Published in final edited form as:
Pharmacoepidemiol Drug Saf. 2013 September ; 22(9): . doi:10.1002/pds.3480.

A Critical Review of Methods to Evaluate the Impact of FDA Regulatory Actions

Becky A. Briesacher, PhD[1,2], **Stephen B. Soumerai, ScD**[3], **Fang Zhang, PhD**[3], **Sengwee Toh, ScD**[3], **Susan E. Andrade, ScD**[1,2], **Joann L. Wagner, MSW**[2], **Azadeh Shoaibi, MS, MHS**[4], and **Jerry H. Gurwitz, MD**[1,2]

[1]Division of Geriatric Medicine, University of Massachusetts Medical School, Worcester, MA

[2]Meyers Primary Care Institute, a joint endeavor of Fallon Community Health Plan, Reliant Medical Group and the University of Massachusetts Medical School, Worcester, MA

[3]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston MA

[4]Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD

Abstract

Purpose—To conduct a synthesis of the literature on methods to evaluate the impacts of FDA regulatory actions, and identify best practices for future evaluations.

Methods—We searched MEDLINE for manuscripts published between January 1948 and August 2011 that included terms related to FDA, regulatory actions, and empirical evaluation; the review additionally included FDA-identified literature. We used a modified Delphi method to identify preferred methodologies. We included studies with explicit methods to address threats to validity, and identified designs and analytic methods with strong internal validity that have been applied to other policy evaluations.

Results—We included 18 studies out of 243 abstracts and papers screened. Overall, analytic rigor in prior evaluations of FDA regulatory actions varied considerably; less than a quarter of studies (22%) included control groups. Only 56% assessed changes in the use of substitute products/services, and 11% examined patient health outcomes. Among studies meeting minimal criteria of rigor, 50% found no impact or weak/modest impacts of FDA actions and 33% detected unintended consequences. Among those studies finding significant intended effects of FDA actions, all cited the importance of intensive communication efforts. There are preferred methods with strong internal validity that have yet to be applied to evaluations of FDA regulatory actions.

Conclusions—Rigorous evaluations of the impact of FDA regulatory actions have been limited and infrequent. Several methods with strong internal validity are available to improve trustworthiness of future evaluations of FDA policies.

Corresponding Author: Becky A. Briesacher, PhD Division of Geriatric Medicine University of Massachusetts Medical School Biotech Four, Suite 315 377 Plantation Street Worcester, MA 01605 Phone: 508-856-3495 Fax: 508-856-5024 Becky.Briesacher@umassmed.edu.

Conflicts of Interest: No authors have conflicts of interest to declare.

Prior Postings and Presentations: Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multifaceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. A related White Paper was developed as part of this initiative and can be found at: http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1038

Keywords

FDA; Regulatory Actions; Evaluation Methodology

Introduction

The U.S. Food and Drug Administration (FDA) is responsible for protecting the public from the adverse effects of medical products, including communicating safety risks that emerge after product approval.[1] The FDA communicates safety risks in a variety of ways. The agency may issue regulatory public health advisories or safety alerts directed to the general public, or dear healthcare professional letters directed to prescribers. It may also modify medical product labeling to include boxed warnings of serious safety risks, or require a Risk Evaluation and Mitigation Strategy (REMS); in some cases, the product may be withdrawn from the market.[2] To date, the FDA has communicated hundreds of safety risks, and these communications are expected to increase with the Sentinel Initiative, its effort to develop a national system for active monitoring of medical product safety.[3-5]

The effectiveness of the FDA's safety risk communications is yet unclear. A recent review of the FDA's safety risk advisories for prescription drugs detected a variety of intended and unintended effects on product use and health outcomes.[6] However that review did not consider the quality of research methods when summarizing the evidence, and acknowledged limitations due to the heterogeneity of study design across evaluations.

The quality of the research methods used to evaluate the impact of FDA regulatory actions has not been assessed, and there are currently no recommendations offered for conducting robust policy evaluations. The current state-of-the art policy evaluations offer quasi-experimental research designs such as interrupted time series and regression discontinuity that share many design elements of experiments including the use of control groups, pre-test post-test measures, intention-to-treat analyses, and explicit mechanisms for inferring counterfactual scenarios (i.e., what would the outcome have been without the FDA action).[7] Novel statistical methods have also been developed to address special circumstances such as sequential analyses when the data on outcomes are still accumulating, or extended Cox models when the hazard ratios of risk will vary with time.[8-10]

The goal of this study was to understand how state-of-the-art methods have been used to evaluate the impact of FDA regulatory actions, and how these methods could be used to improve future evaluations. Our specific objectives were: 1) to characterize prior research approaches used to evaluate FDA regulatory actions, particularly the strengths and weaknesses of previous methodologies; and 2) to recommend designs and statistical approaches that may be useful for evaluating FDA regulatory actions and describe the advantages and disadvantages of each approach. This work was undertaken as part of the Mini-Sentinel program, a pilot project sponsored by the FDA under the Sentinel Initiative to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products.[3,11]

Methods

We conducted this study in two phases. In the first phase (described in greater detail below), we performed a literature review of all published studies and relevant internal documents on FDA regulatory actions (provided by FDA. From these studies, we selected those with adequate research design and approaches that could provide reasonably valid effect estimates in our final summary of the published literature.

