicted Safety Problems (PDUFA I-prePDUFA I)

from models of the form $E(y|Z, X) = g(\beta_1 Z + \text{matched set random effects})$

- Additional **post-marketing blackbox revisions**: -1 (95% CI:-10 to 18)
- Additional **safety based withdrawals**: 2 (95% CI: -1 to 31)
- Additional **adverse event reports per million prescriptions**: -4,443 (95% CI: -53,789 to 44,381)

Posterior predictions based on 2000 draws from 3 independent chains.

Assuming $y|Z, \mathbf{X} \sim \text{Bernoulli}(\pi)$ for withdrawals and blackbox warnings, $y|Z, \mathbf{X} \sim \text{Normal}(\mu, \sigma^2)$ for adverse events, $\beta \sim \text{Normal}(\cdot)$ in all models, and canonical links.

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

22

So, today I'll only present results using the likelihood-style framework. We've used randomization inference on the withdrawals information previously and have found the same general patterns of responses. The overall message of this slide is that we can't distinguish any effects from zero at the 95% level.

For example, the predicted difference in safety based withdrawals between the PDUFA I drugs and the pre-PDUFA I sample is 2: 2 of the 7 observed we'd attribute to PDUFA, while the other 5 are from other reasons. Of course, the uncertainty around this estimate is large (if we resampled a different set of drugs from the imaginary superpopulation of drugs many times, 95% of the time we'd assume that the true estimate would be between -1 and 31 withdrawals). We'd also predict between 10 less and 18 more blackbox revisions because of PDUFA I. And the story for adverse events is similar.

Now, let me engage in some quick self-criticism on a couple of grounds: (1) Notice that our sample only contains approved drugs. We don't observe drugs that were submitted and not approved (because this rarely if ever happens, can talk more about how drug companies deal with problematic clinical trials later --- I don't know a lot about this, but I'd like to know more). But, the theory which is critical of the PDUFA act would not say that we'd expect these drugs to be approved but withdrawn, but that the drugs shouldn't have been approved in the first place if the FDA was functioning as desired. That is, the public policy relevant counterfactual for withdrawals is not really represented in our data. Of course, we never really directly observe the counterfactual, but the whole process of sample choice and matching adjustment was meant to approximate this counterfactual --- what would have happened to these PDUFA drugs which were withdrawn if they had instead been submitted pre-PDUFA.

What about for the other outcomes: I think these outcomes are on firmer ground with regard to whether we can observe the counterfactual. PDUFA can cause safety problems by "premature approval" --- that is, if approval had been extended in time, or perhaps the regulators were not captured by the industry, then perhaps more cautionary documentation about the use of the drugs would have been produced --- measured as a lack of need to change labels after approval --- or perhaps the groups of doctors who had been involved in clinical trials and who are early adopters have more chance to learn more about safe use of the drug (thereby decreasing Adverse Events). That said, while more conceptually clear than the evidence about withdrawals this current analysis does not support that theory.

Another self-criticism is about whether or not all drugs submitted and approved after PDUFA really got a real "dose" of PDUFA (including speed and capture but also other changes due to the law --- doubling in size of FDA etc...). Well, I protected a bit against this excluding any backlog from the treatment group. However, it is possible that these results are biased toward zero since perhaps much of what went on during the first 5 years of PDUFA was a lot like what went on during the 4 years before it. This would be like the spill-over effects that Kathleen mentioned in her talk on pay for performance.

I tried to address this criticism by making two different comparisons, using drugs submitted and approved during the second PDUFA law [during which the deadlines speeded up to 10 months for standard drugs] as my treatment group. (next slide)

Predicted Safety Problems (PDUFA II)

from models of the form $E(y|Z, X) = g(\beta_1 Z + \text{matched set random effects})$

- Additional **post-marketing blackbox revisions**:
vs. PDUFA I -3 (95% CI: -5 to 5)
vs. pre-PDUFA I -7 (95% CI: -4 to 3)
- Additional **safety based withdrawals**:
vs. PDUFA I 0 (95% CI: -1 to 8)
vs. pre-PDUFA I 1 (95% CI: -0 to 24)
- Additional **adverse event reports per million prescriptions**:
vs. PDUFA I 28,601 (95% CI: 4,134 to 52,409)
vs. pre-PDUFA I 58,987 (95% CI: -10,031 to 129,560)

Posterior predictions based on 2000 draws from 3 independent chains.

Assuming $y|Z, \mathbf{X} \sim \text{Bernoulli}(\pi)$ for withdrawals and blackbox warnings, $y|Z, \mathbf{X} \sim \text{Normal}(\mu, \sigma^2)$ for adverse events, $\beta \sim \text{Normal}(\cdot)$ in all models, and canonical links.

I compared this new treatment group to two control groups: (1) the drugs which were considered "Treated" in the previous analysis (ie. submitted and approved during PDUFA I) and (2) the control group from the PDUFA I analysis. The idea here is that by 1997 the PDUFA regime was in full swing, every drug submitted after that implementation date would get a full dose, and therefore my effects wouldn't be biased toward zero.

In the end, though, most of the results are the same as before in broad outlines --- no effects. We do see a CI bounded away from zero for adverse events when compared to PDUFA I, but not when compared to pre-PDUFA I drugs. This lack of congruency doesn't give me much confidence that I've really found a big effect that is not fragile.

In addition, I should note that the Adverse Events data have been a huge headache, generating a lot of missing data based on no pattern that we can discern --- however, since we know that data missing on outcomes can really bias your results unless it is missing at random, I'm a bit worried about these data. Whereas we have about 200 drugs total to analyze for the blackbox and withdrawal data, we only have about 100 drugs when it comes to the adverse event reports because of the trouble we had merging it with FDA data. I'm happy to get advice about this merging after the talk. My discussions with the folks at AHRQ were fairly depressing on this account.

So, let me move on to the end. (next slide)

Summary

- **Choosing** drugs in a small window around the discontinuity of the PDUFA I implementation date + **adjusting** the data using full matching + using shrinkage = strong evidence in favor of ignorability.
- So far we can't distinguish an effect from zero in this preliminary work.
- The policy relevant counterfactual for withdrawals is unclear and perhaps incorrect.

Past analyses of these kinds of data have been plagued with worries about (or ought to have been plagued with worries about) extrapolation and comparing like with like --- our strategy here bases our arguments in favor of lack of observed confounding and comparability first and foremost on decisions we made about the research design and then strengthens these arguments using matching.

Despite the fact that we are compelled by the theoretical arguments about the problems posed by PDUFA, we don't find any evidence to support this theory. As I said earlier, we are asking here merely whether the change in regime was enough to matter --- not taking an indepth look at deadlines --- although we hope to marry these two sides of the analysis soon. And we expect that looking beyond mere comparisons of the regulatory regime will yield results that are more informative about the mechanisms linking the changes at the FDA to possible post-marking drug safety problems. (next slide)

Plans and Questions

Next actions:

- Narrow focus from “effect of PDUFA” to “effect of approval right before the deadline” or “dose of approval time”.
- Explore alternative outcomes: Does quicker approval lead to less learning in the clinical community about use of the drug?
- Estimate effects using randomization inference rather than linear models to avoid making assumptions about $y|Z, \mathbf{X}$.

Questions we have:

- How do we get around the problem of not observing non-approved drugs? Clinical trials data?
- What is the right pre-PDUFA comparison group for the post-PDUFA drugs which had post-marketing problems?
- Where can we find data on this comparison group?

Bowers, Carpenter, Hansen

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

25

Here are some of our current plans for extension of this work and some questions that we are struggling with:

We want to start look more at the hypothesized mechanism by which PDUFA is supposed to have caused problems with post-approval drug safety, and to look at alternative indicators of good performance of the FDA.

I also think a weakness of the results I just presented stems from the fact that we had to play pretend in the likelihood framework --- especially a problem for Adverse Events for which the DGP would be quite a guess (?negative binomial? ?mixture of poissons and Gammas? or what?) So, we'll do randomization inference as well, which requires other assumptions of course, but requiring different ones will allow us to feel more confident that our results reflect the world outside rather than inside our own heads.

We also plan to do formal sensitivity analysis in the style developed by Paul Rosenbaum.

We finally have some questions for you all: We are particularly interested in how to use analyses of the clinical trials pipeline to answer these kinds of questions. Susan Moffitt has been gathering this data. But so far we have not turned to conceptualizing and thinking through what such an analysis would look like (for example, what are the potential outcomes in such an analysis? approval itself? or what?)

Are there are other kinds of data which would provide information about policy relevant counterfactual for different outcomes that would help us evaluate this law per se as well as the many changes in the FDA induced by this law (that go beyond the speed and worries about capture)?