In the second phase of this project, we convened a consensus panel to identify preferred methodologies for evaluating FDA regulatory actions based on their prior work in policy evaluation and findings from phase one literature review. These methodologies were selected based on their ability to control for threats to internal validity, statistical performance, as well as practical considerations. Preference was given to methods that could be used with administrative health plan data, such as found in the Mini-Sentinel Distributed Database.

Methods for Literature Review

Search Strategy

We conducted the literature search utilizing Ovid MEDLINE (see Appendix 1 for search strategy and results). We searched for literature published between 1948 and the first week of August, 2011, which included the term "United States Food and Drug Administration" and an FDA regulatory action (e.g., labeling change (label$), warning (warn$), or REMS); 195 published articles were identified. Additionally, FDA investigators identified relevant seminal literature (published and internal FDA literature) for inclusion in the review. We also manually searched the bibliographies of reviewed papers to identify any additional relevant literature (there was some overlap in the results of these search strategies) resulting in a total of 243 published papers and internal FDA documents for review.

Selection Criteria

We drafted selection criteria for abstract and full-text manuscript review to identify relevant studies. This process was iterative until the entire project team agreed on the selection criteria. Selection criteria are outlined in Appendix 2 and are an adaptation of the Cochrane standards for defining study selection criteria.[12] These included: (a) the adequacy of the research designs to control for threats to validity; (b) the execution of the studies through a careful assessment of their risk of bias; (c) the adequacy of methods for addressing selection bias and confounding; and (d) the omission of reporting biases, including selective reporting of outcomes.

Review of Literature

The review of the literature was a three-step process. First, all available abstracts were reviewed by at least 1 of 3 authors (BAB, SEA, FZ). Second, all literature that met inclusion criteria or could not be excluded on the basis of the abstract underwent a full-text review by 2 of 3 authors (BAB, SEA, FZ). The third step was employed in the case of discrepancies between the two authors' full reviews, in which a third author reviewed the full-text article and the trio discussed their decisions with the workgroup until consensus was reached. Table 1 summarizes the reasons for exclusion.

Methods for Consensus Panel

Selection Criteria and Conduct of Interviews

To identify methods that may be suitable for evaluating the impact of FDA regulatory actions, we conducted a series of interviews with our panel. The panel consisted of six authors (BAB, SBS, FZ, ST, SEA, JHG) with extensive experience in drug policy research and evaluation methodology. Our main methodology was a generalized version of the Delphi method which helps systematically prioritize and synthesize the experiences and knowledge of a panel of experts to achieve group consensus.[13] First, we established the following eligibility criteria: 1) The evaluation methodology should use research designs and analytic methods that produce strong internal validity, as defined by Shadish, Cook, and Campbell (2002).[7] Thus, we selected methods that incorporated control groups and pre-

policy measures to support a counterfactual inference about what would have happened in the absence of the FDA regulatory action. 2) The evaluation methodology should offer pragmatic utility. Therefore, we did not select methods requiring random assignment or primary data collection since both options currently seemed unlikely in the Mini-Sentinel program. As a result, our list of suitable methods to conduct evaluations of FDA regulatory actions assumes that only secondary data are available for the evaluation, in either aggregate-level or person-level data formats. 3) The evaluation methodology may address special circumstances, particularly for the following cases: a) when post-policy data are still accumulating; b) the policy is transient; and c) the data capture is limited to only a few time points. These special circumstance methods are acceptable as only preliminary evaluations until more rigorous evaluations are available. 4) The evaluation methodology may include state-of-the-art approaches that theoretically have potential for unbiased regulatory evaluations, although in practice they require strong assumptions. These methods are acceptable only in sensitivity analyses within the context of a rigorous study design.

Given these criteria, we polled all the panel members for contributions and then created a list of candidate methods. We discussed each of the nominated methods and filtered out redundancies and irrelevant content through an iterative process. We assembled a list of the most suitable methods and then characterized each method by necessary data elements, strengths, and weaknesses. We also identified applied examples published in the literature that could serve as models.