Thanks!

The Foundation, Everyone at the Site, The Aspen Attendees 2005-2007, The FDA Working Group at Harvard (Chris Afendulis, Dan Carpenter, Ben Hansen (at Umich), and Susan Moffitt with important cameos from Justin Grimmer, Clayton Nall, and Evan Zucker), Ed Miller at MEPS/AHRQ.

Finally, I am happy to embrace the social norm of expressing public gratitude. I've already and will continue to express my thanks privately since it would be unseemly to propose a big group hug right now, and since hugs work best with fewer than 100 participants. The last two years have been both very challenging and nurturing and the process of confronting those challenges and basking in the nurturance has been very intellectually, professionally, and personally productive as well as fun. Thanks to you all!

Adjustment Workflow

...given a dataset and choice of analysis sample

Adjustment

- 1 Fit a prognostic score $\psi(\mathbf{X}_C) = E(y_c|\mathbf{X}_C)$ for each outcome using the historical controls.
- 2 Apply the coefficients generated using the historical controls to the analysis sample to get $\hat{\psi}(\mathbf{X})$ for each model.
- 3 Estimate a propensity score including the prognostic scores $\hat{\phi}(\mathbf{X}, \hat{\psi}(\mathbf{X}))$.
- 4 Full match on standardized $\hat{\phi}$ with calipers of 3 sd. stratified by standard vs. priority status
- 5 Assess the extent to which covariate balance given our matched sets can be distinguished from that which would be obtained from random assignment.

Randomization Estimation Workflow

...given matched sets and/or strata

Estimation

- 1 Specify sharp null hypotheses for $A = 0 \dots \text{max compatible } A$
- 2 Adjust the outcomes to reflect no effect if the hypothesis were true.
- 3 Test all these hypotheses for each hypothesized A or use separability (Gastwith, Krieger, Rosenbaum 2000, Bowers and Hansen 2007) to identify two to test.
- 4 The set of hypotheses accepted at some α is a $100(1 - \alpha)$ CI.