Results for Literature Review

The literature review produced a total of 18 independent studies that evaluated the impact of FDA regulatory actions and met the criteria for inclusion.[14-31] The key characteristics of these articles are summarized in Table 2. In summary, these studies were published over a period of two decades (1987 to 2011), and the investigations ranged in size from a small study[18] with only 1,308 individuals to a large population-level study[20] generalizable to 61 million individuals (the study used a mix of sampling methods). The studies evaluated a wide range of patient populations, including children and adolescents in 78%[14-18,20-22,24,25,27-29,31] of the studies and adults over the age of 65 in 50%[14-18,25,28,30,31] of the studies.

FDA Policies and Medical Products Evaluated

Most of the studies evaluated the impact of FDA boxed warnings (50%),[16-18,20,23,25,27,29,31] although other FDA policies were also evaluated including labeling changes,[16,18,19,27,28,30] advisory/safety warnings,[20-26] dear doctor/healthcare professional letters,[17,19,28,30] and product withdrawals.[14,15,19] The studies examined policies relating to a variety of medical products, although antidepressants were most commonly evaluated (44%).[20-25,27,29] Other medical products included antipsychotic agents,[28,30] antidiabetic drugs,[19,26] cisapride,[17] Drug Efficacy Study Implementation (DESI) drugs,[14] pemoline,[18] propoxyphene,[31] terfenadine,[16] and zomepirac.[15]

Data Source and Unit of Analysis

The most common source of data for the FDA policy evaluations was privately-owned proprietary administrative databases available for purchase (61%).[17,18,20-25,27,29,31] The next most common sources were publicly available federal databases such as state Medicaid, Veteran's Administrative, and TRICARE data (33%)[14,15,19,26,28,30] followed by private health insurance or managed care claims data (6%).[16,18] Just over a quarter of the evaluations used individual-level data (28%).[18,19,27,29,30] Assessments of data completeness

or quality were usually not well described but were at least mentioned in most (89%)[14,15,17,18,20-31] of the papers.

Primary Endpoints

The outcomes measured in the evaluations varied depending on the regulatory action of interest; few evaluations included measures related directly to patient health. In evaluations of boxed warnings or withdrawn products (n=12), the studies usually assessed changes in the use of substitute medical products and services (92%).[14-19,23,25,27,29,31] In evaluations of safety warnings,[14-31] some of the studies examined changes in the occurrence of recommended laboratory monitoring (17%).[18,19,27] Adverse events were measured in two studies (11%),[20,31] and contraindicated medical drugs/conditions in two other studies (11 %).[16,17]

Research Design and Analytic Methods

Overall, the study designs fell into the single broad category of quasi-experimental designs. Control groups were used in less than half of the studies (39%).[15,22,24,25,28,30,31] The rigor and sophistication of the analytic methods varied greatly, ranging from segmented regressions in all of the studies[14-31] to unadjusted t-tests of differences (6%).[16] Only two studies (11%)[28,30] described explicit sensitivity analyses to address potential confounders.

Study Outcome Results

Almost half of the 18 studies found no impact[15,18,24,26-28,31] or weak/modest impacts[18,19] of FDA regulatory actions on at least one measure of targeted medical products (see Figure 1). Eleven studies (61%) found significant impacts on at least one of the intended outcomes of the policy.[14,16,17,20-22,26-30] One-third of the studies[14,15,20,21,23,25] also detected unintended consequences of FDA regulatory actions; these included spillover effects to non-indicated patient populations,[23,25] increases in substitute prescribing[14,15] (both positive and negative substitutes), decreases in disease detection and treatment,[21,25] and increases in negative health outcomes.[20]

A comparison of differences between evaluations finding minimal to no impact of FDA regulatory actions and evaluations finding significant impacts revealed few consistent patterns in study design, analytic framework, or measures. For instance the use of interrupted time series was present in both null finding and positive finding evaluations. An exception was the outcomes measure of safety monitoring, which was consistently associated with no impact. Additionally, in all cases where significant intended impacts were detected, the investigators noted intensive communication efforts including widespread media attention as a key to successfully changing prescribing patterns.

Results for Consensus Panel

The consensus panel identified seven research designs and analytic methods suitable for evaluations of FDA regulatory actions (see Table 3). Under the category of research designs and analytic methods with strong internal validity, we identified two quasi-experimental designs ---the interrupted times series and the regression discontinuity design --- and a statistical method ---the extended Cox model. All three approaches offer methodological advantages in their strong ability to control for many potential threats to validity and all of these novel methods can incorporate experimental design elements including intention-to-treat analyses and control groups. Each is briefly described below.

Recommended Study Designs and Analytic Methods

In the case of the interrupted time series design,[7] the pre-intervention data series is used to decrease uncertainty about whether a policy intervention is associated with true change in outcomes. The key assumption of this design is that extrapolating the pre-intervention level and trend correctly reflects the (counterfactual) outcome that would have occurred had the intervention not happened. Interrupted times series may be conducted with person-level or aggregate-level longitudinal data. Data series may be as short as 18 points (9 pre and 9 post) depending on the minimum effect size to be detected, and the variation.[32] But the design becomes stronger with additional time points to establish trends. This study design is among the most robust of the quasi-experimental designs, but it requires a priori knowledge of the diffusion process, especially in case of gradual interventions or delayed causation. In our review, all 18 of the identified studies evaluating FDA regulatory actions incorporated an interrupted time series design.[14-31]

In the case of the regression discontinuity design,[7] changes in the pattern of the outcome at a pre-specified cutoff in an assignment variable are used to indicate a policy impact. This approach is most appropriate for a policy that applies to a patient population defined by a continuous measure with a fixed threshold (e.g., age ≥65). The regression discontinuity design is a reasonably robust quasi-experimental design and may be especially useful when pre-intervention data are lacking or limited. This approach also requires knowledge of the correct functional form between the outcome and assignment variable, and it is less well known in the medical literature. Our literature review did not identify any study with a regression discontinuity design to assess FDA policies.

The extended Cox model[10] uses heaviside functions to yield constant hazard ratios for different time intervals. This statistical method offers the analytic advantages of maximum likelihood (e.g., good large-sample properties) and can be used in various longitudinal study designs. This model has the ability to produce different hazard ratios of outcome that are specific to the preand post-intervention periods. Extended Cox models do not require the proportional hazards assumption for the entire study period and they are robust to right censored data. Our literature review did not identify any study that used an extended Cox model to assess FDA policies.

Methods that May be Suitable for Special Circumstances

We identified three approaches under this category: sequential analysis of gradually accruing data,[9] difference-in-difference-in differences[33] and self-controlled case series.[34,35] These methods may produce valid inferences for some evaluations, especially if they explicitly control for pre-intervention trends. Each is briefly described below.

Sequential analysis of gradually accruing data offers methodological advantages under the special circumstance when post-policy data are still accumulating.[8,9] In this approach, the counterfactuals can be obtained from data from the pre-policy period (i.e., historical controls), or data from a concurrent control group not affected by the policy. A signal is generated if the likelihood ratio --- calculated based on the ratio of the number of observed events to the number of expected events (the counterfactuals) --- exceeds a predetermined value. The key aspect of this method is that the p-values are adjusted for looking at the data in a repeated fashion, or multiple testing. An appropriate example for this method would be an FDA policy that is associated with a potential safety concern that requires timely evidence on the impact of the policy. Ideally, like the interrupted time series design, the validity is increased with 9 or more pre-intervention points in order to establish the counterfactual trend in the absence of the intervention.

The *difference-in-difference-in differences* design may be suitable under the special circumstance of limited time points in the data.[33] A key component of this method is the use of multiple contrasts to isolate the true effect of the intervention. An appropriate example for this method would be an FDA policy that is assessed with only one or two pre- and post-intervention measurements of the outcome, but any observed differences are evaluated against any changes in two relevant comparators.

Self-controlled case series may be useful in the case of a transient policy involving an acute outcome.[34,35] In this method, the study uses data on only the cases to examine the temporal association between a time-varying exposure and outcome. The exact scenario for the best use of this method is unclear; however, it may be suitable for studying the association between a temporal FDA policy (e.g., temporary suspension of a medical product) and an acute event, in which the risk periods are short.

State-of-the-art approach with Theoretical Potential if Assumptions Can be Met

We identified approaches that incorporate confounder summary scores (e.g., propensity scores) under this category. This analytic approach is best used with caution in sensitivity analyses until the plausibility of assumptions are better established. Confounder summary scores are often used in the setting where considerable confounding is suspected, so it must be clear that the score is created with a model that is correctly specified and not biased by important omitted variables. One promising application is within a time series study design and the confounder score is used to help construct the counterfactual estimates.[36]

Discussion

Our synthesis of the literature produced the following conclusions. First, rigorous evaluations of the impact of FDA regulatory actions have been infrequent, especially relative to the large number of actions implemented by the FDA. Second, when FDA regulatory actions were evaluated adequate research design standards were often not employed. For instance, less than a quarter used a control group. Limitations in important measures were common (e.g., inadequate capture of suicidality in administrative data. Third, many of the assessments were limited in scope and often examined only changes in the use of the targeted medical product. Studies that included a broader array of measures such as unintended impacts (e.g., increases in the use of substitute products and services) often found significant effects. Studies with outcome measures relating directly to patient health and adverse events were even rarer. As a result, many evaluations of FDA regulatory actions used suboptimal research designs and analytic methods, making the results limited and susceptible to biased findings. Fourth, among studies employing valid study designs, no impact or weak/modest impact was detectable in 50% of the studies. Intensive media communication was the one consistent factor in all studies finding significant impacts of FDA regulatory actions. Overall, this review revealed considerable gaps in the evidence-base and promising methods to evaluate FDA regulatory policies. This conclusion is consistent with the findings of an earlier review of FDA drug risk communications.[6]