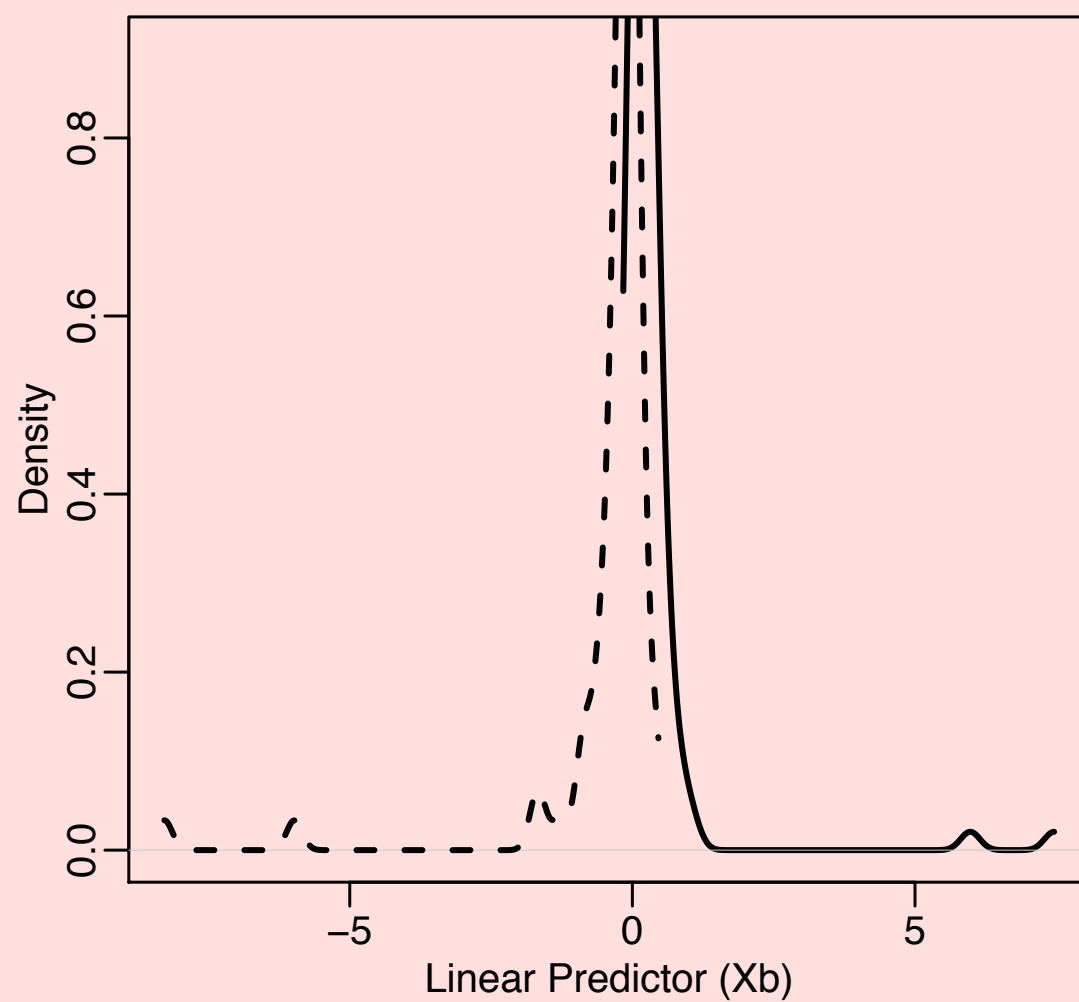
Pre-treatment Covariates

Media Coverage Before PDUFA I, Incidence of Primary Indication, Indicator for Incidence of Primary Indication Missing, Death Rate for Primary Indication, Indicator for Death Rate for Primary Indication Missing, # Hospitalizations for Primary Indication, Order of Entry Into the Market, # Submissions by Firm, Indicator for # Submissions by Firm Missing, # of mentions of the drug during clinical trials in Medline overall, # Medline mentions alluding to the safety of the drug, indicators for the firm of submission, indicators for primary indication.

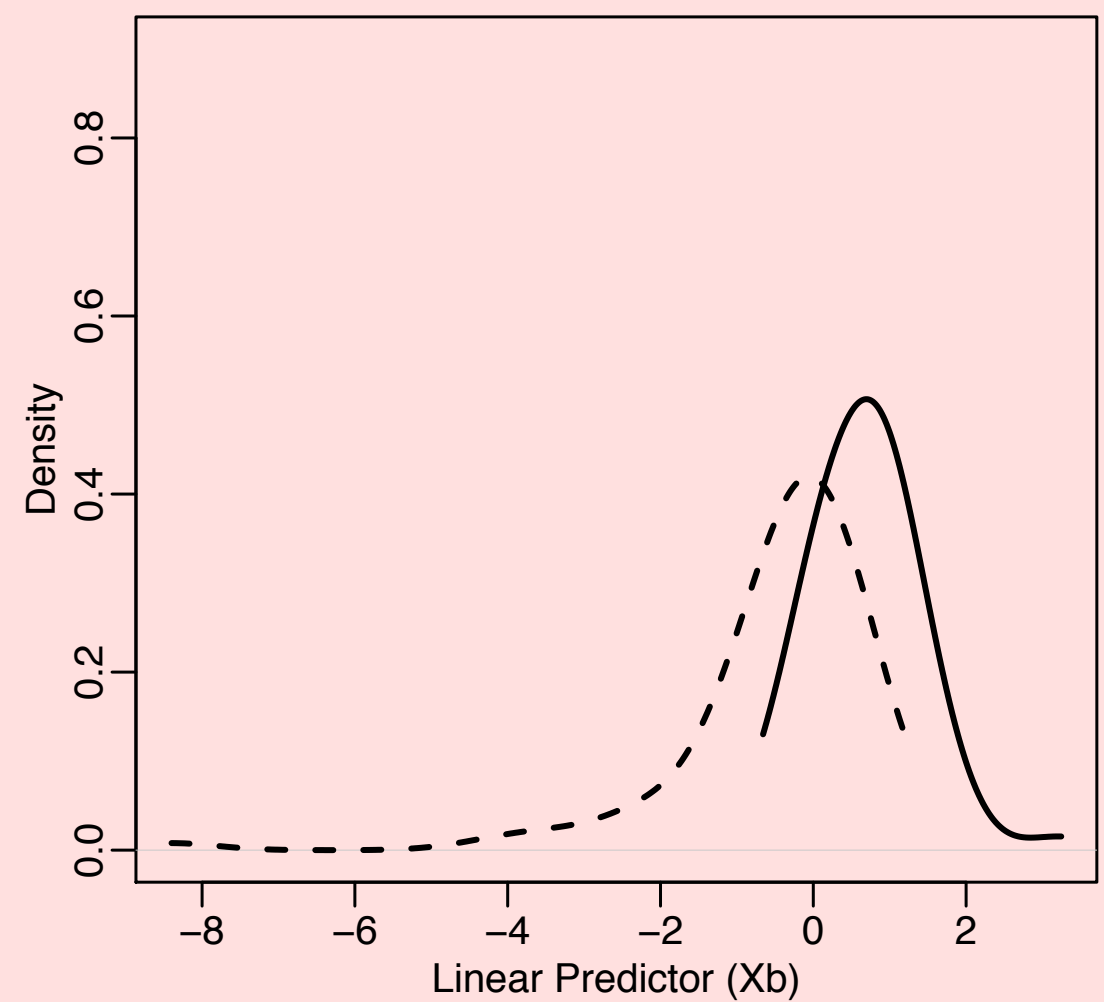
Ignorability and the Distributions of Propensity Scores