Our expert panel review produced the following conclusions. There are proven methods that have been widely used in evaluating non-FDA policies or interventions, especially quasi-experimental research designs, which may be useful for evaluating the impact of FDA regulatory actions. For example, research designs and analytic methods with strong internal validity include the interrupted time series, regression discontinuity design and extended Cox models. Also, aggregate data approaches are necessary for data sharing in the Mini-sentinel and may be used with appropriate and robust methods to evaluate FDA regulatory actions. The methods identified by this panel can address many of the limitations of previously published evaluations. The careful and consistent application of these methods

can provide a valuable opportunity to achieve the goal of less biased assessments of FDA regulatory actions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank the FDA Workgroup Leads and Members: Judith Racoosin (Lead), Azadeh Shoaibi (Lead), Andrea Sutherland, Antonio Paredes, Chris Jones, Danica Marinac-Dabic, David J. Graham, George Neyarapally, George Rochester, Helen Sullivan, Jan Johannessen, Mark Levenson, Marsha Reichman, Melissa Robb, Michael Kieffer, Michie Hunt, Nilsa Loyo-Berrios, Thomas Buttolph.

Sponsors: This work was conducted as part of the Mini-Sentinel program. The Mini-Sentinel program is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

References

1. Murphy S, Roberts R. “Black box” 101: How the Food and Drug Administration evaluates, communicates, and manages drug benefit/risk. J Allergy Clin Immunol. Jan; 2006 117(1):34–39. [PubMed: 16387581]
2. U.S. Food and Drug Administration. [accessed 1 August 2012] Guidance drug safety information-FDA’s communication to the public. 2012. Rev.1:http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation /Guidances/ucm072281.pdf
3. Platt R, Carnahan RM, Brown JS, et al. The U.S. Food and Drug Administration’s Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf. Jan; 2012 21(Suppl 1):1–8. [PubMed: 22262586]
4. Robb MA, Racoosin JA, Sherman RE, et al. The US Food and Drug Administration’s Sentinel Initiative: expanding the horizons of medical product safety. Pharmacoepidemiol Drug Saf. Jan; 2012 21(Suppl 1):9–11. [PubMed: 22262587]
5. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf. Jan; 2012 21(Suppl 1):23–31. [PubMed: 22262590]
6. Dusetzina SB, Higashi AS, Dorsey ER, et al. Impact of FDA drug risk communications on health care utilization and health behaviors: a systematic review. Med Care. Jun; 2012 50(6):466–478. [PubMed: 22266704]
7. Shadish, WR.; Cook, TD.; Campbell, DT. Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton Mifflin Company; Boston, MA: 2002.
8. Kulldorff M, Davis R, Kolczak M, et al. A maximized sequential probability ratio test for drug and vaccine safety surveillance. Seq Anal. 2011; 30(1):58–78.
9. Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. Stat Med. Jan 30; 2010 29(2):284–295. [PubMed: 19941282]
10. Kleinbaum, DG.; Klein, M. Survival Analysis: A self-Learning Text. Second ed. Springer; New York, NY: 2005.
11. Mullard A. Unleashing the mini-sentinel. Nat Rev Drug Discov. Apr; 2012 11(4):255–257. [PubMed: 22460110]
12. Cochrane Handbook for Systematic Reviews of Interventions. The Cochrane Collaboration and John Wiley & Sons Ltd; England: 2008.
13. Dalkey, NC. The Delphi method: An experimental study of group opinion. In: Rourke RL, DL.; Snyder, D., editors. Studies in the Quality of Life: Delphi and Mecision-making. Lexington Books; Lexington, MA: 1972. p. 13-54.
14. Soumerai SB, Ross-Degnan D, Gortmaker S, Avorn J. Withdrawing payment for nonscientific drug therapy. Intended and unexpected effects of a large-scale natural experiment. JAMA. Feb 9; 1990 263(6):831–839. [PubMed: 2404151]