$$\mathbf{y}_T, \mathbf{y}_C \perp \mathbf{Z} | \mathbf{X}, 0 < p(Z = 1 | \mathbf{X}) < 1 \quad (1)$$

Propensity Score



Propensity Score with Shrinkage



Propensity Score

Bowers, Carpenter, Hansen

Propensity Score with Shrinkage

$FDA_{\text{post-PDUFA}} - FDA_{\text{pre-PDUFA}} = ?$

Prognostic Scores (Hansen 2007)

Definition: A *propensity score* is $\phi(\mathbf{X}) = E(Z|\mathbf{X})$ and a *prognostic score* is $\psi(\mathbf{X}) = E(Y_c|\mathbf{X})$. Both reduce the dimensions of $\mathbf{X}$, and thus matching on either is more feasible than directly, and exactly, matching on $\mathbf{X}$.

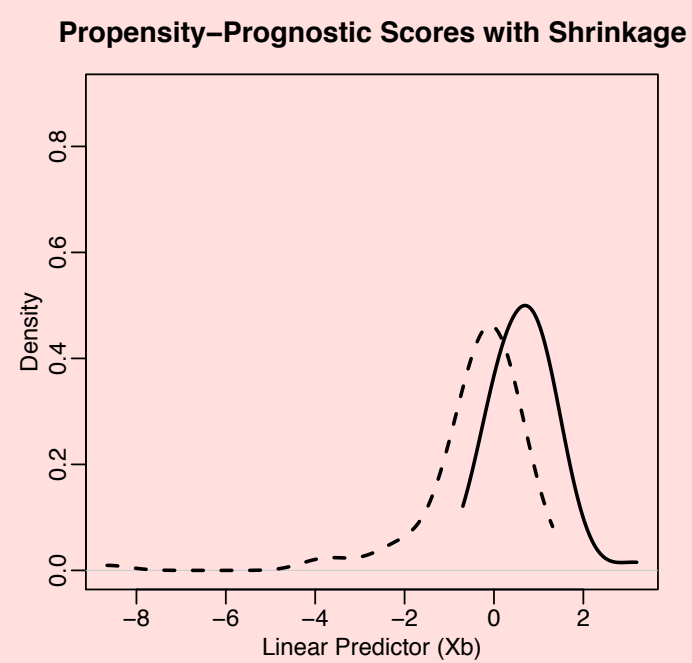
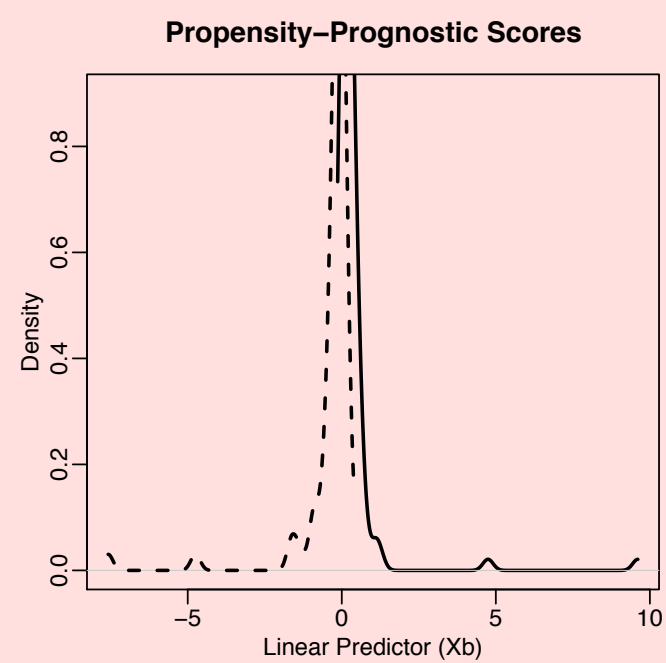
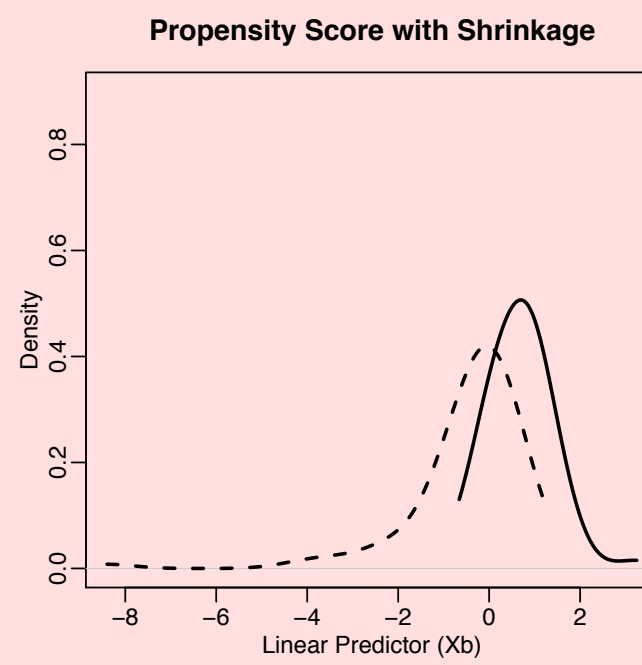
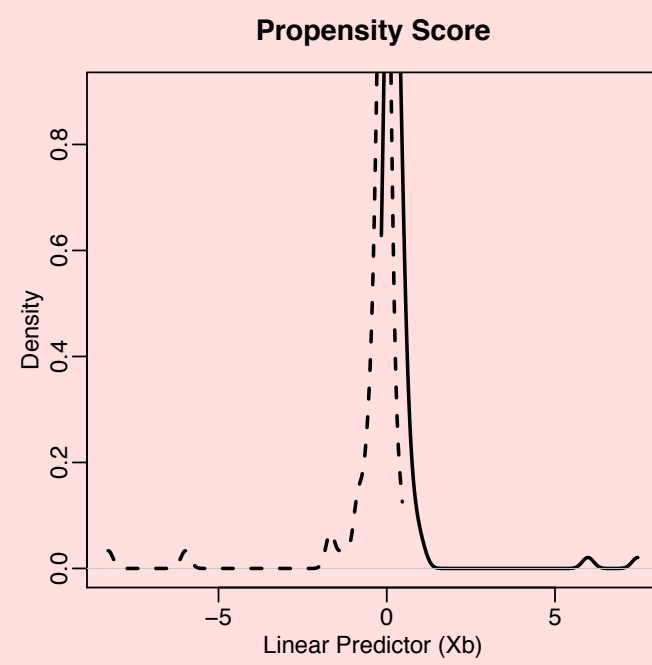
Motivations:

- ① 1290 potential control group drugs excluded from analysis by choice of temporal window and
- ② Separation in glm estimation of $E(Z|\mathbf{X})$ and hope to improve on br1r (Jeffrey's Prior) or bayesglm(Cauchy prior) estimation.

Prognostic Scores In Practice

- Use `glm` or `bayesglm` to predict withdrawals among drugs submitted between 1943 and 1988. We used three different models: (1) same as propensity model; (2) ... but including all continuous variables as 5 knot regression splines; (3) same as 1 but allowing all 2 way interactions.
- Apply those coefficients to the analysis sample (1988 to 1997) to get $\hat{\psi}(\mathbf{X})$ for each model.
- Estimate $\hat{\phi}(\mathbf{X}, \hat{\psi}_1(\mathbf{X}), \hat{\psi}_2(\mathbf{X}), \hat{\psi}_3(\mathbf{X}))$.
- Match on $\hat{\phi}$

Propensity Scores Including Prognostic Scores



Full Matching in Practice

- 1 Generate a matrix of distances between treated and control units. Include Inf to exclude pairs from matching (e.g. pairs for which the distance is larger than a caliper. Here the caliper is 3 sd).

```
> round(ps2distlistbglm[[1]],2)[1:10,1:10]
      174 175  176  188  233  250  251  274  338 399 ...
31  2.09 2.4  1.28 0.81 0.16 1.76 1.71 1.55 0.47 2.7 ...
71  1.83 2.1  1.02 0.55 0.10 1.50 1.45 1.30 0.21 2.4 ...
72  1.77 2.0  0.95 0.49 0.16 1.44 1.39 1.23 0.15 2.3 ...
73  1.66 1.9  0.84 0.38 0.27 1.33 1.28 1.12 0.04 2.2 ...
98   Inf Inf   Inf   Inf   Inf   Inf   Inf   Inf   Inf Inf ...
99   Inf Inf   Inf   Inf   Inf   Inf   Inf   Inf   Inf Inf ...
109 2.83 Inf  2.02 1.55 0.90 2.51 2.45 2.30 1.21 Inf ...
125 2.85 Inf  2.04 1.58 0.92 2.53 2.47 2.32 1.23 Inf ...
147 2.89 Inf  2.07 1.61 0.96 2.56 2.51 2.35 1.27 Inf ...
177 0.92 1.2  0.10 0.36 1.01 0.59 0.54 0.38 0.70 1.5 ...
.   .   .   .   .   .   .   .   .   .   .
.   .   .   .   .   .   .   .   .   .   .
```

- 2 Find the sets of rows (treated) versus columns (controls) which minimize the distance across the whole distance matrix (`fullmatch(ps2distlistbglm)`).
- 3 Test Balance

Matched Set Configuration

70 Sets, No Excluded Observations

```
> table(smallldat[good2bglm,"pdufaF"],thefm2bglm[good2bglm,drop=TRUE])
```

	Pr.1	Pr.10	Pr.11	Pr.12	Pr.13	Pr.14	Pr.16	Pr.17	Pr.18	Pr.2	Pr.25	Pr.3	Pr.30	Pr.31
PostDiscont	5	2	2	1	1	12	1	1	5	1	3	1	4	1
PreDiscont	1	1	1	1	2	1	1	1	1	3	1	1	1	1