15. Ross-Degnan D, Soumerai SB, Fortess EE, et al. Examining product risk in context. Market withdrawal of zomepirac as a case study. JAMA. Oct 27; 1993 270(16):1937–1942. [PubMed: 8411550]
16. Thompson D, Oster G. Use of terfenadine and contraindicated drugs. JAMA. May 1; 1996 275(17):1339–1341. [PubMed: 8614120]
17. Weatherby LB, Nordstrom BL, Fife D, et al. The impact of wording in “Dear doctor” letters and in black box labels. Clin Pharmacol Ther. Dec; 2002 72(6):735–742. [PubMed: 12496755]
18. Willy ME, Manda B, Shatin D, et al. A study of compliance with FDA recommendations for pemoline (Cylert). J Am Acad Child Adolesc Psychiatry. Jul; 2002 41(7):785–790. [PubMed: 12108802]
19. Cluxton RJ Jr. Li Z, Heaton PC, et al. Impact of regulatory labeling for troglitazone and rosiglitazone on hepatic enzyme monitoring compliance: findings from the state of Ohio medicaid program. Pharmacoepidemiol Drug Saf. Jan; 2005 14(1):1–9. [PubMed: 15546159]
20. Gibbons RD, Brown CH, Hur K, et al. Early evidence on the effects of regulators’ suicidality warnings on SSRI prescriptions and suicide in children and adolescents. Am J Psychiatry. Sep; 2007 164(9):1356–1363. [PubMed: 17728420]
21. Libby AM, Brent DA, Morrato EH, et al. Decline in treatment of pediatric depression after FDA advisory on risk of suicidality with SSRIs. Am J Psychiatry. Jun; 2007 164(6):884–891. [PubMed: 17541047]
22. Nemeroff CB, Kalali A, Keller MB, et al. Impact of publicity concerning pediatric suicidality data on physician practice patterns in the United States. Arch Gen Psychiatry. Apr; 2007 64(4):466–472. [PubMed: 17404123]
23. Valuck RJ, Libby AM, Orton HD, et al. Spillover effects on treatment of adult depression in primary care after FDA advisory on risk of pediatric suicidality with SSRIs. Am J Psychiatry. Aug; 2007 164(8):1198–1205. [PubMed: 17671282]
24. Morrato EH, Libby AM, Orton HD, et al. Frequency of provider contact after FDA advisory on risk of pediatric suicidality with SSRIs. Am J Psychiatry. Jan; 2008 165(1):42–50. [PubMed: 17986680]
25. Libby AM, Orton HD, Valuck RJ. Persisting decline in depression treatment after FDA warnings. Arch Gen Psychiatry. Jun; 2009 66(6):633–639. [PubMed: 19487628]
26. Stewart KA, Natzke BM, Williams T, et al. Temporal trends in anti-diabetes drug use in TRICARE following safety warnings in 2007 about rosiglitazone. Pharmacoepidemiol Drug Saf. Nov; 2009 18(11):1048–1052. [PubMed: 19655338]
27. Busch SH, Frank RG, Leslie DL, et al. Antidepressants and suicide risk: how did specific information in FDA safety warnings affect treatment patterns? Psychiatr Serv. Jan; 2010 61(1):11–16. [PubMed: 20044412]
28. Morrato EH, Druss B, Hartung DM, et al. Metabolic testing rates in 3 state Medicaid programs after FDA warnings and ADA/APA recommendations for second-generation antipsychotic drugs. Arch Gen Psychiatry. Jan; 2010 67(1):17–24. [PubMed: 20048219]
29. Valluri S, Zito JM, Safer DJ, et al. Impact of the 2004 Food and Drug Administration pediatric suicidality warning on antidepressant and psychotherapy treatment for new-onset depression. Med Care. Nov; 2010 48(11):947–954. [PubMed: 20856141]
30. Kales HC, Zivin K, Kim HM, et al. Trends in antipsychotic use in dementia 1999-2007. Arch Gen Psychiatry. Feb; 2011 68(2):190–197. [PubMed: 21300946]
31. Soumerai SB, Avorn J, Gortmaker S, et al. Effect of government and commercial warnings on reducing prescription misuse: the case of propoxyphene. Am J Public Health. Dec; 1987 77(12): 1518–1523. [PubMed: 3674250]
32. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. J Clin Epidemiol. Nov; 2011 64(11):1252–1261. [PubMed: 21640554]
33. Woolridge, J. Econometric Analysis of Cross Section and Panel Data. The MIT Press; Cambridge, MA: 2002.
34. Whitaker H. The self controlled case series method. BMJ. 2008; 337:a1069. [PubMed: 18755768]

35. Whitaker HJ, Farrington CP, Spiessens B, et al. Tutorial in biostatistics: the self-controlled case series method. Stat Med. May; 2006 30; 25(10):1768–1797.

36. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. J Eval Clin Pract. Dec; 2011 17(6):1231–1238. [PubMed: 20973870]

37. Serumaga B, Ross-Degnan D, Avery AJ, et al. Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. BMJ. 2011; 342:d108. [PubMed: 21266440]

38. Kozhimannil KB, Adams AS, Soumerai SB, et al. New Jersey's efforts to improve postpartum depression care did not change treatment patterns for women on medicaid. Health Aff (Millwood). Feb; 2011 30(2):293–301. [PubMed: 21289351]

39. Grootendorst PV, O'Brien BJ, Anderson GM. On becoming 65 in Ontario. Effects of drug plan eligibility on use of prescription medicines. Med Care. Apr; 1997 35(4):386–398.

40. Briesacher BA, Soumerai SB, Field TS, et al. Medicare part D's exclusion of benzodiazepines and fracture risk in nursing homes. Arch Intern Med. Apr 26; 2010 170(8):693–698. [PubMed: 20421554]

41. Zhang Y, Adams AS, Ross-Degnan D, et al. Effects of prior authorization on medication discontinuation among Medicaid beneficiaries with bipolar disorder. Psychiatr Serv. Apr; 2009 60(4):520–527. [PubMed: 19339328]

42. Yih WK, Kulldorff M, Fireman BH, et al. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink project. Pediatrics. May; 2011 127(Suppl 1):S54–64. [PubMed: 21502252]

43. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Stat Methods Med Res. Feb; 2009 18(1):7–26. [PubMed: 18562396]

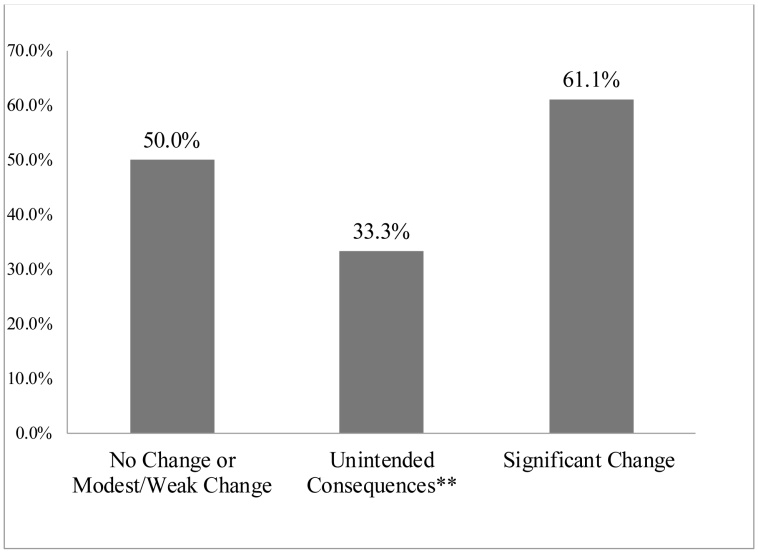
44. Afendulis CC, He Y, Zaslavsky AM, et al. The impact of Medicare Part D on hospitalization rates. Health Serv Res. Aug; 2011 46(4):1022–1038. [PubMed: 21306369]

45. Morrato EH, Newcomer JW, Kamat S, et al. Metabolic screening after the American Diabetes Association's consensus statement on antipsychotic drugs and diabetes. Diabetes Care. Jun; 2009 32(6):1037–1042. [PubMed: 19244091]

46. Wharam JF, Graves AJ, Landon BE, et al. Two-year trends in colorectal cancer screening after switch to a high-deductible health plan. Med Care. Sep; 2011 49(9):865–871. [PubMed: 21577162]

Key Points

- Rigorous evaluations of the impact of FDA regulatory actions have been limited and infrequent.
- Methods with strong internal validity are available to improve future evaluations of FDA policies.



*Among 18 studies with methods and designs meeting minimal criteria for generating valid evidence; sum of bars exceed 100% due to multiple outcomes.
**Increased use of substitute medical products or services, or negative health outcomes.

Figure 1.
Summary of Evidence* on the Impact of FDA Regulatory Actions on Prescribing Patterns and Recommended Monitoring

Table 1

Selection Criteria Review Results

	N
Reviewed Articles or Abstracts	243
Excluded Articles by Criteria	
Review/descriptive/opinion/audits/pre-post	89
FDA regulation not focus of study	80
No outcomes endpoint	29
Not medical product	15
Prospective/primary data collection (surveys)	7
Not US data	5
Final Articles	18

Table 2

Literature Review Results (n= 18)