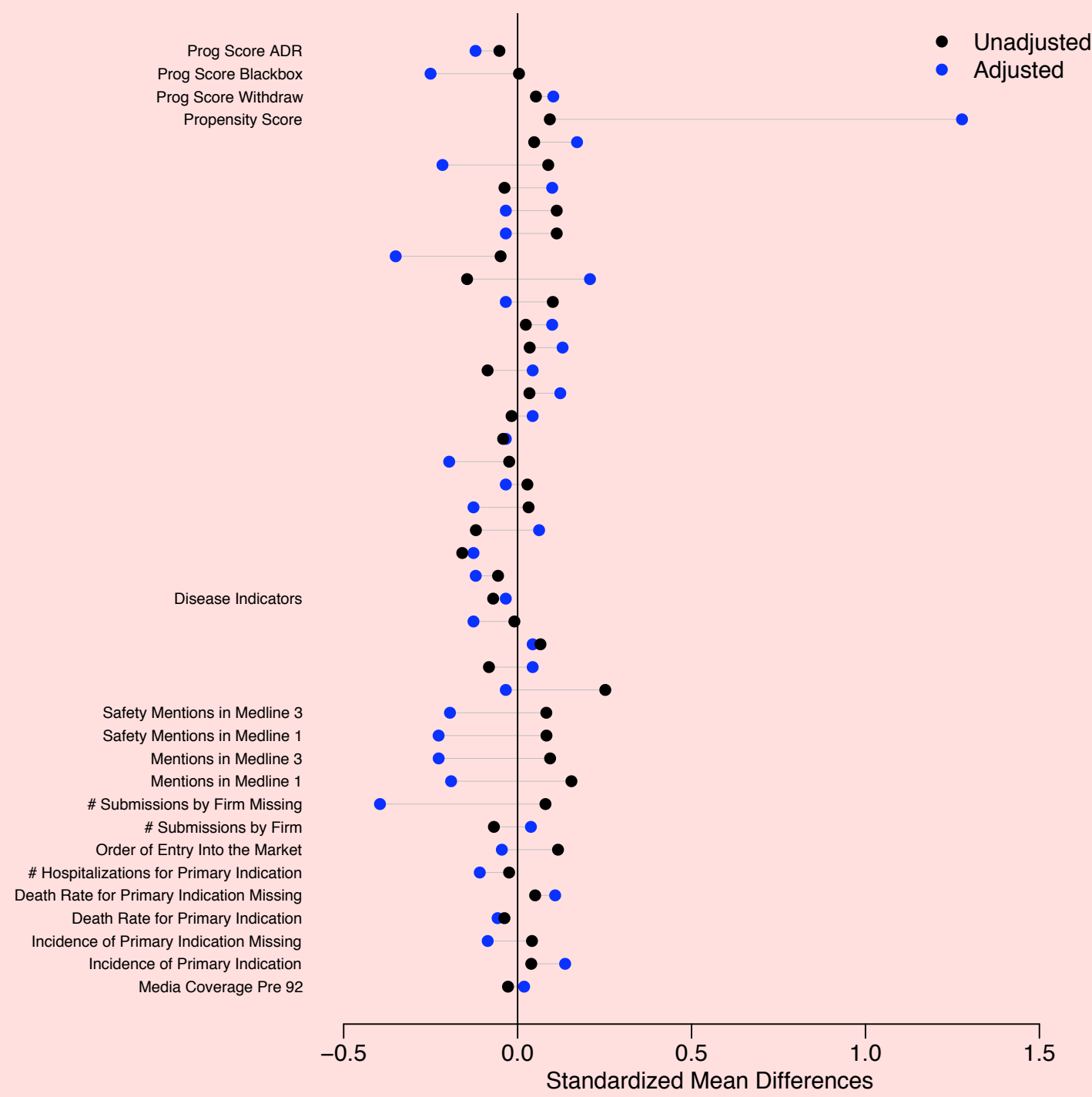
	Pr.38	Pr.4	Pr.42	Pr.43	Pr.46	Pr.49	Pr.6	Pr.7	Pr.8	Pr.9	St.1	St.10	St.100	St.101	St.11
PostDiscont	1	1	1	1	1	1	1	1	1	3	4	1	1	1	1
PreDiscont	1	1	1	2	1	1	1	5	4	1	1	1	2	1	2

	St.13	St.14	St.15	St.17	St.18	St.19	St.2	St.21	St.24	St.25	St.27	St.28	St.3	St.30
PostDiscont	1	3	9	2	1	1	12	4	4	1	4	3	9	3
PreDiscont	2	1	1	1	3	1	1	1	1	1	1	1	1	1

	St.32	St.34	St.4	St.48	St.49	St.5	St.50	St.51	St.54	St.56	St.69	St.70	St.86	St.95
PostDiscont	2	2	5	1	1	12	1	1	1	1	1	1	2	1
PreDiscont	1	1	1	2	1	1	1	3	1	3	1	2	1	1

	St.96	St.97	St.98	St.99
PostDiscont	1	1	1	1
PreDiscont	1	1	1	2

Balance on Covariates



PDUFA took time to implement

During the first few years of PDUFA I, we eliminated backlogs of original applications and supplements that had formed in earlier years when the program had fewer resources. Phased in over the 5 years of PDUFA I, the goals were to review and act on 90 percent of priority new drug applications (NDAs), biologics license applications (BLAs), and efficacy supplements (i.e., submissions for products providing significant therapeutic gains) within 6 months of submission of a complete application; to review and act on 90 percent of nonpriority original NDAs, BLAs, and efficacy supplements within 12 months, and on resubmissions and manufacturing supplements within 6 months. Over the course of PDUFA I, we exceeded all of these performance goals. (Federal Register, 16 Jan 2007, 1743-1753)

Worries about the Linear Model as Commonly Practiced

- ① Extrapolation, Fragility and Sensitivity to specification, (i.e. strong assumptions required for causal inference)
- ② Incorrect coverage of CIs (i.e. strong assumptions required for probabilistic inference)
- ③ For BOTH: Worry about data snooping, requires assumptions about $Y|f(Z, X)$
- ④ Requires stories about Super Populations.

Distributions of Balance Measures

Standardized, Adjusted Mean Diff's and eQQs