Data			
	Years	range	1974-2008
	Unit of analysis		
	Person-level		28%
	Aggregate		72%
	Data types *		
	Medicaid/VA/TRICARE data		33%
	Private health insurance/managed care organization claims		6%
	Proprietary administrative databases		61%
	Registries/outcomes databases		17%
	Electronic medical record/chart reviews		0%
	Data evaluated for completeness/missing		89%
Sample			
	Sample size	range	1,308 – 61,000,000
	Special Populations *		
	Including pediatric		78%
	Including geriatric		50%
FDA Policy *			
	Boxed warnings		50%
	Labeling change		33%
	FDA advisory/safety warnings		39%
	Dear healthcare professional letters		22%
	Withdrawals		17%
Primary Endpoints *			
	Use of targeted medical product		89%
	Substitute product(s) or services		61%
	Laboratory monitoring		17%
	Adverse event(s)		11%
	Contraindicated product(s)/condition(s)		11%
Research Design *			
	Interrupted time series/Time series, including control		39%
	Interrupted time series/Time series, no control		61%
	Pre-post, including control		17%

Statistics are reported in percentages unless otherwise indicated.

* Not mutually exclusive.

Table 3

Novel Research Designs and Analytic Methods Recommended for Evaluating FDA Regulatory Actions

Methods with Strong Internal Validity					
Research Design	**Analytic Method**	**Application Example**	**Necessary Data Elements**	**Strengths**	**Weaknesses**
Interrupted Time Series Designs with control group[7]	Segmented regression analysis. Can include time series estimators (PROC AUTOREG, ARIMA in SAS) using aggregate data or GEE/GLMM using individual data. Specification of model is: 1 parameter for level change and 1 parameter for change in slope. Requires correction for autocorrelation.	For aggregate data: Serumaga 2011.[37] Soumerai 1987.[31] For individual data: Kozhimannil 2011.[38]	Can use person-level or aggregate-level longitudinal data. Data must provide multiple pre- and post-intervention time-points. Short times series (more than 8 pre- and post-time point but <100) are useful but the design becomes stronger with additional time points to establish trends.	Appropriate for policy change that occurs at particular time point(s). One of the most robust of all quasi-experimental designs. Especially strong for testing abrupt changes in levels and slopes.	Requires knowledge of diffusion process, especially in case of gradual interventions or delayed causation. Visual inspection analyses without statistical testing are useful but insufficient for establishing effect.
Regression Discontinuity Designs[7]	Various.	Grootendorst 1997.[39]	Person-level data. Requires an assignment variable (preferably a continuous measure) observed on entire sample.	Appropriate for policy that applies to a defined patient population with fixed eligibility criterion based on a threshold across continuous measure (e.g., age ≥65, or blood pressure values ≥x). A reasonably robust quasi-experimental design. Useful when pre-exposure data are lacking or limited.	Requires knowledge of function form of relationship between outcome and assignment variable. Has less power for detecting small effects. Not well-known in medical literature.

Methods with Strong Internal Validity					
Research Design	**Analytic Method**	**Application Example**	**Necessary Data Elements**	**Strengths**	**Weaknesses**
Various	Extended Cox models with heaviside function[10] (segmented survival analyses).	Briesacher 2010.[40] Zhang 2009.[41]	Person-level longitudinal data.	Appropriate for policy with binary outcome that is expected to change over time. Offers advantages of maximum likelihood (e.g., good large-sample properties). Does not require proportional hazard assumption. Allows for multiple estimates of exposure and outcome relationship as a function of time intervals. Robust to rightcensored data.	Requires accurate specification of time intervals.
Sequential Analysis of gradually accruing data[8,9]	Poisson or Bernoulli maximized sequential probability ratio test, Flexible Exact Sequential Analysis, group sequential methods.	Yih 2011.[42]	Aggregate-level data. Requires rich historical data or appropriate concurrent control groups.	Best suited for early detection of suspected problem as data are accumulating.	Data must be continuously collected at regular intervals over time.
Self-controlled Case Series[43]		Whitaker 2008.[34]	Person-level longitudinal data.	Best suited for studying association between transient exposure (e.g., temporary policy) and acute event (i.e., risk periods are short).	Highly sensitive to time interval between exposure and the event. Less appropriate for longterm monitoring or variable risk period.
Difference-in-difference-in-differences[33]	Various.	Afendulis 2011.[44]	Person-level or aggregate-level data.	Improvement over difference-in difference models. Best suited for studies with limited measurement of pre-post observation.	No controls for historical trends.

Methods with Strong Internal Validity					
Research Design	**Analytic Method**	**Application Example**	**Necessary Data Elements**	**Strengths**	**Weaknesses**
Various	Summary Confounding Scores (propensity scores, disease risk scores).	Morrato 2009.[45] Wharam 2011.[46]	Person-level.	Best suited as sensitivity analyses to augment more rigorous study design.	Does not control for unmeasured confounding.